

# DIGITAL PUBLIC

# HUMANITIES

# OPEN

# CULTURE

# RETI SOCIALI

ASSOCIAZIONE per  
l'INFORMATICA UMANISTICA  
e la CULTURA DIGITALE



# AIUCD 2021

**AIUCD 2021 - DH per la società: e-guaglianza, partecipazione, diritti e valori nell'era digitale.**

Raccolta degli abstract estesi della 10<sup>o</sup> conferenza nazionale

**AIUCD 2021 - DHs for society: e-quality, participation, rights and values in the Digital Age.**

Book of extended abstracts of the 10<sup>th</sup> national conference

# TECH

# ECONOMY

# E-PARTICIPATION

# TECNOLOGIE

# ASSISTIVE





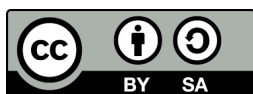
ISBN: 9788894253559

Copyright ©2021 AIUCD  
Associazione per l'Informatica Umanistica e la Cultura Digitale



Il presente volume e tutti i contributi sono rilasciati sotto licenza  
Creative Commons Attribution Share-Alike 4.0 International license ([CC-BY-SA 4.0](https://creativecommons.org/licenses/by-sa/4.0/)).  
Ogni altro diritto rimane in capo ai singoli autori.

This volume and all contributions are released under the  
Creative Commons Attribution Share-Alike 4.0 International license ([CC-BY-SA 4.0](https://creativecommons.org/licenses/by-sa/4.0/)).  
All other rights retained by the legal owners.



Federico Boschetti, Angelo Mario Del Grosso, Enrica Salvatori (edd.), *AIUCD 2021 - DH per la società: e-guaglianza, partecipazione, diritti e valori nell'era digitale. Raccolta degli abstract estesi della 10ª conferenza nazionale*, Pisa, 2021.  
Federico Boschetti, Angelo Mario Del Grosso, Enrica Salvatori (edd.), *AIUCD 2021 - DHs for society: e-quality, participation, rights and values in the Digital Age. Book of extended abstracts of the 10th national conference*, Pisa, 2021.

Immagine di copertina | cover image: Chiara Mannari, Università di Pisa.  
Editing: Mario Valori  
Gestione bibliografica | Bibliographic standardization: Gaia Sitri

Ogni link citato era attivo al 14 gennaio 2021, salvo ove diversamente indicato.  
All links have been visited on 19th January 2021, unless otherwise indicated

Si prega di notificare all'editore ogni omissione o errore si riscontri, al fine di provvedere alla rettifica.  
Please notify the publisher of any omissions or errors found, in order to rectify them.  
**[aiucd.segreteria \[at\] aiucd.org](mailto:aiucd.segreteria@aiucd.org)**



Gli abstract estesi pubblicati in questo volume hanno ottenuto il parere favorevole da parte di valutatori esperti della materia, attraverso un processo di revisione anonima mediante double-blind peer review sotto la responsabilità del Comitato Scientifico di AIUCD 2021.

All the extended abstracts published in this volume have received favourable reviews by experts in the field of DH, through an anonymous double-blind peer review process under the responsibility of the AIUCD 2021 Scientific Committee.

Il programma della conferenza AIUCD è disponibile online  
The AIUCD 2021 conference program is available online  
<https://aiucd2021.labcd.unipi.it/>

### **Comitato di programma - Programme committee**

Enrica Salvatori (UNIFI – LabCD – AIUCD) Conference Chair  
Angelo Mario Del Grosso (CNR-ILC – AIUCD ) Conference co-Chair  
Vittore Casarosa (CNR-ISTI)  
Francesca Frontini (CNR-ILC – CLARIN-ERIC)  
Monica Monachini (CNR-ILC – CLARIN-IT)  
Gianluca Miniaci (UNIFI)  
Angelica M. Puddu (UNIFI – KRINO)  
Maria Simi (UNIFI)  
Simona Turbanti (UNIFI)  
Gigliola Vaglini (UNIFI)

### **Comitato scientifico - Scientific committee**

Federico Boschetti (CNR-ILC – VeDPH-UNIVE) General Chair  
Luca De Biase (UNIFI)  
Roberto Delle Donne (UNINA)  
Alessandro Lenci (UNIFI)  
Barbara Leporini (CNR-ISTI)  
Cristina Marras (CNR-ILIESI)  
Flavia Marzano (LINK CAMPUS UNIVERSITY)  
Anna Monreale (UNIFI)  
Susanna Pelagatti (UNIFI)  
Ginevra Peruginelli (CNR-IGSG)  
Maria Chiara Pievatolo (UNIFI)  
Gino Roncaglia (ROMA TRE)  
Arjuna Tuzzi (UNIPD)  
Giulia Venturi (CNR-ILC)

### **Enti organizzatori / Organizing institutions:**

AIUCD, LabCD dell'Università di Pisa, CLARIN, ILC-CNR, ISTI-CNR

## **Lista dei revisori - List of the reviewers**

Agnese Addone, Alessandra Donnini, Alessandro Lenci, Alessandro Perissinotto, Alina Sirbu, Ana Iglesias Maqueda, Andrea Bellandi, Angelica Lo Duca, Angelo Mario Del Grosso, Anna Galluzzi, Anna Maria Tammaro, Antonio Narzisi, Armando Stellato, Barbara Guidi, Barbara Leporini, Benedetta Iavarone, Caterina Senette, Chiara Mannari, Christian D'Agata, Claudio Forziati, Cristiano Amendola, Cristina Marras, Cristina Pattuelli, Daniela Fogli, Daniela Rotelli, Daria Spampinato, Deborah Paci, Dino Buzzetti, Dominique Brunato, Edmondo Grassi, Elisa D'Argenio, Emanuela Reale, Emanuele Luigi Colazzo, Emiliano Giovannetti, Emma Pietrafesa, Enrica Salvatori, Eva Sassolini, Fabiana Guernaccini, Fabio Ciotti, Fabio Ciraci, Fabio Pacini, Federico Boschetti, Federico Meschini, Francesca Di Donato, Francesca Frontini, Francesca Levi, Francesca Pratesi, Francesca Tomasi, Francesco Grisolia, Francesco Romano, Franz Fischer, Gino Roncaglia, Giorgio Guzzetta, Giorgio Maria Di Nunzio, Giovanni Michetti, Giovanni Scarafile, Giuliana Vitiello, Giulio Mori, Giulio Rossetti, Giuseppe Andrea L'abbate, Hansmichael Hohenegger, Ioanna Miliou, Irene Sucameli, Javier Gomez, Jisu Kim, Jorge Morato, Juan Carlos Torrado Vidal, Laura Giarré, Laura Pollacci, Letizia Cortini, Lucia Ferlino, Luigi Bambaci, Luigi Catalani, Luigi Siciliano, Marco Manca, Maria Chiara Pievatolo, Maria Claudia Buzzi, Maria Simi, Mariasole Bondioli, Marilena Daquino, Marina Buzzi, Marina Pietrangelo, Marion Hersh, Massimiliano Gambardella, Massimiliano Grava, Massimo Magrini, Massimo Zancanaro, Matteo Sanfilippo, Maurella Della Seta, Maurizio Caminito, Maurizio Lana, Maurizio Vivarelli, Michela Natilli, Michele Coscia, Mirko Tavosanis, Nadia Sansone, Nicola Lettieri, Paola Galimberti, Paolo Bottoni, Paolo Mogorovich, Paolo Monella, Paolo Rossi, Pierluigi Feliciati, Rachele Sprugnoli, Raffaele Gareri, Riccardo Del Gratta, Roberto Delle Donne, Roberto Pellungrini, Roberto Rosselli Del Turco, Rolando Minuti, Ruggero G. Pensa, Salvatore Citraro, Sara Conti, Sebastiano Faro, Serge Noiret, Silvia Corbara, Silvia Mirri, Silvia Piccini, Simona Turbanti, Simone Reborà, Simonetta Montemagni, Stefano Chessa, Stefano Rollo, Susanna Pelagatti, Timothy Tambassi, Tito Orlandi, Tiziana Mancinelli, Tiziana Possemato, Tommaso Agnoloni, Vindice Deplano, Vittore Casarosa.









# Prefazione

Il decimo convegno annuale dell'Associazione per l'Informatica Umanistica e la Cultura Digitale ha nell'edizione di quest'anno un titolo peculiare e importante: *DH per la società: e-guaglianza, partecipazione, diritti e valori nell'era digitale*.

Questo è in linea con lo spirito più profondo dell'Associazione, che fin dalle sue origini ha voluto essere un luogo dove filologi, bibliotecari, storici, giuristi, informatici e ingegneri informatici si incontrano e si confrontano (informatica umanistica) per incidere su aspetti rilevanti della società (culture digitali). Abbiamo deciso quindi di discutere su domande forti invitando ognuno a uscire dalla propria *comfort zone* per praticare l'impegno civile necessario a far parlare accademia e società.

Il direttivo dell'associazione e gli organizzatori si sono di fatto impegnati in questa occasione per fornire un aiuto concreto al cambiamento epocale in atto. Il mondo dei negozi sotto casa, delle lezioni tenute solo in presenza, della tv, dell'ufficio postale, del telefono ha lasciato spazio ad aziende di commercio elettronico su scala mondiale, alla didattica a distanza (DAD), alle *newsletter*, ai *social network*, alle *chat* sugli *smartphone*. È una rivoluzione in atto già da tempo e, oggi, abbiamo urgente e profondo bisogno di strumenti concettuali e di ricerca per gestirla. Perché le rivoluzioni non portano solo benefici, ma anche disagi e diseguaglianze. Le *Digital Humanities* (DH) sono il campo di studi in cui si elaborano e si applicano gli strumenti per gestire questo cambiamento. La conferenza vuole quindi rappresentare proprio un momento di approfondimento e di riflessione dell'informatica umanistica come luogo privilegiato di incontro tra i diversi bisogni della società contemporanea, perché siamo convinti che si debba restituire all'umanista il ruolo di chi interpreta e accompagna il cambiamento, di chi favorisce la cultura aperta, la partecipazione, i diritti, i nuovi e antichi valori.

In sostanza c'è la necessità di passare da una fase in cui le *Digital Humanities* sono prevalentemente concentrate sui metodi e sui mezzi, ad una fase in cui esse? tornino a ragionare sui fini, vale a dire sugli obiettivi di ricerca e sui benefici prodotti dal raggiungimento di tali obiettivi per la società nel suo complesso. Nel decimo libro della *Repubblica*, Platone fa notare che «per ogni oggetto esistono tre arti: quella che ne farà uso, quella che lo realizzerà e quella che lo imiterà ..» «Ma la virtù, la bellezza, la perfezione di ogni singolo oggetto riguardano soltanto l'uso per il quale ciascuno di essi è fabbricato», e allora «chi adopera ogni singolo oggetto deve per forza averne la maggiore esperienza e riferire al fabbricante i pregi e i difetti che si rivelano all'uso; ad esempio un flautista dà spiegazioni al costruttore di flauti sugli strumenti che gli servono nel suo mestiere e gli ordinerà come fabbricarli» (601d-e, trad. Caccia). L'umanista digitale non costruisce né flauti né pifferi, anche se talvolta viene scambiato ingiustamente per un pifferaio magico, ma collabora alla progettazione di strumenti che servono ad abbattere le barriere fra la conoscenza o l'esperienza estetica e la nuova società inclusiva. Non si deve tuttavia cadere nel tranello secondo cui il digitale per sé, grazie ad un indefinito potere taumaturgico, abbatte le barriere. Al contrario, uno strumento digitale può innalzare barriere nuove, se non è stato progettato bene, se non è stato pensato per essere così duttile da adattarsi alle esigenze specifiche di chi lo usa.

Ecco perché ad alcuni di noi stanno particolarmente a cuore, fra le molteplici tracce del convegno, il tema della cultura aperta – o cultura dell’apertura – e il tema dell’accessibilità. La cultura dell’apertura ci spinge ad essere esigenti, a voler essere sempre più liberi di usare gli strumenti che ci vengono consegnati in modi inattesi rispetto alle previsioni dei costruttori stessi: noi i libri elettronici non li vogliamo solo leggere (per quello basta il libro a stampa!) o leggere ovunque (anche in questo caso, un libro tascabile può bastare); noi vogliamo costruire indici, estrarre e mettere in relazione dati, visualizzare nuvole di parole, usufruire dei risultati di raffinate analisi semantiche..

Infine, ragionare tutti insieme sul tema dell’accessibilità alle risorse digitali ci fa vedere come negli ultimi decenni, silenziosamente e quasi invisibilmente, la società sia diventata di fatto più inclusiva, dove le persone con disabilità sensoriali, motorie o cognitive, o i discenti nelle varie fasi dell’età evolutiva, non sono affatto beneficiari passivi di risorse e strumenti creati per loro, ma sono protagonisti attivi nella progettazione di risorse, strumenti, percorsi ideati e, grazie a questo, migliori per tutti.

L’incarico di organizzare il decimo convegno dell’Associazione AIUCD fu dato a Pisa un anno fa e, in particolare, annunciato durante la sessione di chiusura della nona conferenza annuale dell’associazione, svoltasi a Milano dal 15 al 17 gennaio 2020, presso l’Università Cattolica del Sacro Cuore. Per gli amanti della cabala, erano le 17 circa del 17 gennaio 2020, ed era un venerdì. Già allora era manifesto il presagio che il nostro sarebbe stato un lavoro complicato: ma all’epoca la straordinaria emergenza sanitaria dovuta alla diffusione del virus SARS-COV-2 era ancora soltanto una epidemia locale di una regione della Cina, e la sola sfida che ci sembrava di dover affrontare era relativa alla proposta del nuovo sguardo che le DH avrebbero avuto verso la società.

Da quel giorno fino ad oggi, molte istituzioni del variegato mondo della ricerca pisana hanno lavorato in grande sinergia per preparare al meglio l’evento associativo più importante per la comunità italiana di informatica umanistica.

Nonostante le notevoli difficoltà incontrate a causa del COVID-19 e alle conseguenti misure di sicurezza via via deliberate, possiamo dire che i “numeri” del convegno danno la misura di un serio interesse verso i temi proposti e di un conseguente indice di successo.

Le date e i dati principali della decima conferenza si possono riassumere in 118 contributi inviati, dalla data di pubblicazione della call for paper – il 19 giugno 2020 – fino al termine stabilito del 2 ottobre 2020. Dei contributi inviati, 96 lavori sono stati presentati come comunicazione orale e i restanti 22 presentati come poster. Il processo per l’invio delle proposte ha previsto la scelta di uno tra i sei possibili temi della conferenza: *Digital public humanities*, *Open culture*, *Reti sociali*, *Tech-economy*, *e-Participation* e infine *Tecnologie assistive per l’inclusione*.

La distribuzione dei contributi ha visto 35 proposte inviate per il tema *Digital public humanities*; 43 per *Open Culture*; 14 *Reti sociali*; 3 *Tech-economy*; 6 *e-Participation*; e 17 *Tecnologie assistive per l’inclusione*.

Promuovendo una politica di inclusione intesa a concedere ampio spazio di discussione e promozione per il maggior numero di attività di ricerca possibile, sono state accettate 100 proposte (85% circa), revisionate da un totale di 125 colleghi per un totale di 318 revisioni (per una media di 2.7 revisioni per contributo).

Delle 100 proposte, 60 sono state accettate come presentazioni orali (63% circa delle proposte inviate per una presentazione orale) e 40 come poster, di cui 18 proposte erano originariamente state inviate come poster (82% circa sul totale delle proposte inviate come poster) e 22 erano state pensate dagli autori come talk (23% circa del totale delle proposte originariamente inviate come presentazioni orali).

Il numero totale degli autori relativi ai contributi inviati è stato pari a 295, di cui 268 sono autori di proposte accettate.

Nella giornata di pre-convegno del 19 gennaio sono stati inoltre organizzati quattro workshop:

1. *Introduzione alle edizioni digitali: preparazione con codifica XML TEI e visualizzazione con il software EVT*, a cura di Roberto Rosselli Del Turco;
2. *Narrativa e divulgazione scientifica delle DH: l'esperienza dei QUARANTIP*, a cura di Krino;
3. *Qui CLARIN-IT: posso aiutarti? / This is CLARIN, how can we help you?* a cura di Monica Monachini e Francesca Frontini;
4. *ALDiNa, Archivi Letterari Digitali Nativi*, a cura di Tiziana Mancinelli, Emanuela Carbè e Federico Boschetti

Infine l'ultimo numero, quello dei circa 650 iscritti al convegno: frutto non solo dell'apertura delle tematiche, ma, indubbiamente, anche della nuova modalità di partecipazione a distanza. Questo credo ci dovrà far molto riflettere su cosa fare in futuro, a pandemia speriamo archiviata, quando dovremo ragionare su come conciliare il piacere e l'utilità dell'incontro in presenza con il dovere morale di favorire una partecipazione democratica.

La conferenza, allargandosi a temi fino ad oggi non molto dibattuti entro l'associazione, ha comportato diverse aperture verso i ricercatori attivi negli ambiti della sanità, dell'economia, della giurisprudenza, delle reti sociali. Un'apertura che ci ha ovviamente arricchito, ma che ci ha anche posto nuove sfide. Le DH da anni sono una metadisciplina dai confini sfumati e in via di definizione, una galassia dalle numerose e diverse componenti: quale può essere la forza aggregante che le tiene insieme? Forse ce lo potranno dire i giovani partecipanti al convegno: numerosi studenti e neolaureati quest'anno hanno mandato le loro proposte che sono state valutate e accettate dai revisori. Un'ulteriore apertura di cui siamo particolarmente orgogliosi.

Ogni comunità di studiosi e studenti con cui AIUCD viene quotidianamente in contatto ha il proprio linguaggio specialistico e le proprie buone pratiche. Speriamo che la condivisione di tali linguaggi e di tali pratiche promuova la co-evoluzione delle diverse comunità che hanno sentito spontaneamente la necessità di accorciare le rispettive distanze. Paradossalmente – ma forse non tanto – la distanza fisica imposta dalle circostanze ha promosso, nelle estenuanti videoconferenze dell'ultimo anno, il contatto virtuale di persone appartenenti a mondi epistemologicamente paralleli.

Non ci bastava aver proposto un tema nuovo e nuove aperture. Ci ha messo lo zampino anche questa terribile pandemia, costringendoci a organizzare un evento completamente on line. Dal 12 marzo 2020 il nostro paese è entrato in lockdown e tutti gli eventi di socialità e collettivi sono stati di fatto fortemente limitati o comunque sono stati trasformati in eventi virtuali con partecipazione remota. La stessa sorte è toccata alla conferenza AIUCD2021.

Ci siamo trovati quindi subito a dover valutare le piattaforme più adatte e robuste per ospitare il convegno in modalità remota. Ovviamente non potevamo replicare in rete un convegno tradizionale: l'Associazione sa bene che il mezzo digitale può e deve mutare le forme di comunicazione. Si è trattato quindi di un'esperienza decisamente diversa, più interattiva, meno formale e con "ritualità" diverse rispetto al passato.

Chiusi nelle proprie case, isolati nelle proprie città, i nostri autori si sono trovati costretti a fare ciò che fino a poco tempo fa poteva essere, per l'umanista, solo un'opzione: molti si sono trovati a dover usare in modo massiccio strumenti per la scrittura collaborativa, a dover modificare quindi non solo le proprie abitudini ma anche i propri paradigmi cognitivi. In questo spirito di collaborazione sono state organizzate anche le sessioni del convegno, prendendo a modello l'ultima conferenza CLARIN.

Fortunatamente abbiamo avuto fin da subito un grande appoggio e una grande disponibilità dei mezzi e delle risorse proprio dall'infrastruttura CLARIN, che sia nella sua declinazione Italiana, che è partner della conferenza, sia nella sua declinazione di coordinamento europeo ci ha permesso di organizzare l'evento completamente online.

Per questo il comitato organizzatore e di programma della conferenza vuole rivolgere un sentito e accorato ringraziamento agli amici di CLARIN-IT, soprattutto nelle persone di Monica Monachini, coordinatrice del nodo Italiano di CLARIN, e di Francesca Frontini, da poco entrata nel board europeo di CLARIN-ERIC. In più, un doveroso ringraziamento anche per i tanti colleghi di CLARIN-ERIC, soprattutto nella persona di Franciska de Jong, direttrice esecutiva, e di Maria Eskevich (Central Office Coordinator), che ha tenuto i contatti con gli organizzatori della conferenza.

Ugualmente grati siamo ai giovani studenti di Krino per il supporto nelle giornate del convegno, all'Università di Pisa, all'ISTI-CNR e all'ILC-CNR, co-organizzatori assieme ad AIUCD, e infine allo staff di Media Events di UNIPI per il servizio di post-produzione video.

Federico Boschetti, Angelo Mario Del Grosso ed Enrica Salvatori

# **Sessione I**

## **Pierre de Coubertin**



# RESTORE: smaRt accESs TO digital heRitage and mEmory

Emiliano Degl'Innocenti<sup>1</sup>, Alessia Spadi<sup>2</sup>, Federica Spinelli<sup>3</sup>

<sup>1</sup> OVI-CNR – emiliano.deglinnocenti [at] cnr.it

<sup>2</sup> OVI-CNR – spadi [at] ovi.cnr.it

<sup>3</sup> OVI-CNR – spinelli [at] ovi.cnr.it

## SINTESI

Il progetto RESTORE (smaRt accESs TO digital heRitage and mEmory) ha preso avvio il 1 giugno 2020 e prevede una durata di due anni. Al progetto, coordinato dall'Istituto Opera del Vocabolario Italiano del CNR, partecipano l'Archivio di Stato e il Museo di Palazzo Pretorio di Prato, la Soprintendenza Archivistica e Bibliografica della Toscana oltre all'azienda SPACE SpA.

Il progetto - cofinanziato dalla Regione Toscana - ha come scopo principale il recupero, l'integrazione e l'accessibilità di dati e oggetti digitali prodotti negli ultimi vent'anni dai suoi partner, al fine di costituire una base di conoscenza sulla storia della città e delle sue istituzioni, lo sviluppo del suo tessuto economico e imprenditoriale, il ruolo delle donne nello sviluppo della rete del welfare cittadino. A partire dalla figura del mercante Francesco di Marco Datini, della sua famiglia e del suo entourage, dalla dimensione locale sarà possibile ricostruire una parte rilevante della storia delle città dell'Europa e del Mediterraneo del XIV secolo, inclusi gli aspetti commerciali ed economici.

Il contesto di riferimento per le attività scientifiche e di sviluppo infrastrutturale è costituito dalla collaborazione (attraverso il CNR-OVI) con le infrastrutture DARIAH-ERIC (ESFRI1 Landmark per le discipline umanistiche e le scienze sociali) ed E-RIHS (progetto ESFRI per la scienza del patrimonio), oltre che dalla Cloud Europea della Ricerca (EOSC).

## ENGLISH ABSTRACT

The RESTORE project (smaRt accESs TO digital heRitage and mEmory) started on June 1, 2020, and has a duration of two years. The project, coordinated by the Istituto Opera del Vocabolario Italiano of the CNR, includes the State Archives and the Museum of Palazzo Pretorio in Prato, the Archival, and Bibliographic Superintendence of Tuscany as well as the SPACE SpA company.

The project - co-financed by the Tuscany Region - has as its main purpose the recovery, integration, and accessibility of data and digital objects produced in the last twenty years by its partners, in order to build a knowledge base on the history of the city and its institutions, the development of its economic and entrepreneurial system, the role of women in the development of the welfare network. Starting from the figure of the merchant Francesco di Marco Datini, his family, and his entourage, from the local dimension it will be possible to reconstruct a significant part of the history of European and Mediterranean cities of the 14th century, including commercial and economic aspects.



The reference landscape for the scientific and infrastructural development of the project's activities is constituted by the collaboration (through the CNR-OVI) with the DARIAH-ERIC (ESFRI Landmark for the Humanities and Social Sciences) and E-RIHS (ESFRI project for the Heritage Science), as well as the European Research Cloud (EOSC).

## **PAROLE CHIAVE**

corpora testuali annotati; collezioni; archivi; metodologie e tecnologie per la produzione, rappresentazione, conservazione e valorizzazione del patrimonio culturale; modelli e strumenti per la rappresentazione semantica e la gestione della conoscenza in ambito umanistico

## **INTERVENTO**

### **1 INTRODUZIONE**

RESTORE, progetto finanziato nel contesto del bando POR-FSE 2014-20 della Regione Toscana, si fonda su un partenariato tra enti di ricerca, istituti culturali e PMI, con l'obiettivo di recuperare e mettere a sistema le risorse digitali prodotte dai partner in diversi contesti, attività ed epoche, migliorandone la fruizione da parte di ricercatori, professionisti delle industrie culturali e creative e cittadini interessati alla storia della città di Prato e delle sue istituzioni, allo sviluppo del suo tessuto economico e imprenditoriale, al ruolo delle donne nello sviluppo della rete del welfare cittadino e - attraverso le vicende di Margherita Datini - dell'Ospedale degli Innocenti di Firenze e della Fondazione dei Ceppi.

Il sito web del progetto è disponibile all'indirizzo – <http://restore.ovi.cnr.it>; tutti gli URL di siti e risorse citati nel testo risultavano accessibili alla data del 26 novembre 2020.

### **2 STATO DELL'ARTE E OBIETTIVI**

L'ecosistema delle risorse digitali che a vario titolo riguardano il patrimonio culturale è ampio e molto ricco, ma soffre di un livello di frammentazione talmente alto da rischiare di comprometterne il valore: i contenuti di qualità – prodotti da biblioteche, archivi e centri di ricerca - oltre a risultare spesso difficilmente reperibili e scarsamente accessibili, sono anche caratterizzati da un livello molto basso di interoperabilità, giacché questi soggetti lavorano spesso in contesti operativi e scientifici molto specifici, quando non in totale isolamento. La tendenza a generare moli cospicue di informazioni digitali, che già nel 2002 aveva portato Martin Hilbert [5] ad individuare una svolta nell'accumulazione della conoscenza a livello globale (i.e. “the beginning of the digital age”<sup>1</sup>: per la prima volta nella storia dell'umanità vi sono più dati prodotti e conservati nel dominio digitale che nella controparte analogica), accelera di anno in anno.

---

1 [https://en.wikipedia.org/wiki/Digital\\_data#cite\\_ref-Hilbertvideo2011\\_3-0](https://en.wikipedia.org/wiki/Digital_data#cite_ref-Hilbertvideo2011_3-0)

Una simile ricchezza ed eterogeneità di informazioni suggerisce, anche per il settore del Patrimonio Culturale, una attenta valutazione dei criteri da utilizzare per la selezione, strutturazione, arricchimento e pubblicazione di dati, servizi e strumenti - di alta qualità, scientificamente validati - che altrimenti rimarrebbero intrappolati nei sistemi informativi dei relativi creatori: spesso scarsamente interoperabili e, in alcuni casi, resi del tutto inaccessibili dalla mancanza di standard di riferimento.

Alcune fra le più recenti tendenze della ricerca storica [1] valutano positivamente l'impatto che un approccio innovativo - basato sulla datizzazione delle fonti, la loro integrazione in un ecosistema digitale fortemente interoperabile e l'utilizzo di strumenti e metodi digitali di analisi dei dati - possono avere sul tentativo di superare alcuni dei limiti dell'approccio storicistico tradizionale che, a causa di barriere disciplinari, linguistiche, geografiche e semantiche, risulterebbe incapace di restituire un quadro complessivo dell'evoluzione di una comunità (i.e.: una città, una nazione, ecc.) ed una descrizione sufficientemente approfondita delle dinamiche di interrelazione che ne caratterizzano le diverse entità (i.e.: persone, eventi, istituzioni, concetti, ecc.).

Il progetto RESTORE intende fornire una ricostruzione multilivello della realtà sociale pratese a partire dal Medioevo e - in prospettiva - fino al Novecento, attraverso la realizzazione di un sistema informativo basato principalmente sull'analisi di una porzione rilevante del patrimonio documentario cittadino, permettendone la consultazione da remoto con strumenti di ricerca avanzati che la rendano facilmente fruibile e riutilizzabile sotto forme diverse (ricerche scientifiche, ricerche genealogiche, didattica, divulgazione scientifica).

Anche in Italia, a fronte di numerosi interventi, supportati da finanziamenti pubblici erogati nell'ultimo decennio, sono state prodotte collezioni digitali (i.e.: immagini, metadati, corpora testuali, strumenti di consultazione, thesauri e tassonomie specializzate) caratterizzate da scarsa o assente interoperabilità e/o dalla mancanza di adeguate strategie di sostenibilità sul medio-lungo periodo: accessibili attraverso portali dedicati oppure, nei casi peggiori, totalmente inaccessibili.

### 3 LE RISORSE

#### **Fonti archivistiche, collezioni museali, corpora testuali**

L'Archivio di Stato di Prato<sup>2</sup> conserva un vasto patrimonio di fonti provenienti da soggetti civili, religiosi e assistenziali che, integrate con quelle di altre istituzioni pratesi fra cui quelle relative al patrimonio artistico del Museo di Palazzo Pretorio<sup>3</sup>, consentono di ricostruire la storia della Comunità cittadina, del suo territorio e della sua popolazione, dal XIV secolo e in prospettiva fino ai giorni nostri.

Per i due fondi più antichi e ricchi di informazioni storiche sulla popolazione, ovvero Datini<sup>4</sup> e Ospedale Misericordia e Dolce<sup>5</sup>, si dispone della descrizione archivistica informatizzata, delle immagini di-

---

2 <http://archiviodistato.prato.it>

3 <http://www.palazzopretorio.prato.it/it/>

4 Archivio di documenti di amministrazione e carteggi del mercante Francesco di Marco Datini (1335-1410) che testimonia, attraverso la sua vastissima attività in campo industriale, commerciale e bancario uno spaccato di economia e vita sociale dell'intero bacino Mediterraneo. Il progetto di digitalizzazione è iniziato nel 1999.  
<http://datini.archiviodistato.prato.it/il-progetto/>

5 Ente di assistenza che si è preso cura di viandanti, poveri e bambini abbandonati a partire dal XIII secolo.  
<https://sias.archivi.beniculturali.it/cgi-bin/pagina.pl?TipoPag=comparc&Chiave=432072>  
Le risorse digitali relative a questo fondo sono consultabili sul sito

gitali dei documenti, della trascrizione dei testi e di un corpus testuale lemmatizzato prodotto dall'OVI<sup>6</sup>.

In passato, alcuni finanziamenti hanno permesso la realizzazione di strumenti digitali di ricerca (CD-ROM Per la tua Margherita, sito web Datini online, ecc.), che col passare del tempo risultano scarsamente accessibili, poco interoperabili quando non addirittura del tutto inutilizzabili. In generale inoltre, tali progetti risultano scollegati fra sé e dalla descrizione archivistica presente sul sito dell'Archivio di Stato di Prato e del Sistema Informativo degli Archivi di Stato, privando così tali lavori del valore aggiunto dato dall'integrazione delle informazioni.

Obiettivo di RESTORE è strutturare queste risorse secondo l'approccio FAIR<sup>7</sup> [7] e integrarle con nuovi dati e oggetti digitali provenienti da altri istituti, tra cui l'OVI e il Museo di Palazzo Pretorio<sup>8</sup>. Quest'ultimo conserva il patrimonio artistico cittadino, il cui nucleo più consistente proviene proprio dall'Ospedale Misericordia e Dolce, con opere realizzate espressamente per l'istituzione ospedaliera o raccolte da famiglie pratesi o enti religiosi e confluite in essa.

Il Museo ha realizzato campagne di digitalizzazione delle proprie opere e di informatizzazione del catalogo, utilizzando una piattaforma commerciale aderente agli standard di settore<sup>9</sup> e ha selezionato una serie di fonti, rilevanti per gli scopi del progetto, ancora da digitalizzare. Se la lettura delle immagini può facilitare l'interpretazione o lo studio di testimonianze archivistiche, l'attenzione ai documenti, aiuta a sciogliere nodi attributivi, stilistici e/o di committenze, altrimenti difficilmente risolvibili: per questo motivo si rende necessaria la creazione di una base di conoscenza integrata, in grado di abbattere le esistenti barriere disciplinari, istituzionali e tecnologiche e renderle accessibili con modalità innovative, in modo da far conoscere e valorizzare, rendendola più facilmente navigabile e comprensibile, una ricca porzione documentale altrimenti poco nota, in grado di gettare luce su temi fortemente radicati nella storia pratese e - in genere - toscana (il tessuto produttivo, la rete assistenziale, il ruolo delle donne in entrambi questi contesti, l'infanzia abbandonata), che toccano da vicino tutti gli strati della cittadinanza e mirano a coinvolgere la popolazione in un percorso culturale di conoscenza e riappropriazione delle proprie origini, permettendo di usufruire della conoscenza del passato per costruire l'inclusione sociale nel presente.

---

<http://www.archiviodistato.prato.it/accedi-e-consulta/aspoSt005/tree>

6 <http://ovi.cnr.it>

7 Dati Ricercabili, Accessibili, Interoperabili e Riutilizzabili:  
<https://www.force11.org/group/fairgroup/fairprinciples>

8 <http://www.palazzopretorio.prato.it/it/le-opere/>

9 L'Archivio di Stato di Prato ha come riferimento gli standard informatici promossi dall'ICA (International Council for Archives), come ISAD e ISAAR che si traducono negli standard di codifica EAD-XML per la descrizione archivistica ed EAC-XML per la gestione delle liste di autorità dei soggetti produttori. Il Museo di Palazzo pretorio ha come riferimento gli standard catalografici promossi dall'Istituto Centrale per il Catalogo e la Documentazione.  
<http://www.iccd.beniculturali.it/it/standard-catalografici>

Nel dettaglio le risorse che verranno trattate per prime sono:

- il fondo Datini, costituito da 150.000 lettere e circa 600 registri da cui si possono ricavare informazioni sulle persone coinvolte, sui costi e le tipologie delle merci, sui luoghi deputati agli scambi ecc., è ad oggi l'archivio mercantile per il Medioevo più grande al mondo;
- il fondo dell'Ospedale Misericordia e Dolce, con le sue 7000 unità archivistiche, presenta tutte le articolazioni delle funzioni di un ente di assistenza: dal sostegno al viandante, alla cura del povero e dell'ammalato, fino all'accoglienza dei gettatelli, bambini abbandonati e allevati, grazie allo stesso Spedale, dall'intera Comunità pratese;
- la collezione del Museo di Palazzo Pretorio è composta da circa 3000 opere tra dipinti, sculture, disegni, tessuti, abiti, mobili e altri manufatti che raccontano la storia artistica e culturale della città.

OVI metterà a disposizione un corpus che raccoglie le lettere edite dell'ampio carteggio dell'Archivio Datini<sup>10</sup> [6], lemmatizzate selettivamente: nel corpus sono registrate le varietà grafiche e morfologiche di una serie di termini rilevanti al fine di contribuire alla ricostruzione della vita economica e giuridica, nonché dei rapporti sociali dell'epoca. Sono stati lemmatizzati i nomi propri di persona, comprensivi di eventuali soprannomi e cariche specifiche (nel caso in cui l'indicazione rimandi ad un preciso personaggio storico individuato); di luogo, nomi di città, paesi, contrade, località, vie, piazze, porte, chiese, monasteri, palazzi, ospedali, enti e istituzioni, ecc. Gli elementi lemmatizzati sono distribuiti entro ventidue categorie che comprendono, oltre ai nomi di persona e di luogo, termini e verbi afferenti al campo religioso o all'agricoltura, le parti del corpo, i nomi della settimana, i termini generici mese e anno, ecc. inclusi: abbigliamento e arredi, alimenti, animali, arti e mestieri, calendario, economia diritto e politica, edilizia e architettura, medicina, monete, navigazione, parentele, pelletteria e tessili, ecc. Il corpus lemmatizzato del carteggio Datini, realizzato dal CNR-OVI tra il 2003 e il 2005, è allestito con lo stesso programma che gestisce il corpus "Tesoro della Lingua Italia delle Origini (TLIO)", in una versione appositamente dedicata e interrogabile via web.

Il corpus consta di:

- 2.511 testi
- 45259 forme
- 977.034 occorrenze di cui 126.663 lemmatizzate
- 6.510 lemmi e 22 iperlemmi.

---

10 Il corpus, realizzato tra il 2003 e il 2005, è consultabile sul sito dell'Istituto Opera del Vocabolario (CNR-OVI)

[http://aspweb.ovi.cnr.it/\(S\(p3pz2q55amjuzs450kpqyy55\)\)/CatForm01.aspx](http://aspweb.ovi.cnr.it/(S(p3pz2q55amjuzs450kpqyy55))/CatForm01.aspx)

Il lavoro è stato attuato attraverso:

- l'identificazione e la digitalizzazione delle lettere edite da immettere nella banca dati e che vanno a costituire un corpus organico;
- le operazioni di OCR e la generazione di files su cui inserire, dopo la correzione dell'output, le codifiche del software GATTO 3.3 (elaborato dall'Istituto CNR Opera del Vocabolario Italiano da D. Iorio-Fili e concesso gratuitamente);
- la messa a punto di una versione personalizzata dello stesso software, con le funzioni richieste dall'Archivio di Stato di Prato;
- la creazione di una banca dati di tutti i testi, con possibilità di ricerca delle forme presenti;
- la lemmatizzazione di termini appartenenti ad alcuni settori di lessico: storico-economico, commerciale (incluso il commercio marittimo), militare, storico-giuridico, tecnicismi;
- l'identificazione e indicizzazione di antroponomi e toponimi;
- la creazione di iperlemmi (lemmi di secondo livello) per categorie di lemmi di particolare significato per la ricerca;
- la creazione di una versione compatibile per web, richiamabile dal sito web del Progetto Datini.

Grazie alle competenze dell'OVI e al supporto tecnologico di SPACE SpA<sup>11</sup>, RESTORE analizzerà le criticità individuate dagli istituti culturali coinvolti, nel contesto della raccolta, aggregazione, arricchimento semantico e messa a disposizione di informazioni ad alto valore aggiunto - relative al patrimonio culturale summenzionato – e costituirà una base di conoscenza popolata con informazioni scientificamente affidabili, rese accessibili, interoperabili e riusabili, secondo i principi FAIR. RESTORE è un progetto pilota che – facendo ampio ricorso alle soluzioni tecniche ed alle componenti software sviluppate dai progetti PARTHENOS<sup>12</sup>, SSHOC<sup>13</sup>, IPERION HS<sup>14</sup> e IDEHA<sup>15</sup> – elaborerà modelli e soluzioni per la gestione, il trattamento e la fruizione del patrimonio culturale e documentario, replicabili in maniera sostenibile (sia tecnologicamente che in termini di risorse necessarie al mantenimento) presso altre istituzioni.

---

11 <http://www.spacespa.it/chi-siamo/>

12 PARTHENOS Joint Resources Registry  
<http://146.48.123.109/>

13 Social Sciences and Humanities Open Marketplace  
<https://www.sshopencloud.eu/ssh-open-marketplace>

14 Integrated Platform for the European Research Infrastructure:  
<http://www.iperionhs.eu/about/>

15 Innovation for Data Elaboration in Heritage Areas:  
[https://www.ispc.cnr.it/it\\_it/2020/06/03/ideha/](https://www.ispc.cnr.it/it_it/2020/06/03/ideha/)

La base di conoscenza risultante da queste attività sarà interoperabile con una ampia rete di analoghe iniziative a livello nazionale e internazionale, attraverso la collaborazione con il cluster tematico Social Sciences and Humanities Open Cloud (SSHOC)<sup>16</sup> e con le infrastrutture digitali di ricerca e di servizi attive nell'ambito delle discipline umanistiche (DARIAH<sup>17</sup>, EUROPEANA<sup>18</sup>) e della scienza del patrimonio (E-RIHS<sup>19</sup>).

## 4 ATTIVITÀ PREVISTE

Le attività di RESTORE sono iniziate nel mese di giugno del 2020 e proseguiranno per 24 mesi. Nella prima fase (da giugno a dicembre 2020) saranno analizzate le risorse messe a disposizione dai partner e le differenti problematiche connesse alla loro integrazione, in funzione della costituzione della base di conoscenza.

Oltre a definire un modello concettuale adeguato alla gestione delle informazioni provenienti dai vari contesti di ricerca e produzione (i.e.: biblioteche, archivi, musei etc.) [3, 4], che le renda utilizzabili per processi di apprendimento automatico ed estrazione della conoscenza realizzati seguendo diversi approcci (i.e.: NLP<sup>20</sup>, NER<sup>21</sup> etc.), verranno implementati tool per la mappatura sintattica e semantica degli schemi e standard più rilevanti in ciascun settore (i.e.: TEI<sup>22</sup> per i testi; EDM<sup>23</sup> oltre a MAG<sup>24</sup>, MODS e METS<sup>25</sup> per contenuti prodotti da biblioteche; EAD<sup>26</sup> ed EAC<sup>27</sup> per gli archivi; CIDOC-CRM<sup>28</sup> per contenuti prodotti in altri ambiti del Digital Cultural Heritage, ecc.) e la trasformazione dei dataset [2].

---

16 <https://www.sshopencloud.eu/>

17 <https://www.dariah.eu/>

18 <https://www.europeana.eu/it>

19 <http://www.e-rihs.eu/>

20 Natural Language Processing  
[https://parthenos.d4science.org/web/parthenos\\_vre](https://parthenos.d4science.org/web/parthenos_vre)

21 Named Entities Extraction:  
[https://parthenos.d4science.org/web/parthenos\\_vre](https://parthenos.d4science.org/web/parthenos_vre)

22 Text Encoding Initiative:  
<https://tei-c.org/>

23 Europeana Data Model:  
<https://pro.europeana.eu/page/edm-documentation>

24 Metadati Amministrativi e Gestionali  
<https://www.iccu.sbn.it/export/sites/iccu/documenti/manuale.html>

25 <https://www.loc.gov/standards/mods/presentations/mets-mods-morgan-ala07>

26 Encoded Archival Description  
<https://www.loc.gov/ead/>

27 Encoded Archival Context  
<https://eac.staatsbibliothek-berlin.de>

28 CIDOC Conceptual Reference Model (CRM)  
<http://www.cidoc-crm.org/>

Allo stato attuale (novembre 2020) - in considerazione del caso d'uso costituito dalle risorse relative a Datini e al suo entourage - dopo una valutazione comparativa di diverse soluzioni open source per la gestione e la pubblicazione di dati, gli asset digitali dei partner sono stati importati e descritti in CKAN<sup>29</sup>. Si tratta, come detto, di un prodotto open source ben documentato e largamente utilizzato da progetti e istituzioni per la pubblicazione di grandi quantità di dati<sup>30</sup>. CKAN supporta nativamente numerosi standard e formati digitali differenti<sup>31</sup>, ulteriormente estendibili tramite appositi plugin, e offre la possibilità di interagire con i contenuti e le funzionalità della base di conoscenza (ricerca, importazione, esportazione, autenticazione, ecc.) attraverso API<sup>32</sup>, oltre a supportare RDF e SPARQL, requisito fondamentale in vista della costituzione di un endpoint per la pubblicazione dei dati.

Su questa base, a partire dal 2021 e fino al termine del progetto pilota basato su Datini e il suo entourage, le attività principali saranno volte alla definizione dell'architettura definitiva del sistema e all'elaborazione di strategie per l'acquisizione, il mapping e la modellazione dei dati forniti dai partner. Parallelamente verranno messe a punto le modalità di estrazione della conoscenza e di esposizione delle informazioni della base di conoscenza risultante, in vista della loro accessibilità, fruizione e riutilizzo, sia attraverso la piattaforma digitale di RESTORE che mediante servizi innovativi (museo virtuale, itinerari tematici smart, ecc.) sviluppati da altri operatori della filiera culturale e creativa.

Al termine del progetto, la piattaforma tecnologica di RESTORE potrà essere aperta a ulteriori partner ed estesa in termini di funzionalità offerte e tipologia di risorse digitali gestite: da questo punto di vista i punti di riferimento saranno costituiti dalle infrastrutture ESFRI attive nel settore delle discipline umanistiche (DARIAH-ERIC) e della scienza del patrimonio (E-RIHS), il cui coinvolgimento garantirà la piena aderenza alle linee guida europee in tema di accessibilità, interoperabilità, riutilizzo e creazione di valore aggiunto per la società a partire dai risultati raggiunti dalla ricerca.

---

29 <https://ckan.org/>

30 <https://ckan.org/about/instances/>

31 <https://ckan.org/features/>

32 <https://docs.ckan.org/en/2.9/api/>

## BIBLIOGRAFIA

1. Bod, Rens. *Le scienze dimenticate: come le discipline umanistiche hanno migliorato il mondo*. Roma: Carocci, 2019.
2. Degl’Innocenti, Emiliano, Roberta Giacomi, Susanne Haaf, Steven Krauwer, Klaus Illmayer, Antonio Davide Madonna, e Maurizio Sanesi. «PARTHENOS D2.4 - Report on the Assessment of Interoperability, Services and Tools», 25 aprile 2018. <https://doi.org/10.5281/ZENODO.2662530>.
3. Gangemi, Aldo, e Valentina Presutti. «Content ontology design patterns as practical building blocks for web ontologies». In *International Conference on Conceptual Modeling*, 128–141. Springer, 2008.
4. ———. «Ontology design patterns». In *Handbook on ontologies*, 221–243. Berlin: Springer, 2009.
5. Hilbert, Martin. «How much information is there in the “information society”?» *Significance* 9, n. 4 (2012): 8–12.
6. Toccafondi, Diana, e Gianni Cascone. «Per la tua Margherita. Lettere di una donna del’300 al marito mercante. Margherita Datini a Francesco di Marco, 1384-1401». Archivio di Stato di Prato, 2002.
7. Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. «The FAIR Guiding Principles for Scientific Data Management and Stewardship». *Scientific Data* 3, n. 1 (15 marzo 2016): 160018. <https://doi.org/10.1038/sdata.2016.18>.





# **GKPS: portale per la condivisione delle conoscenze nell'ambito del rischio clinico**

Giorgia Tani  
FTGM-CNR – gtani [at] ftgm.it

## **SINTESI**

GKPS - Global Knowledge sharing platform for Patient Safety - è un portale nato dall'incontro tra il digitale e la sanità che mira alla condivisione della conoscenza nell'ambito del rischio clinico realizzato in collaborazione con il centro di gestione del rischio clinico della regione Toscana (GRC) e l'Organizzazione mondiale della sanità (OMS).

Il rischio clinico è la disciplina che studia la probabilità che un paziente subisca un danno durante la prestazione delle cure o a causa delle cure sanitarie prestate. Su scala mondiale tali problematiche risentono dei differenti livelli economici, organizzativi, tecnologici e culturali.

La realizzazione di una piattaforma web risulta quindi particolarmente adatta a facilitare lo scambio di informazioni all'interno della comunità internazionale, offrendo uno strumento moderno, accessibile a tutti e dai costi di sviluppo e gestione contenuti. La raccolta via web delle esperienze fatte sul campo e la condivisione delle conoscenze e degli strumenti disponibili rappresentano un'eccellente opportunità che l'OMS ha deciso di supportare.

## **PAROLE CHIAVE**

sicurezza del paziente, rischio clinico, condivisione, internazionalità

## **INTERVENTO**

### **1 INTRODUZIONE**

Il progetto nasce nel 2016 grazie a una collaborazione tra la Fondazione Toscana Gabriele Monasterio del CNR di Pisa, il centro di gestione del rischio clinico della regione Toscana (GRC) e l'Organizzazione mondiale della sanità (OMS) con lo scopo di migliorare la qualità delle cure e la sicurezza dei pazienti, per ottenere una proiezione dei punti deboli e di forza della situazione sanitaria mondiale e condividere l'apprendimento dato dal reporting degli incidenti sulla sicurezza dei pazienti e dall'applicazione di buone pratiche per la prevenzione dei rischi.

La Global Knowledge Sharing Platform for Patient Safety (GKPS) è da intendere come uno spazio pubblico, dove raccogliere, condividere e gestire sistematicamente le conoscenze sulla sicurezza dei pazienti derivanti da differenti contesti economici, geografici e culturali con una metodologia strutturata e sostenibile.

GKPS mira anche a rafforzare la rete di persone che in tutto il mondo lavorano per migliorare la sicurezza e la qualità dell'assistenza sanitaria. Uno degli obiettivi è quello di affrontare i bisogni delle persone sia nei paesi sviluppati che in quelli in via di sviluppo, sottolineando che nessuno dovrebbe essere lasciato indietro e che nella prospettiva GKPS, la sicurezza e la qualità delle cure non dovrebbero essere un lusso.

Un portale di questo genere si rivolge a utenti che operano nell'ambito del rischio clinico e discipline affini, esperti che talvolta potranno essere sia autori che fruitori. Il target in generale viene identificato con lo specialista del settore, anche se ciò non preclude l'accesso alla consultazione da parte di altre figure professionali o di comuni cittadini [1].

## 2 LA PIATTAFORMA

Il sito è interamente scritto in lingua inglese e si articola in diverse sezioni che possono essere raggruppate in due macro-sezioni, la prima di carattere descrittivo che contribuisce a fornire informazioni generali sulla mission del progetto, sul rischio clinico, sugli enti che sono coinvolti e supportano il progetto, sulle politiche editoriali e sui termini d'uso (la compongono nello specifico: *GKPS project*, *About us*, *Partners*, *Contact*, *Links*, *Editorial policy*, *Terms of use*); l'altra offre l'accesso alle informazioni condivise di esperienze e risorse (*Safety incidents and practices*, *Resources on patient safety*, *Resources for RLSs*) e permette all'utente di condividere le proprie esperienze personali sull'argomento (pagina di *Submission*).

A queste se ne aggiunge una interamente dedicata all'emergenza sanitaria COVID-19 (formata dalle pagine *Safety practices on COVID-19* e *Health workers protection*), in cui è possibile consultare e scaricare esperienze relative alle soluzioni messe in atto per garantire la sicurezza degli operatori sanitari e dei pazienti, e buone pratiche implementate a livello di struttura, regionale e nazionale per far fronte alla pandemia. In questa sezione la raccolta e la traduzione in inglese delle pratiche italiane per la sicurezza dei pazienti sono state realizzate in collaborazione con l'Agenzia Nazionale Italiana per i Servizi Sanitari Regionali - Agenas.

All'utente che si trova a navigare all'interno del sito vengono dunque proposte numerose possibilità di interazione con esso, prima tra tutte la ricerca di conoscenze condivise nell'ambito del rischio clinico. Questo tipo di informazione si può trovare sotto forma di resoconto dell'esperienza sul campo oppure di "risorsa", cioè conoscenze di diversa tipologia offerte in forma di file scaricabile. Nel primo caso si tratta di incidenti o di buone pratiche, cioè interventi o azioni realmente accaduti, di diversa complessità, che potrebbero dimostrarsi efficaci nell'evitare, prevenire o mitigare le conseguenze indesiderate dell'assistenza sanitaria. L'utente che si trova nella pagina di ricerca di esperienze legate al rischio clinico può affinare la sua ricerca attraverso l'impostazione di una serie di parametri, tramite una mappa selezionabile o da un form contenente vari filtri. Oltre alla selezione dell'area geografica o dei singoli stati (selezione reciprocamente esclusiva), si possono applicare ulteriori filtri riguardanti il tipo di esperienza, buone pratiche e/o incidenti e una ricerca testuale.

Le pagine di ricerca delle risorse, invece, sono due e si dividono in: risorse sulla sicurezza dei pazienti e risorse RLS. Queste ultime differenziano dalle prime per una maggiore specificità del loro obiettivo, raccolgono infatti materiali di formazione sui sistemi di reporting e apprendimento che possono essere articoli, rapporti, casi di studio, podcast e webinar. Nonostante questa differenziazione in base all'argomento, dal punto di vista della formattazione e della modalità di ricerca dei contenuti sono uguali.

L'utente può filtrare la ricerca per tipologia di risorsa (pubblicazioni, materiali di supporto, linee guida, documenti e articoli...), lingua, argomento, data di pubblicazione, nazionalità e livello (globale, regionale, nazionale, locale, istituzionale) oppure tramite un'area di testo che cerca una corrispondenza nel titolo e nei tag.

Le singole risorse che compongono la lista dei risultati derivanti dall'impostazione dei filtri di ricerca hanno tutte la stessa struttura e forniscono le stesse informazioni nonostante possano appartenere a tipologie diverse. Le informazioni che l'utente visualizza per ogni risorsa tra quelle in elenco sono: la tipologia di risorsa associata a un'icona esplicativa, il titolo, le informazioni bibliografiche, un breve abstract e una lista di eventuali tag. Cliccando poi sulla singola risorsa di interesse, questa è corredata di una descrizione solitamente più dettagliata rispetto all'abstract e di un pulsante di *download* che consente di scaricare la risorsa stessa.

La piattaforma, oltre a fornire materiali liberamente consultabili e scaricabili, è adatta, come precedentemente anticipato, anche ad ospitare materiale sottomesso da terze parti. Chiunque voglia contribuire con testimonianze di esperienze può farlo iscrivendosi con il ruolo di autore e accedendo così al *back end* della piattaforma che è adatto a gestire l'iter di sottomissione e revisione delle proposte inviate dagli utenti. All'interno del portale si possono sottomettere incidenti e buone pratiche, entrambe con una struttura piuttosto complessa e un numero di campi elevato. Le risorse, invece, non sono soggette alla procedura di sottomissione, ma sono completamente gestite dagli editori.

### 3 METODI

Per la realizzazione del progetto è stato deciso di utilizzare WordPress<sup>1</sup>, un CMS (Content Management System) a oggi fra i più potenti e utilizzati, con un'attiva community alle spalle. Questo software gratuito e open source consente la creazione, l'organizzazione e la gestione dei contenuti tramite una semplice e pratica interfaccia di *back end*. Il risultato del lavoro svolto nel *back end*, prende forma nel *front end* laddove i contenuti sono fruibili come un tradizionale sito web.

Inoltre, WordPress consente una profilazione che permette l'organizzazione gerarchica dei permessi associati ai ruoli. WordPress prevede sei ruoli standard (super amministratore, amministratore, editore, autore, contributore, utente registrato) a cui possono essere eventualmente aggiunti nuovi profili. Nel caso di GKPS la profilazione personalizzata è di fondamentale importanza per garantire il corretto funzionamento del processo di invio e revisione delle esperienze proposte da utenti esterni, diversi dai gestori della piattaforma. Ciò si realizza grazie ad un plugin, *Advanced Access Manager*, che permette di modificare, limitare o aumentare i poteri consentiti a un ruolo standard di WordPress o addirittura creare ruoli, sia *ex novo*, sia ereditando i diritti di un altro ruolo e procedendo per successive modifiche<sup>2</sup>.

---

1 <https://it.wordpress.org/>

2 <https://it.wordpress.org/plugins/advanced-access-manager/>

Tre sono i ruoli più importanti creati tramite questo plugin: l'autore, il revisore e l'editore. L'autore si identifica nell'utente che vuole sottomettere la sua esperienza al giudizio dei revisori, i quali si occupano di verificare la correttezza dei dati, mentre l'editore si occupa della loro pubblicazione.

Il primo passo da affrontare per permettere all'autore l'invio di contenuti è quello di offrire un modello per l'inserimento dei dati richiesti. Sono state prese in considerazione varie ipotesi sia da *back end* che da *front end*.

Il vantaggio di utilizzare il *back end* risiede soprattutto nel fatto che ad ogni utente viene associata una propria area, quindi un utente che avesse sottoposto cinque articoli, potrebbe seguire nell'ambiente a lui dedicato (*dashboard*) l'andamento di tutti e cinque i suoi lavori. La soluzione da *front end* può sembrare più immediata, ma in genere offre strumenti di gestione più limitati. Si nota, inoltre, che parte delle funzionalità che si possono offrire lato *back end* sono native di WordPress e quindi non richiedono uno specifico sviluppo.

Dopo aver valutato tutte le ipotesi si è scelto di sviluppare una soluzione che avesse come obiettivo quello di rendere intuitivi e immediatamente rintracciabili i campi dove inserire le informazioni e che fosse capace di strutturare in modo automatico queste informazioni associando loro un *template* per la corretta visualizzazione dei contenuti lato *front end*. Diversi sono ovviamente i campi custom relativi agli incidenti e alle pratiche.

Tra i plugin indispensabili per lo sviluppo di GKPS figurano poi: Custom Post Type UI<sup>3</sup> e Advanced Custom Fields<sup>4</sup>. Il primo consente di creare dei post custom che per qualche caratteristica hanno bisogno di essere differenziati dai post tradizionali, mentre il secondo permette di creare gruppi di campi personalizzati, al cui interno si trovano diversi tipi di campi singoli (*text*, *textarea*, *number*, *date*, *checkbox*, *radio*, ecc...), i quali possono essere associati a uno o più tipi di post, standard o custom.

Quindi per ogni tipo di esperienza (incidente o pratica), sono stati creati tre tipi di post personalizzati con il plugin *Custom Post Type UI*: uno ("Incident Subm" e "Practice Subm") che serve all'autore per sottomettere la propria esperienza, uno ("Incident Rev" e "Practice Rev") serve invece al revisore per controllare la proposta dell'autore, e uno ("Incident Pub" e "Practice Pub") all'editore che provvede a pubblicare l'articolo dopo che ha ricevuto l'approvazione da parte del revisore. I post di sottomissione, revisione e pubblicazione possiedono la stessa struttura, anche se a questi ultimi, a differenza degli altri due, è associato il *template* di formattazione.

Nonostante le numerose estendibilità di WordPress, che concernono sia gli aspetti grafici che le funzionalità, qualche volta i temi e i plugin non sono sufficienti a sopperire alle numerose e necessarie personalizzazioni, tantoché in questi casi è necessaria la scrittura e l'integrazione di codice *ad hoc*. È questo il caso del *template* associato agli articoli di incidenti e buone pratiche, realizzato tramite HTML e progettato graficamente in CSS, ma anche dei cosiddetti *breadcrumbs*, i "percorsi a briciole", dei link di navigazione che riproducono il percorso disegnato dal sito web e consentono all'utente di in quale sezione del portale si trova<sup>5</sup>.

---

3 <https://it.wordpress.org/plugins/custom-post-type-ui/>

4 <https://www.advancedcustomfields.com/>

5 <https://www.serverplan.com/blog/breadcrumb/#gref>

Sono state inoltre realizzate alcune specifiche funzionalità atte a migliorare o a correggere il comportamento del tema grafico scelto. Questi sviluppi sono stati funzionali a migliorare l'esperienza del fruitore. Sempre in questa ottica si è scelto di integrare e personalizzare una libreria Javascript, non un plugin di WordPress, per poter realizzare una mappa interattiva, utile nella ricerca dei contenuti, che tenesse conto della particolare tassonomia adottata dall'OMS che, ad esempio, non suddivide il mondo nei cinque noti continenti, bensì in macroaree differenti che un comune utente difficilmente potrebbe immaginare<sup>6</sup>. Anche l'implementazione delle funzioni di ricerca delle diverse tipologie di materiale condiviso ha richiesto molteplici personalizzazioni secondo specifiche richieste dei committenti che non potevano essere effettuate attraverso gli strumenti standard messi a disposizione da WordPress.

## 4 RISULTATI

Il punto di forza di GKPS è sicuramente la capacità di spaziare tra i contenuti, offrendo informazioni che vanno dalle tematiche più tradizionali del rischio clinico fino a quelle più attuali, come la pandemia del COVID-19. Talmente ampia è la sua risonanza mondiale che si ritaglia un'area a sé stante all'interno della piattaforma, risultando più funzionale e intuitivo per gli utenti, ma anche continuando a rispettare gli obiettivi originali del portale.

Attualmente la piattaforma è al vaglio dell'OMS che, particolarmente oberato dalla situazione internazionale dovuta all'emergenza sanitaria, sta verificando con rigore scientifico le diverse sezioni del sito da pubblicare e i contenuti che le popolano.

Finora la piattaforma ha le caratteristiche per la quale è stata pensata, ha avuto modo di essere testata, anche se non stressata a causa di contenuti abbastanza esigui, durante una live-session del convegno "Second Expert Meeting for the collaborative design of the Global Knowledge sharing platform for Patient Safety (GKPS)" con risultati soddisfacenti.

Il risultato è una piattaforma adatta a gestire diverse tipologie di contenuti attraverso un articolato, ma preciso processo di sottomissione delle esperienze. All'utente che vuole partecipare attivamente con l'invio di esperienze relative al rischio clinico verrà assegnato il ruolo di autore, il quale dovrà attenersi alle linee guida indicate nel portale. Un utente particolarmente esperto può candidarsi al ruolo di *reviewer*. Nel caso venisse accolto come revisore, dovrà anch'egli seguire le norme di comportamento messe a disposizione dal portale. Ogni articolo viene affidato dagli editori a uno o più revisori che hanno il compito di acconsentirne o negarne la pubblicazione. Nel caso in cui questo non venga immediatamente approvato, può essere oggetto di confronto tra l'autore e il revisore. In questo caso il revisore segnala le parti a suo giudizio incomplete o inesatte che verranno modificate, ampliate o corrette. L'autore, preso atto dei commenti del revisore, potrà correggerlo e inviarlo di nuovo. Qualora il documento risultasse idoneo, il revisore informerà l'autore e l'editore, il quale dichiarerà l'articolo pubblicabile.

---

6 [https://it.wikipedia.org/wiki/ISO\\_3166-1\\_alpha-2](https://it.wikipedia.org/wiki/ISO_3166-1_alpha-2)

## **BIBLIOGRAFIA**

1. Centro di Gestione del Rischio Clinico (Regione Toscana). *Relazione di attività del Centro Gestione Rischio Clinico*. Firenze: Regione Toscana, 2020. <https://www.regione.toscana.it/documents/10180/11888593/Relazione+annuale+GRC+18-19.pdf/e9f7701f-6366-7c4b-09de-8aa-de07525d4?t=1597056880584>.

# Di cosa parliamo quando parliamo di FAIR?

Francesca Frontini<sup>1</sup>, Anas Fahad Khan<sup>2</sup>

<sup>1</sup> ILC-CNR – francesca.frontini [at] ilc.cnr.it

<sup>2</sup> ILC-CNR – fahad.khan [at] ilc.cnr.it

## SINTESI

Nel 2016 un consorzio di scienziati afferenti a diverse istituzioni e discipline enuncia i principi FAIR; in questi quattro anni l'importanza e la portata del programma FAIR è divenuta sempre più evidente. L'adesione a tali principi nelle discipline umanistiche sembra farsi largo, ma non senza difficoltà e interrogativi. Questo lavoro propone una riflessione sulle implicazioni della proposta FAIR per la gestione dei dati scientifici, confrontandola con la sua effettiva ricezione nella comunità delle DH in Italia e in Europa.

## PAROLE CHIAVE

principi FAIR, open data, dati della ricerca, politiche della ricerca, EOSC

## INTERVENTO

### 1 INTRODUZIONE

Nel marzo 2016, un consorzio di scienziati afferenti a diverse istituzioni e discipline pubblica “FAIR Guiding Principles for scientific data management and stewardship” [11], una serie di principi e linee guida identificati per la gestione dei dati scientifici.

Quello che, in un clima sempre più favorevole agli *open data* e alla condivisione dei risultati scientifici, poteva a prima vista sembrare l'ennesimo richiamo a già note buone pratiche di data curation, si rivela ben presto come un programma più concreto e ambizioso. L'acronimo è accattivante e se vogliamo anche rassicurante - FAIR sta per *findable* (trovabile), *accessible* (accessibile), *interoperable* (interoperabile), *reusable* (riutilizzabile), e l'assonanza con la parola inglese *fair*, richiama evidentemente le categorie semantiche di equità, giustizia, pari condizioni (si pensi ad espressioni come *fair play*, o *fair trade*) [1]. Tuttavia, i principi FAIR<sup>1</sup> enunciati nel manifesto stilano una lista di requisiti di stampo tecnico ben precisi e stringenti; l'iniziativa appare poi da subito dotata di un notevole sostegno istituzionale.

---

1 [www.force11.org/group/fairgroup/fairprinciples](http://www.force11.org/group/fairgroup/fairprinciples)



Le FAIR *data policies* sono infatti ormai diventate anche una parte importante dell'iniziativa European Open Science Cloud (EOSC) della Commissione Europea, che mira a sviluppare un'infrastruttura di servizi al fine di promuovere la scienza aperta. Ad esempio, il progetto FAIRsFAIR<sup>2</sup>, finanziato dall'UE, ha l'obiettivo di sviluppare un'infrastruttura della conoscenza e implementare soluzioni pratiche per assistere gli scienziati (e qui possiamo intendere il termine "scenziato" in un senso estremamente ampio, che comprende i ricercatori che lavorano nelle scienze sociali e umanistiche) nel rendere i loro dati più FAIR. Vi è poi EOSCFAIR4, uno dei gruppi di lavoro EOSC, e parte integrante della sua struttura di governance, che è dedicato a "implementare i principi dei FAIR data" e ha l'obiettivo esplicito di promuovere "l'interoperabilità interdisciplinare". Già oggi, nei *data management plan* dei progetti europei e nazionali (in alcuni paesi) è di fatto necessario far riferimento alla FAIRizzazione dei risultati della propria ricerca.

Tutto questo ci fa capire che i principi FAIR sono destinati a diventare imprescindibili per il ciclo di vita dei dati scientifici, anche per chi lavora nel settore delle scienze umane. Tuttavia, in questo ambito l'adesione ai principi FAIR sembra più lenta. Una possibile motivazione si trova nel recente documento pubblicato dal gruppo di lavoro EOSCFAIR dal titolo "Six Recommendations for Implementation of FAIR Practice" [4]<sup>3</sup> che rileva gli atteggiamenti delle varie comunità disciplinari; i risultati mostrano che molti ricercatori vedono i principi FAIR come adatti solo ai dati quantitativi e digitali<sup>4</sup>. Questo è forse anche dovuto al fatto che, almeno all'inizio, la comprensione delle implicazioni pratiche e tecnologiche dietro ai principi astratti è stata parziale e limitata ad una lettura superficiale e generica dei principi espressi.

Nonostante questo ritardo, osserviamo come oggi un certo numero di progetti e pubblicazioni in ambito DH facciano riferimento ai principi FAIR, e pensiamo che sia utile cercare di studiare la ricezione di questo nuovo paradigma all'interno della nostra comunità.

Nella nostra presentazione, delinearemo prima di tutto in breve le idee chiave dei principi FAIR, partendo da un esempio concreto, quello proposto da Barend Mons e riguardante i dati biomedici, mostrando prima di tutto la differenza tra i principi FAIR nella loro formulazione attuale e, ad esempio, pratiche più generiche di *Open Data*. Ripercorreremo in seguito la ricezione dei principi FAIR analizzando alcune pubblicazioni dal 2016 in poi, in Italia e all'estero. Trarremo infine alcune conclusioni preliminari al fine di aprire una discussione sulle conseguenze dei cambiamenti in atto per la nostra comunità scientifica.

---

2 [www.fairsfair.eu](http://www.fairsfair.eu)

3 <https://docs.google.com/document/d/1VKbI29Dc5jHhkVGZkUTD-L8IUeqJllvsSJiNeurjc6k/>

4 FAIR "was made for quantitative data (e.g., social sciences and humanities) or that it is not suitable for the study of real world objects because that is different from the study of digitised objects (e.g., humanities but much less in natural history collections)".

## 2 FAIR DATA & SERVICES, FAIR DIGITAL OBJECTS, KNOWLETS – COSA VUOL DIRE FAIR NELLE SCIENZE BIOMEDICHE

Non passeremo qui in rassegna i principi uno per uno, ma già ad una prima lettura si nota il riferimento a protocolli e standard informatici di stoccaggio, accesso e metadatozione dei dati, nonché alle licenze di utilizzo. Inoltre, e soprattutto, il riferimento alle macchine, oltre che agli umani, come utenti finali di tali dati.

Il fatto che non si tratti semplicemente dell'esposizione di idee astratte risulta visibile anche nel sostegno che da subito l'iniziativa ha ricevuto a livello istituzionale e scientifico. Ne è la prova la creazione del consorzio internazionale GoFAIR<sup>5</sup>, nel cui *steering committee* siedono rappresentanti dei governi francese, olandese e tedesco, e che si prefigge di implementare concretamente i principi FAIR nei vari settori disciplinari. Ed è infatti in Europa che la promozione dei principi FAIR sta ricevendo il grande impulso, anche in connessione al già citato EOSC.

Ma qual è l'obiettivo ultimo del programma FAIR? A prima vista esso sembra avvicinarsi a idee propugate dai fautori dell'*Open Data* - che punta a rendere disponibili con licenze aperte i dati della ricerca, ma non è esattamente così. FAIR non vuol dire esattamente *open*; i dati FAIR possono anche essere dati chiusi, o aperti solo per alcuni. Alla base del programma FAIR vi è la possibilità di interrogare in maniera standardizzata e combinare grandi set di dati di provenienza diversa, con l'obiettivo di imprimere un'accelerazione alle scoperte scientifiche in vari ambiti. Il paradigma FAIR ha dunque una forte attinenza ai principi del Web Semantico e dei *Linked Data* (non necessariamente Open), in quanto raccomanda la descrizione di metadati e dati secondo modelli formali e vocabolari condivisi.

Forse l'ambito in cui questo programma trova la sua applicazione più chiara è quello delle scienze biomediche. Uno dei più noti sostenitori dei principi FAIR è Barend Mons<sup>6</sup> professore al Leiden University Medical Center, nei Paesi Bassi. Nella sua recente presentazione a EOSS 2020 dal titolo "The FAIR Guiding Principles in Times of Crisis"<sup>7</sup> egli illustra il suo programma per un riutilizzo più efficace dei dati scientifici nel suo settore, ed in particolare per l'uso dei big data per identificare i trattamenti più efficaci per curare i sintomi del COVID-19. Considerando la mole di pubblicazioni sull'argomento e il fatto che una ricerca ha bisogno di essere confermata indipendentemente più e più volte prima di essere considerata un dato scientifico assodato, è facile capire come in questo settore al momento la semplice lettura di articoli scientifici da parte di umani non sia più sufficiente. Secondo il prof. Mons la visione FAIR dovrebbe permettere di costruire un vero e proprio *Internet of FAIR Data & Services*<sup>8</sup>, composto di *FAIR Digital Objects*<sup>9</sup>.

---

5 <https://www.go-fair.org/>

6 [https://en.wikipedia.org/wiki/Barend\\_Mons](https://en.wikipedia.org/wiki/Barend_Mons)

7 <https://eross2020.inf.unibz.it/programme/>

8 <https://www.go-fair.org/resources/internet-fair-data-services/>

9 <https://fairdo.org/>

In particolare, i risultati della ricerca dovrebbero essere pubblicati non solo sotto forma testuale, ma anche sotto forma di grafi di conoscenza, in cui ogni informazione, ad esempio la proprietà di un dato farmaco, o i suoi effetti secondari, possano essere trasformati in *knowlets*, unità di conoscenza strutturata, standardizzata, FAIR.

Esplorando un enorme grafo di conoscenza composto dagli studi sull'uso dell'idrossiclorochina per la cura delle complicazioni del COVID-19, Mons mostra come sia possibile “vedere” che, nel complesso, gli articoli che riscontrano effetti positivi sono pochi.

Ascoltando la presentazione ci si rende conto che, almeno nelle scienze della vita, i principi FAIR non sembrano rappresentare solo buone pratiche di gestione dei dati, ma un programma ambizioso che mira a modificare ed in qualche modo “automatizzare” il modo di fare ricerca. Va comunque detto che anche su questa visione a lungo termine i sostenitori di FAIR non sono tutti d'accordo; nel già citato “Six Recommendations for Implementation of FAIR Practice” si sottolinea che il FAIR per le macchine, per quanto importante, rappresenta un obiettivo molto difficile da raggiungere.

### **3 RICEZIONE DEI PRINCIPI FAIR NELLE SCIENZE UMANE – UN’ANALISI DI CORPUS**

A nostro avviso la portata del programma FAIR è ben lungi dall'essere stata pienamente compresa e interiorizzata nelle scienze umane e sociali, dove prevale una visione dei principi FAIR forse ancora un po' “ingenua”, che in buona sostanza si riduce alla adesione a buone pratiche di *data curation* già note ad alcuni, ma ora *à la page* e riassunte da uno slogan accattivante. Al fine di verificare questa intuizione, ci serviremo di una analisi di articoli ed abstract sottomessi alle conferenze AIUCD e DH (ADHO) a partire dal 2016<sup>10</sup>, anno della pubblicazione dei principi FAIR. I risultati preliminari (si veda la Tabella 1) mostrano la penetrazione dei principi FAIR nella comunità DH (almeno europea) a partire dal 2017. In quell'anno un ruolo importante è stato svolto dal progetto PARTHENOS, che ha contribuito a calare i principi nel contesto delle humanities<sup>11</sup>. Come settore trainante emergono le scienze del patrimonio (*Digital Cultural Heritage*); in quest'ultimo ambito l'esistenza di ontologie formali in RDF come CIDOC-CRM ha probabilmente facilitato l'adesione a vocabolari condivisi, indispensabili ad una vera FAIRizzazione dei (meta)dati. Tuttavia, negli altri ambiti il riferimento ai *Linked Data* in relazione FAIR è raro e per molti progetti essi non costituiscono un principio ispiratore. Il riferimento a principi FAIR (magari solo a un principio) appare in alcuni casi chiosa *ex post*, che serve a giustificare delle scelte fatte a monte. Un articolo può ad esempio dichiarare l'adesione ai principi di riusabilità e interoperabilità sottolineando la scelta di uno standard di codifica del testo o una licenza aperta, tralasciando però quelli di trovabilità e accessibilità che richiederebbero il deposito del dataset in un centro dati riconosciuto (“data is deposited in a trusted repository”) che permette l'*harvesting* di metadati.

---

10 Questo lavoro preliminare è stato condotto sugli atti delle conferenze, disponibili online, tuttavia intendiamo approfondire la ricerca utilizzando l'Index of Digital Humanities Conferences, reso disponibile di recente [7].

11 Si veda ad esempio il ‘PARTHENOS – Joint PARTHENOS-CLARIN-DARIAH Workshop on Applying the FAIR PrinciplesPARTHENOS’, svoltosi nel 2017, oltre che i riferimenti in bibliografia.  
<http://www.parthenos-project.eu/joint-parthenos-clarin-dariah-workshop-on-applying-the-fair-principles/>

In DH sono quindi ancora rari i lavori nei quali un dataset viene presentato argomentando punto per punto l'adesione ai principi FAIR sulla base delle scelte tecnologiche. Più frequenti invece gli articoli, seminari o panel dove la discussione intorno a FAIR avviene nel contesto di grandi infrastrutture di ricerca o progetti infrastrutturali [ad esempio 5; 10], cosa che sottolinea come l'implementazione di tali principi non possa essere demandata ai singoli progetti, ma necessiti di interventi strutturali e strumenti ben precisi.

| Menzioni dei principi FAIR                 | 2016 | 2017 | 2018   | 2019  | 2020   |
|--|------|------|--|---|--|
| AIUCD articoli ed abstract (inclusi panel) | -    | -    | 2 (terminologia; open data e infrastrutture) | 2 (cultural heritage; Linguistic Linked Data) | 3 (principi FAIR per DH; edizioni digitali; geolinguistic) |
| DH abstract (paper, panel, seminari)       | -    | 1    | 1  | 11  | 4  |

Tabella 1 – Paper e abstract delle conferenze AIUCD e DH che menzionano i principi FAIR, con indicazioni delle tematiche.

## 4 CONCLUSIONI PRELIMINARI

Nel problematizzare la ricezione dei principi FAIR da parte della comunità DH, non intendiamo proporre un'evangelizzazione acritica dei riluttanti umanisti al programma FAIR, ma piuttosto mettere in evidenza anche le contraddizioni e le difficoltà che da più parti cominciano ad essere espresse in particolare nel campo delle DH. L'implementazione dei requisiti FAIR, con tutta la loro complessità tecnica, richiede investimenti importanti, che tendono a favorire i grandi centri di ricerca dei paesi più ricchi, dotati di infrastrutture e finanziamenti ad hoc. In questo senso FAIR non vuole sempre dire *fair*, come ha sottolineato Antonio Rojas Castro nella sua recente presentazione proprio sulla difficoltà di applicarli da in progetti e dataset provenienti dal *Global South* [3].

Ci sono comunque ragioni per essere ottimisti. L'adozione del paradigma dei *Linked (Open) Data* in una serie di discipline di SSH mostra che ci sono aree in cui una forte convergenza verso linguaggi, modelli e vocabolari standard aiuta a supportare il tipo di interoperabilità sintattica e semantica - all'interno e tra discipline - che è alla base dei principi FAIR. Questo è vero in campi come la linguistica e aree correlate come le risorse linguistiche e la lessicografia, ma anche negli studi medievali, per i dati geospaziali e quelli del patrimonio culturale ecc. A questo si aggiunge l'uso ormai diffuso delle linee guida TEI che aiutano a codificare diversi tipi di risorse utilizzando un modello unificato basato su XML. Come sottolinea Mons, i lavori pionieristici sulla standardizzazione dei metadati delle risorse linguistiche sono stati all'avanguardia anche rispetto alle scienze naturali. Questo mostra che esiste una solida base di culture e pratiche su cui FAIR può essere implementato nelle SSH.

## 5 RINGRAZIAMENTI

Questo abstract è ispirato dai lavori svolti dagli autori nel contesto del progetto MSH-SUD Montpellier CommonData, e dei progetti H2020 SSHOC (grant agreement 823782) e TRIPLE (grant agreement 863420).

Ringraziamo i nostri colleghi di ILC-CNR per le stimolanti discussioni sulle tematiche trattate.

## BIBLIOGRAFIA

1. Amiel, Philippe, Francesca Frontini, Pierre-Yves Lacour, e Agnès Robin. «Pratiques de gestion des données de la recherche: une nécessaire acculturation des chercheurs aux enjeux de la science ouverte?. Résultats d'une enquête exploratoire dans le bassin montpelliérain (juin 2018)». *Cahiers Droit, Sciences & Technologies*, n. 10 (2020): 147–168. <https://doi.org/https://doi.org/10.4000/cdst.2061>.
2. Barbuti, Nicola. «Ripensare i dati come risorse digitali: un processo difficile?» *L'INFORMATICA UMANISTICA*, s.d., 19.
3. Castro, Antonio Rojas. «FAIR enough? Building DH Resources in an Unequal World», 2020.
4. Deniz Beyan, Oya, Chue Hong, Neil, Cozzini, Stefano, Hoffman-Sommer, Marta, Hooft, Rob, Lembinen, Liisi, Marttila, Juuso, e Teperek, Marta. «Seven Recommendations for Implementation of FAIR Practice». Zenodo, 23 giugno 2020. <https://doi.org/10.5281/ZENODO.3904140>.
5. Dumouchel, Suzanne, e Elena Giglia. «CO-OPERAS IN: Integration And Cooperation To Face Fragmentation And Address Complexity In The SSH». DataverseNL, 2019. <https://doi.org/10.34894/AJF13Q>.
6. Eichmann-Kalwara, Nickoal, Jeana Jorgensen, e Scott B. Weingart. *Representation at Digital Humanities Conferences (2000-2015)*. JSTOR, 2018.
7. Eichmann-Kalwara, Nickoal, Weingart Scott B., e Lincoln Matthew. «The Index of Digital Humanities Conferences.» Carnegie Mellon University, 2020. <https://doi.org/https://doi.org/10.34666/k1de-j489>.
8. PARTHENOS, Hollander, Hella, Morselli, Francesca, Uiterwaal, Frank, Admiraal, Femmy, Trippel, Thorsten, e Di Giorgio, Sara. «PARTHENOS Linee guida per l'applicazione dei principi FAIR alla gestione e al riuso dei dati», 2 settembre 2019. <https://doi.org/10.5281/ZENODO.3363243>.
9. Sbarbati, Stefano. «The FAIR Principles and the EOSC Concept in PARTHENOS Community – PARTHENOS Project», 2020. <http://www.parthenos-project.eu/the-fair-principles-and-the-eosc-concept-in-parthenos-community>.
10. Wessel, Leon. «CLARIN for DH Scholars», s.d.
11. Wilkinson, Mark D., Dumontier Michel, Jan Aalbersberg IJsbrand, Gabrielle Appleton, Myles Axton, Baak Arie, e Niklas Blomberg. «The FAIR Guiding Principles for Scientific Data Management and Stewardship». *Scientific Data* 3, 2016. <https://doi.org/10.1038/sdata.2016.18>.

# L'Atlante delle stragi. Da Digital History a Digital Public History

Igor Pizzirusso  
Istituto nazionale Ferruccio Parri – pizzirussoigor [at] gmail.com

## SINTESI

Nato raccogliendo le ricerche locali attivate in alcuni territori a partire dal 1999, l'Atlante delle stragi naziste e fasciste ha avuto una gestazione lunga, che ha poi avuto un'accelerazione grazie al finanziamento del Fondo italo-tedesco per il futuro. Online dal 6 aprile 2016, la banca dati è stata fin da subito un importante strumento per i ricercatori, ma anche un fenomenale memoriale pubblico, aperto e per questo capace di intercettare le frequenti sollecitazioni provenienti da intere comunità locali, da associazioni territoriali o da familiari delle vittime, che tutt'oggi contribuiscono all'integrazione e al perfezionamento del corpus dei dati e quindi a migliorare la qualità della ricerca e dello strumento stesso.

## PAROLE CHIAVE

stragi naziste e fasciste, digital history, digital public history, public history, resistenza, nazismo, fascismo, seconda guerra mondiale

## INTERVENTO

### 1 INTRODUZIONE

On line dal 6 aprile 2016, l'Atlante delle stragi naziste e fasciste<sup>1</sup> ha avuto una gestazione molto lunga, che affonda le radici vent'anni prima della data di pubblicazione.

Frutto di una ricerca ampia e articolata che ha coinvolto la rete degli istituti della Resistenza, l'ANPI e circa 130 studiosi sparsi in tutto il territorio nazionale, l'Atlante è divenuto presto uno strumento "pubblico", non solo per la capillarità di diffusione che gli ha consentito il web. Memorie locali e famigliari per molto tempo sopite o rimaste in circoli ristretti hanno finalmente trovato il luogo dove poter trovare visibilità, inserite oltretutto in un quadro d'analisi nazionale.

---

1 <http://www.straginazifasciste.it/>

## 2 NASCITA DEL PROGETTO

Il tema dei massacri compiuti dai tedeschi in Italia nel periodo 1943-1945 torna prepotentemente alla ribalta per due eventi “pubblici”, di natura più giornalistica che storica: la scoperta del cosiddetto “armadio della vergogna” (1994) e il processo a Erich Priebke (1995) per la strage delle Fosse Ardeatine. Nel frattempo si erano comunque attivate diverse ricerche sul territorio, ma è solo a partire dal 1999 che si sviluppano indagini storiche approfondite su determinati scenari regionali: la Campania, che subisce la prima ondata di violenze dopo l’armistizio e fino alla ritirata tedesca sulla linea Gustav; la Toscana e l’Emilia-Romagna, che invece sperimentano la tremenda azione repressiva nazista in particolare durante la stasi del fronte di guerra sulla Linea Gotica.

Quando prende avvio la ricerca nazionale, nel 2014<sup>2</sup>, queste ricerche sono ovviamente le prime a confluire nel progetto dell’Atlante e a costruire la prima ossatura del database.

## 3 LA DIGITAL HISTORY DIVENTA PUBLIC HISTORY

La scelta di costruire una banca dati e non un volume cartaceo non è immediata, ma ci si accorge presto di quanto sia essenziale. Il valore aggiunto in termini di data analysis è inestimabile, ma non va sottovalutata nemmeno la facilità e l’immediatezza di intervento e modifica che è tipica degli strumenti digitali, assolutamente essenziale quando si tratta di divulgare una materia così fluida e incerta come lo sono le stragi naziste (e fasciste) in Italia.

Materia fluida e incerta perché anche molto vissuta e partecipata a livello locale, dove da sempre si agitano ricerche autonome di giovani storici o appassionati, comunità che hanno vissuto e rielaborato la tragedia, associazioni di famigliari o di vittime. Sono infatti numerose le segnalazioni pervenute nel tempo (e tutt’oggi) all’Istituto nazionale. Si tratta per lo più di aggiunte o modifiche spesso (ma non sempre) pertinenti, che se da un lato non alterano il quadro complessivo nazionale (né i quadri regionali o provinciali) dall’altro costituiscono tasselli fondamentali per emendare errori o inesattezze e restituire così un racconto (e una memoria) quanto più possibile completa e fedele dei fatti e delle vicende, singole, famigliari o di intere comunità.

Tutto ciò rende l’Atlante un chiaro esempio di public history, intesa nel senso più forte della definizione. La banca dati, in quanto prodotto on line, ha una diffusione capillare, almeno potenzialmente; rende insomma accessibile la storia a un pubblico più vasto, teoricamente illimitato. Non solo. L’Atlante consente al pubblico stesso di diventare protagonista del processo di produzione del contenuto storico, naturalmente con la costante e sapiente mediazione degli storici dell’Istituto nazionale Ferruccio Parri. Una produzione che potremmo definire “sociale” in una duplice accezione: da un punto di vista pratico, se la intendiamo come “collaborazione”; da un punto di vista etico, se la intendiamo come coinvolgimento delle comunità e come valorizzazione di istanze provenienti “dal basso”.

È per altro curioso ragionare sul fatto che il progetto non sia nato con questo specifico intento. Si trattava infatti di una ricerca di tipo “classico” che si è deciso di trasferire in digitale soprattutto per pragmatismo. L’approdo al web ha però trasformato un prodotto potenzialmente per specialisti in qualcosa di diverso e “per tutti”.

---

2 [http://www.straginazifasciste.it/?page\\_id=9](http://www.straginazifasciste.it/?page_id=9)

Naturalmente l'Atlante rimane un raffinatissimo strumento di ricerca storica; ma è anche un memoriale virtuale delle vittime delle violenze naziste e fasciste del periodo 1943-1945, un luogo dove intere comunità trovano rappresentanza e le singole storie si mettono in relazione con altre singole storie per comporre il quadro della storia con S maiuscola.

Nel processo ha certamente aiutato un tema (le stragi) che è in strettissimo contatto con le memorie locali e famigliari. Un elemento che deve far riflettere sul perché non tutti i prodotti digitali siano in grado di innescare meccanismi di questo genere. Sui motivi, insomma, per cui non tutto il digital è automaticamente public. Ma questa è un'altra storia...

## **4 LA STORIA CON S MAIUSCOLA: I RISULTATI DELLA RICERCA**

La storia dell'Atlante è invece una storia che parla soprattutto di risultati, quantitativi e qualitativi. Il primo conseguito dalla banca dati è stato quello di confermare molte delle argomentazioni e dei fatti noti legati agli scenari più indagati (Emilia-Romagna, Campania e Toscana). E tuttavia, anche in questi quadri già ampiamente studiati, l'Atlante funge da importante pungolo alla correzione o all'integrazione. Soprattutto però, esso ha avuto un ruolo di congiunzione e sintesi, riuscendo a unire i suddetti quadri locali in un quadro d'insieme su scala nazionale, rilevando perciò le continuità, i tratti comuni, le concomitanze nell'occupazione nazista e fascista, ma anche e soprattutto le determinanti diversità; e ha dato inoltre l'impulso per mettere in discussione anche teorie consolidate, come quella sulle stragi "eliminazioniste" dell'Emilia e della Toscana, che si sono scoperte anch'esse legate a precise logiche belliche più che di sterminio indiscriminato.

L'Atlante ha permesso quindi di elaborare nuove tesi, delineando scenari più completi e razionali della violenza nazista e – come proprio il database ha permesso di evidenziare – fascista. Il dato che emerge dalla ricerca è esplicito: il 20% del totale degli episodi è di esclusiva matrice italiana, senza alcuna partecipazione tedesca (e si arriva a oltre il 50% in alcune regioni del Nord, come la Lombardia).

## **5 L'INDAGINE E I DATI**

Naturalmente risultati tanto efficaci sono stati possibili soltanto grazie a una concettualizzazione preventiva della ricerca, che ha individuato con estrema precisione quali fossero i dati da ricercare per ogni singolo episodio e in quale modo fosse più giusto e utile categorizzarli ed etichettarli.

Si è cercato perciò di distinguere le stragi in base alla matrice (nazista, fascista o nazifascista), alle violenze perpetrate, ai perpetratori (persone e reparti militari), alla tipologia delle vittime, alle modalità di uccisione e alle violenze connesse, con la convinzione che l'insieme di queste informazioni avrebbe permesso di rimarcare quali fossero gli obiettivi principali, le modalità, le finalità della politica di violenza tedesca e italiana nel periodo 1943-1945. Conseguenza di ciò sono state alcune precise scelte compiute fin dalle primissime fasi di progettazione.



Sotto questo aspetto, due delle decisioni più innovative, e per questo interessanti, sono state:

- 1 Includere tra le vittime anche i partigiani inermi. Ovviamente non sono stati presi in considerazione coloro che sono morti in combattimento, bensì solo coloro che vennero uccisi dopo la cattura o il disarmo. Anche queste uccisioni, spesso simboliche, appartenevano alla politica di violenza nazifascista.
- 2 Includere anche le uccisioni singole, perché spesso lo stillicidio di “episodi minori” accompagna una fase cronologica o geografica della violenza o tratteggia la prassi di alcuni reparti. Ogni singolo omicidio rappresenta un tassello di un quadro più grande e solo ricomponendo e raggruppando anche questi casi isolati si ottiene un quadro generale esaustivo.

L'aspetto più interpretativo, e pertanto più complesso, è stato ed è indubbiamente quello legato alle tipologie, che hanno a che fare con gli scopi della violenza: rappresaglia e rastrellamento, ad esempio, sono certamente delle modalità con cui si perpetra una strage, ma anche delle finalità.

Fare chiarezza su questo aspetto ha per altro permesso di sgombrare il campo anche da equivoci persistenti nella vulgata comune, che identificava ad esempio come rappresaglia cioè che invece aveva caratteristiche e obiettivi diversi.

## **6 CONSULTARE L'ATLANTE**

L'Atlante si può consultare un quattro modi distinti:

1. Usando la mappa, che fornisce la georeferenziazione di ogni episodio e consente la navigabilità sul territorio. E' il metodo migliore per usare l'Atlante quando ci si trova “in loco”, quasi fosse una app.
2. Ricerca per località o toponimo, estremamente precisa e quindi dedicata a chi sa già quale luogo o quale strage cercare.
3. Ricerca semplice per luogo e data, utilissima per tratteggiare quadri di insieme territoriali nel breve o lungo periodo.
4. Ricerca avanzata, che unisce i vari filtri di ricerca (le categorizzazioni di cui sopra) e consente una profondità di analisi dedicata soprattutto a storici e ricercatori.

Ci sono poi una quinta e una sesta modalità, legate rispettivamente ai materiali iconografici (tuttora in corso di implementazione) e ai singoli reparti autori delle stragi, per i quali è possibile ricostruire sulla mappa la scia di sangue.

## **7 CONCLUSIONE**

L'Atlante delle stragi naziste e fasciste costituisce dunque un esempio eccellente di public history a livello nazionale, forse anche per la sua dimensione assolutamente globale: si può indagare il particolare (il comune o la specifica zona sulla mappa) o le macroaree (regionali o provinciali). Si dà quindi grande importanza alle vicende dei singoli territori, alle memorie di comunità, alle singole vite spezzate dalla violenza nazista e fascista, ma al contempo le si inserisce e le si inquadra in uno scenario nazionale.

## **ULTERIORI RIFERIMENTI**

1. Pezzino, Paolo. *Zone di Guerra, geografie di sangue*. A cura di Gianluca Fulveti. Bologna: Il Mulino, 2016.



# **The discovery platform GOTRIPLE: An EOSC service for social sciences and humanities research**

Francesca Di Donato<sup>1</sup>, Suzanne Dumouchel<sup>2</sup>, Monica Monachini<sup>3</sup>, Stefanie Pohle<sup>4</sup>

1 ILC-CNR - francesca.didonato [at] ilc.cnr.it

2 Huma-num-CNRS - suzanne.dumouchel [at] huma-num.fr

3 ILC-CNR - monica.monachini [at] ilc.cnr.it

4 Max Weber Stiftung - pohle [at] maxweberstiftung.de

## **ABSTRACT**

In this paper, we present TRIPLE – Transforming Research through Innovative Practices for Linked Interdisciplinary Exploration – an on-going project funded as part of the European Horizon 2020 programme INFRAEOSC-02-2019 “Prototyping new innovative services” (2019-2023). The project’s main objective is to develop a multilingual and multicultural discovery solution for the social sciences and humanities (SSH), which will provide a single access point that allows users to explore, find, access and reuse materials such as literature, data, projects and researcher profiles at European scale. The paper first provides an overview of TRIPLE’s main goals and impacts. It then describes the methodology adopted for the design and development of the project platform, GOTRIPLE. Finally, it contextualises the project within the European research landscape, and more specifically, in the European Open Science Cloud (EOSC) ecosystem. In the conclusion, some current challenges and open issues are presented.

## **KEYWORDS**

EOSC, open science, scholarly communication, discoverability, OPERAS, TRIPLE

# TALK

## 1 INTRODUCTION

TRIPLE - Transforming Research through Innovative Practices for Linked Interdisciplinary Exploration - is a project funded under the European Commission program INFRAEOSC-02-2019 “Prototyping new innovative services”<sup>1</sup>. It involves a consortium of 19 partners from 13 European countries<sup>2</sup>.

The TRIPLE project’s main objective is to design and develop a multilingual and multicultural discovery platform dedicated to social sciences and humanities (SSH) resources, which enables researchers to discover and reuse SSH macro-typologies data, i.e. research data and publications, profiles, and projects. The GOTRIPLE platform is conceived to be composed of a core component, built upon the Isidore search engine<sup>3</sup> developed by Huma-num [1], and integrated with a variety of innovative tools, which include a web annotation service (Pundit<sup>4</sup> [3]), a Trust Building System, and a recommender system. The front-end visualizations are based on open technologies developed by Open Knowledge Maps<sup>5</sup>. Moreover, in order to access data about projects, a crowdfunding platform will be connected to GOTRIPLE. A Forum complements the innovative services by offering an agora for channeling internal and external discussions among users.

GOTRIPLE platform is not a stand-alone application but is designed to be one of the innovative services provided by OPERAS - the research infrastructure supporting open scholarly communication in the social sciences and humanities in the European Research Area<sup>6</sup>. OPERAS services adopt a common Authentication and Authorization Infrastructure (AAI), thanks to the implementation of Virtual Organization common identity, which uses federated authentication mechanisms.

Apart from GOTRIPLE, OPERAS services include:

- OPERAS Metrics Service<sup>7</sup>: The service collects usage metrics and altmetrics from many different sources (such as Google Books, Matomo Analytics, World reader, etc.) about the usage of monographs.
- OPERAS Publishing Service Portal (Pathfinder)<sup>8</sup>: The service will be composed of two parts: a presentation of the services as a catalogue and a wizard that walks researchers through a series of questions towards the service offering that best fits their needs.

---

1 <https://cordis.europa.eu/project/id/863420>

2 <https://www.gotriple.eu>

3 <https://www.huma-num.fr/service/isidore>

4 <http://thepund.it>

5 <https://openknowledgemaps.org>

6 <http://operas-eu.org>

7 <https://www.operas.unito.it/services/metrics-service/>

8 <https://www.operas.unito.it/services/publishing-service-portal-pp/>

- OPERAS Certification Service<sup>9</sup>. It collects the variety of peer-reviewing practices from hundreds of monograph publishing houses, categorizes them, and provides a single access point to the list of certified peer-reviewed monographs available in Open Access in the world.
- OPERAS Research for Society<sup>10</sup>: By developing the practices of academic blogging for SSH researchers, offered by the hypotheses.org platform, it is designed to be an interactive platform between SSH researchers and the society at large.
- OPERAS XML Toolbox: The toolbox provides tools to achieve interoperability between two standards, XML JATS, adopted by the academic publishing industry, and certain flavours of XML TEI (Text Encoding Initiative) adopted by the humanities research community for books and digital editions.

Within the OPERAS-P project<sup>11</sup>, active discussions and collective think-tanks on the development of sustainable models of governance for infrastructures, business models for open scholarly publishing, groundbreaking concepts to address the fairification of SSH data, multilingualism, the future of scholarly writing as well as quality assessment of novel research outputs are on-going. As a next step, OPERAS aims at becoming a mature community, based on a set of stable national nodes, grounded on innovative plans for future development, and providing a set of services compatible with the European Open Science Cloud (EOSC)<sup>12</sup>.

The following section presents TRIPLE's approach, focusing on the methodology we are following in the design and development of GOTRIPLE, based on a user-centered design approach. In section 3 we then contextualize TRIPLE as an OPERAS service in the European landscape. In the conclusion, some critical aspects are listed, which are offered for discussion.

## 2 METHODOLOGY

TRIPLE's work plan is based on a strict and strong interaction between different activities: the core is made by a double-helix structure: one is dedicated to data acquisition and the other one to the co-design processes [2]. Both of them are like the TRIPLE DNA - which produces the discovery platform and its innovative services. This double-helix structure nurtures as well as other activities such as the definition of business models for the platform's long-term sustainability and the integration of GOTRIPLE into the EOSC. Other aspects of the work plan include project management and the communication and dissemination of the project outputs and results. These interactions are described in Figure 1, where activities are grouped under eight work packages.

---

9 <https://www.operas.unito.it/services/certification-service-doab/>

10 <https://www.operas.unito.it/services/research-for-society/>

11 <https://cordis.europa.eu/project/id/871069>

12 <https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud>

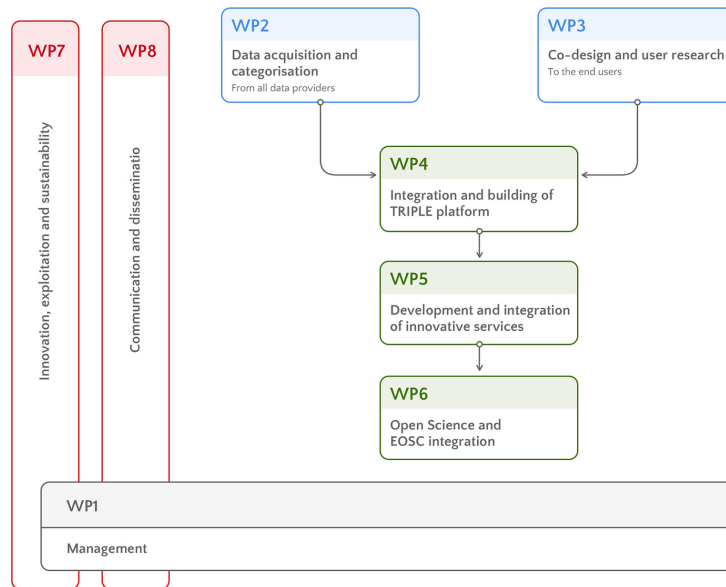


Figure 1 – TRIPLE pert chart

The methodology adopted by the TRIPLE project is based on user-centered design techniques. TRIPLE's approach in fact puts users' needs at the core of the design process. We are using a mix of social sciences and design research approaches in order to study the future GOTRIPLE users and to elicit their needs as well as to co-design the main functionalities of the platform. In particular, TRIPLE has tried to identify the user needs and to better understand the working practices of SSH researchers and researchers in order to design the GOTRIPLE platform. Through an online questionnaire over 925 survey responses could be collected. Current and future user needs also emerged from approx 40 qualitative interviews. From the the analysis of the interviews, we started defining the personas which are associated with their scenarios. Additionally, to obtain a comprehensive picture of existing working practices of researchers (e.g. what tools they are already using, how they structure collaborations, etc.), to gather further information on user needs and to better identify potential user communities, a benchmarking of TRIPLE based on the results of the questionnaire is on-going. The consortium will also conduct a number of co-design/co-creation workshops and design research activities on the user journey in order to increase the appropriation capacity of SSH researchers towards the GOTRIPLE platform. Here, the plan is to use a double approach, based on hands-on workshops with the goal to sketch prototypes of interfaces in an agile and iterative approach.

The second helix in the construction of GOTRIPLE is data acquisition through efficient coordination with data providers to ensure high-quality indexation of content in six languages in addition to those already managed by ISIDORE (French, English, Spanish), i.e. Portuguese, Italian, German, Polish, Croatian, Greek. This objective is divided into two parts: to achieve effective data acquisition during the timeframe of the project; and to set up a framework that will be used after the end of the project to continue the data acquisition process by adding new content from the current data sources, adding/removing data sources and providing guidelines for further additions of languages through machine learning and thesauri alignments.

We aim at overseeing the effective gathering of metadata and enrichment in the platform. A data acquisition plan defines the general specifications. On the one hand, support is provided to beneficiaries, OPERAS members, and potential providers to comply with the platform specifications and OpenAire<sup>13</sup>. On the other hand, data discoverability is enhanced through the use of controlled vocabularies. A minimal set of categories is used for machine learning for all the languages of the platform. A mapping of existing classifications in the field is going to be realised, and new ones are created, especially to facilitate searching in new languages on the platform. Links with the most used controlled vocabularies are provided through automated semantic annotation and completed with disambiguation tools. This work will make it possible to provide discipline-based metrics able to document SSH production scope and visibility.

At the beginning of the project, the first envisaged scenario had the objective to ensure the global harvesting process in the GOTRIPLE platform, including a presentation of three protocols used for harvesting (OAI-PMH protocol, OAI ResourceSync, and Sitemaps protocol pointing to RDFa metadata); the metadata of the resource (Dublin Core Metadata Element Set, DCMI Metadata Set), embedded metadata in the files (pdf, jpg); a description of the global harvesting process: harvesting of metadata, classification and annotation, enrichment with vocabularies, then indexing and conversion of metadata in a structured way in RDF model.

However, by using feedback experience from the ISIDORE platform, this scenario has changed for a new one, i.e a push/pull model collecting metadata through existing aggregators (ISIDORE, OpenAIRE, Narcis etc) with a delivery platform and not anymore acquiring metadata by a harvesting process. Nevertheless, the enrichment process, including normalization, classification and annotation, is still made by the TRIPLE pipeline, which gathers data provided by aggregators.

All of the information above converged into the “Data acquisition plan”<sup>14</sup>, which develops a full methodology to allow data providers and aggregators to be interoperable with both the GOTRIPLE platform according to the technical specifications and especially by adopting the TRIPLE data model and with OpenAIRE at the EOSC level. Indeed, on this second level of interoperability, the OpenAire guidelines will serve as a basis for the management and interoperability of metadata. Accordingly, the European Publication Guidelines and Aggregation Scheme is used as a management framework for cultural content. Adapted support (tutorials, webinars, workshops, documentation) will be proposed to fulfill the needs of each category of stakeholder to prepare their metadata. The OPERAS Executive Assembly, which gathers National Nodes, will be instrumental for the achievement of this task, particularly the University of Coimbra, which leads the Multilingualism Special Interest Group of OPERAS<sup>15</sup>, and IBL-PAN, that reaches out to Central European countries and languages.

---

13 <https://explore.openaire.eu/>

14 Cf. TRIPLE deliverable D4.1 - Data Acquisition Plan. The deliverable, and all the public project deliverables released, will be published on Zenodo.

15 <https://www.operas.unito.it/special-interest-groups/multilingualism/>



### 3 TRIPLE IN THE CONTEXT OF OPEN SCIENCE RESEARCH

As mentioned before, GOTRIPLE is not conceived to be a standalone platform, but a fundamental brick of the OPERAS Research Infrastructure (RI), complementing its tools and services; moreover, it is an SSH service of EOSC, and linked with the SSHOC project. Since the beginning of the project, existing research infrastructures for SSH (ERICs), namely DARIAH, CLARIN, and CESSDA<sup>16</sup>, have been members of the project consortium. Moreover, the partnership with EGI, actively involved in the definition of the European Open Science Cloud (EOSC), ensures full compliance of the platform with EOSC and its marketplace. More specifically, while planning GOTRIPLE's integration into the EOSC, we are working on two levels of integration.

The first level of integration concerns the integration of services in the EOSC Marketplace. This integration will be performed via the protocol for the integration of OPERAS RI services into the EOSC [4]. In the context of the EOSC, onboarding resources means that providers connect them with and through the EOSC, with a wide range of communities interested in knowledge production, and by following a protocol for integration. This protocol consists of a standardized administrative process aiming at collecting information about the resource in question and a validation process assessing the compliance of the provided information from service providers, thanks to a set of inclusivity criteria.

Onboarding services to the EOSC will also be possible through regional or thematic portals and their marketplaces as well, but current and established workaround onboarding is centered around the central route, namely, the integration through the EOSC Portal Marketplace, which is currently under development in the EOSC Enhance project<sup>17</sup>. At a later stage, GOTRIPLE is also expected to be integrated within the SSHOC Marketplace, which is planned to be released in Autumn 2021. Within the framework of the EOSC Enhance project, this protocol will be extended to datasets and other scholarly resource types (such as training resources), while integration through the regional and thematic clouds will enable giving access to specialized, context-dependent knowledge structures, e.g. implementing language services in regional portals or detailed, domain-specific ontologies and vocabularies in the thematic portals. In the EOSC, data repository and data discovery services, like GOTRIPLE, are onboarded as services within the EOSC Marketplace for the benefit of SSH Researchers.

A second level integration concerns the integration of EOSC Core services and can be achieved by implementing some of the EOSC Federation Services into our Services, such as the GOTRIPLE platform and innovative services. Such EOSC Federation Services include, but are not limited to, the Authorization and Authentication Infrastructure (AAI), the Helpdesk, the Accounting Service, the Monitoring Service, etc. The AAI and the Helpdesk are important additions to the GOTRIPLE platform.

Further activities of the TRIPLE project are focused on Open Science training.

---

16 <https://www.dariah.eu>

<https://www.clarin.eu/>

<https://www.cessda.eu/>

17 <https://www.eosc-portal.eu/enhance>

Planned to start at the time of writing this article, this activity will produce, adapt and reuse general and specific guidelines, as well as provide training mainly in the form of webinars - also linking to existing active projects which are already providing them such as OpenAIRE, FOSTER, DARIAH, CLARIN, and the SSHOC Training Community<sup>18</sup> - to the project members, in order to ensure they all have competencies on Open Science practices and workflows. Webinars will tackle the different technical aspects of GOTRIPLE, covering the following topics: multilingualism and vocabularies, web annotation, FAIR data, certification, visualization, API data, and data reuse.

## 4 CONCLUSION

In the process of development of GOTRIPLE we are facing four specific difficulties. The first one concerns GOTRIPLE's integration into the EOSC, while the latter is still under definition through a complex iterative process, which involves many different stakeholders. As an example, we can mention the onboarding process. At this point of maturity of the EOSC Portal, the onboarding workflow is under continuous development, which means that although drastic changes are not expected, certain aspects of the workflow and the currently used Service Description Template - which is work-in-progress - are expected to change.

The release of EOSC-related outputs by the six EOSC Working Groups is the result of a participatory process to which the TRIPLE's consortium - are asked to provide comments and feedback very frequently. In order to be able to monitor the production of the main documentation, a reporting methodology has been established by assigning a responsible partner to each working group [3]. This process allows the team to have a constant update on relevant outcomes and the progress of different EOSC Working Groups. Another difficulty derives from the richness and variety of the EOSC ecosystem, which is composed of several projects, research infrastructures, and initiatives, and actions, where a hierarchy of authoritative sources is not always clear. We are addressing this problem thanks to the involvement in our consortium of partners such as EGI, CLARIN, CESSDA and DARIAH, whose different areas of expertise and perspectives, as well as their participation in relevant related projects, helps to understand the complexity of the scenario which composes the EOSC ecosystem. In this perspective, a strong collaboration with the SSHOC consortium is expected to develop common knowledge of EOSC requirements, to ensure the cohesion between different SSH services and even to consider deeper links between the services such as mixed governance models and business models.

The second main difficulty concerns how to deal with the pandemic situation. In order to deal with this problem, we are implementing some strategies, such as creating informal sessions where we have open discussions on various topics (TRIPLE Café) and publishing an internal monthly newsletter to inform the different partners about the progress of the project.

---

18 <https://www.openaire.eu/frontpage/webinars>  
<https://www.openaire.eu/workshops>  
<https://www.fosteropenscience.eu/>  
<https://www.dariah.eu/activities/training-and-education/>  
<https://www.clarin.eu/content/funding-call-participation-clarin-trainer-network-programme>  
<https://sshopencloud.eu/training>

The third main difficulty is related to the ambition of the TRIPLE project: having a multilingual and interdisciplinary platform to highlight multiculturalism practices. To achieve this goal, important work has to be done on the alignment of vocabularies (for each discipline of the SSH - currently 27 following the MORESS categories and in nine languages). In itself, this task is a challenge that will contribute to the whole SSH community.

Last but not least, TRIPLE specificity is to harvest many small SSH data repositories in the different European countries in order to make them more visible and reusable by the scientific community.

However, the practices of the repositories are very different, and there is no harmonization yet between them on their data model. This means for the TRIPLE consortium to support the data providers and aggregators in their transition to a more common data model, the future TRIPLE data model.

## 5 ACKNOWLEDGEMENTS

The authors want to thank Yoann Moranville (DARIAH), Paula Forbes (Abertay University), and Mélanie Bunuel (Huma-num - CNRS) for their contributions to this paper.

The project is funded by the European Commission, under Grant Agreement No. 863420.

## REFERENCES

1. Dumouchel, Suzanne. «The EOSC as a knowledge marketplace: the example of ISIDORE», 2018. <https://hal-amu.archives-ouvertes.fr/hal-01708784/>.
2. Dumouchel, Suzanne, Emilie Blotière, Gert Breidfuss, Yin Chen, Francesca Di Donato, Maria Eskevich, Paula Forbes, Haris Georgiadis, Arnaud Gingold, e Elisa Gorgaini. «GOTRIPLE: A User-Centric Process to Develop a Discovery Platform». *Information* 11, n. 12 (2020): 563. <https://doi.org/10.3390/info11120563>.
3. Grassi, Marco, Christian Morbidoni, Michele Nucci, Simone Fonda, e Francesca Di Donato. «Pundit: Creating, Exploring and Consuming Semantic Annotations». In *Proceedings of the 3rd International Workshop on Semantic Digital Archives*, 65–72. SDA, 2013.
4. OPERAS-P Consortium. «OPERAS-P Deliverable: Protocol for the integration of OPERAS RI services into EOSC», 28 agosto 2020. <https://doi.org/10.5281/ZENODO.4005677>.
5. TRIPLE. «TRIPLE Deliverable: Report on the General Interoperability Requirements», 30 settembre 2020. <https://zenodo.org/record/4005677>.

# **Sessione II**

## **Claire Simon**



# Cultural Heritage for All with Virtual Reality: early findings of a Scenario-Based Design approach

Barbara Balbi<sup>1</sup>, Alessandra Marasco<sup>2</sup>

1 MIBACT – SABAP (Comune di Napoli) – barbara.balbi [at] beniculturali.it

2 CNR-IRISS – a.marasco [at] iriss.cnr.it

## ABSTRACT

This paper describes the findings of the early stage of a research aimed to design a virtual reality (VR) solution as a compensatory tool for improving the accessibility of cultural heritage sites for visitors with mobility impairments. For the research, the Scenario-Based Design method (SBD) is applied to identify user requirements of the VR interface within a User-Centered Design approach. Based on this method, different **personas** and corresponding **scenarios** are designed to represent the needs, motivations, and behaviors of the main target groups of users and the applications of compensatory technological solutions. Further, the SBD is used in the prototyping phase as a tool for the co-design activity, which involved users with specific accessibility needs, designers, and other stakeholders to identify the requirements of the VR solution. The paper presents the preliminary findings relating to the functional and experiential requirements and discusses the methodological and practical implications to support the use of this approach for the co-design of digital solutions for cultural heritage accessibility.

## KEYWORDS

cultural heritage, accessibility, virtual reality, scenario-based design, user experience, human-computer interaction

## TALK

### 1 BACKGROUND

The design of digital applications to improve physical, intellectual, and sensorial access to cultural heritage visitors with different accessibility needs is a growing research area, grounded on the need to guarantee the right to culture as stated by the Universal Declaration of Human Rights [2; 4; 5;]. Studies in this area address the design and impact of virtual reality (VR) as a tool to improve accessibility at heritage sites [6;1;18;8;11]. In this direction, User-Centered Design (UCD) [9] is supported as the preferred approach, in line with the principles of Design for All [2] and design for user empowerment [10]. Among the UCD methods, Scenario-Based Design (SBD) [16] has been used for the development of tourism digital experiences [12] and in digital support systems for museums visits and exhibitions [7; 15].

This methodology is useful to engage users, designers, and other stakeholders in collaborative design as an effective bridge between designers and the end-users [12]. It provides the possibility of visualizing a case of use in a narrative (i.e. a storyboard) of a typical situation in which the end user interacts with the digital solution. In this way, it facilitates the users to imagine possible difficulties or advantages in using the digital application and the designers to adequately consider the contextual and social factors for its sustainability. In this sense, the scenario becomes the design tool that can be modified, enriched, extended in the co-designing activities in parallel to the technological implementation.

However, there is a lack of research focusing on the use of this specific approach to support the identification of the dimensions of the visitor experience to be enriched with VR and the requirements for an effective interaction with the digital solution in the context of cultural heritage accessibility. This paper aims to provide an initial contribution in this direction by describing the findings of a study that adopted SBD as a tool to identify the user and functional requirements of the VR interface from the concept generation to the prototype design.

## 2 MATERIALS AND METHODS

This study is conducted within a research project aimed to develop a VR mobile application to improve visitors' experiences at heritage sites where physical barriers cannot be easily eliminated to allow full accessibility [12]. Based on UCD, the research involved users with permanent and temporary mobility impairments, including the elderly, in all the stages of the design cycle (Fig. 1), in order to fully take into account their expectations, needs, and requirements in relation to the use of VR as a compensative experience of inaccessible cultural heritage. In the explorative stage of research, a multidisciplinary approach at the crossroad of tourism experience design and innovation, human-computer interaction, and heritage valorization has been applied to develop the personas and preliminary scenarios as tools to support the co-design activity in the prototype phase [11]. Personas have been developed on the profile features of the main target groups for the VR solution, in relation to three main features (as exemplified in Fig. 2): 1) accessibility needs [17], 2) engagement with cultural heritage, 3) attitude and use of technologies in tourism/leisure activities, including VR. The four personas represent the protagonists of four corresponding preliminary scenarios. Each scenario was matched with three elements, as detailed in Table 1: a) level of site accessibility; b) users' motivation for visiting the heritage site; and c) expertise or attitude in technology use.

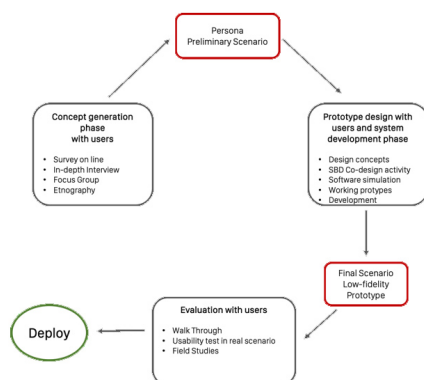


Figure 1: the UCD Cycle for the research

**Alfonso**

Alfonso is a member of an association that supports the personal empowerment of people with disabilities and thus is aware of every critical issues of physical accessibility. He moves on a folding wheelchair. Concerning travels and cultural visits, he's very downhearted, since his experiences were never completely positive, but he still searches for accessible places and experiences and collaborates, with his association, to their organization. He is very critical concerning the scarce attitude and competences of the staff at cultural heritage sites and the lack of appropriate information on the accessibility of public structures. However, his previous cultural experiences with devices for virtual and immersive visits impressed him favourably.

The first point of his association's manifesto is:  
*Accessibility is an holistic concept that has significant importance for the life's quality of the whole humanity, for a complete society is inclusive and accessible in physical, mental and cultural terms.*

Age: 55

Traits: Informed and conscious

Employment: President of the association

Curiosity for innovations: 5/5

Willingness to cooperate: 4/5

Involvement: 4/5

From the solution he expects: not a special solution only for people with disabilities nor a completely different visit, but a way to fully share the visit experience with others.

Figure 2: description of a person

| Accessibility of the site         | Personas' motivation and attitude towards technology  | Scenarios' Storyboard   |
|-----------------------------------|---|---|
| Totally inaccessible              | Persona 1: High motivation to visit heritage, high attitude to use technology (included VR) | Persona 1 is a young woman with a permanent mobility impairment would go for a cultural visit and a friend proposes her to visit an amazing church that unfortunately is totally inaccessible... (to be continued).   |
| Partially accessible              | Persona 2: Medium motivation, medium attitude to use technology, low experience with VR.    | Persona 2 (Fig. 2) is a man with a permanent mobility impairment, member of an association for people with disabilities, and he is looking for a site for the annual journey day. He finds an interesting partially accessible cultural site ... (to be continued). |
| Partially accessible              | Persona 3: Low motivation, high attitude in use of technology, included VR.                 | Persona 3 is a young boy, with a broken leg, who is visiting a site with his classroom, but he is very bored and the partially accessibility of the heritage site is a barrier for him... (to be continued).  |
| Totally or partially inaccessible | Persona 4: Medium motivation, low attitude in use of technologies, no VR experience.        | Persona 4 is an old woman with a great desire of visiting a cultural site with her grandniece, but her mobility is limited, and the site is not totally accessible... (to be continued).  |

Table 1: the three design features for preliminary scenarios.

As an example, in the second scenario Alfonso, a man with a permanent mobility impairment has to define the annual recreative program for his association, in which a cultural visit is planned; the corresponding scenario (Fig.3) described needs, reactions, problems that can arise in this typical situation.

### 3 THE CO-DESIGN ACTIVITY

The preparatory tools described above were used in the participatory design activities, aimed to obtain the guidelines for the development of the VR solution in terms of experiential and functional users' requirements. In order to carry on the co-design phase, pairs of users with different accessibility needs, designers and stakeholders, have been created to work together in envisioning the use of VR as a medium to visit the inaccessible cultural site, using the preliminary scenarios as a basis to create and visualize a personalized story. The co-design activity involved ten participants -the corresponding users to personas, the designers of VR interface, the designers of experience, and a stakeholder- as detailed: president of a disability association, user with permanent mobility impairment/ user with permanent mobility impairment / user with temporary mobility impairment/ old user with mobility impairment/ accessible tourism expert from industry/ hardware developer/ UI Designer/ experience Designer/ UX Designer/ accessible tourism expert from industry. The activity was conducted during a Focus Group organized in a single day at the headquarters of a disabled people association.



It was divided into three moments: a) the empathize phase (participants' self-presenting, brainstorming on the accessibility of cultural heritage); b) evaluation of different digital solutions for cultural visits (a test of Oculus RIFT device with a VR cinematic experience of a cultural site; an app for the virtual guided tour of a museum with smartphone and cardboard; a 3D model desktop base visit of an ancient cathedral) for an in-depth understanding of the requirements of the VR solution; c) co-design of scenarios (combination of pairs users/designers, introduction of preliminary scenarios structured as storyboards, co-design).



Figure 3: an excerpt from preliminary scenarios

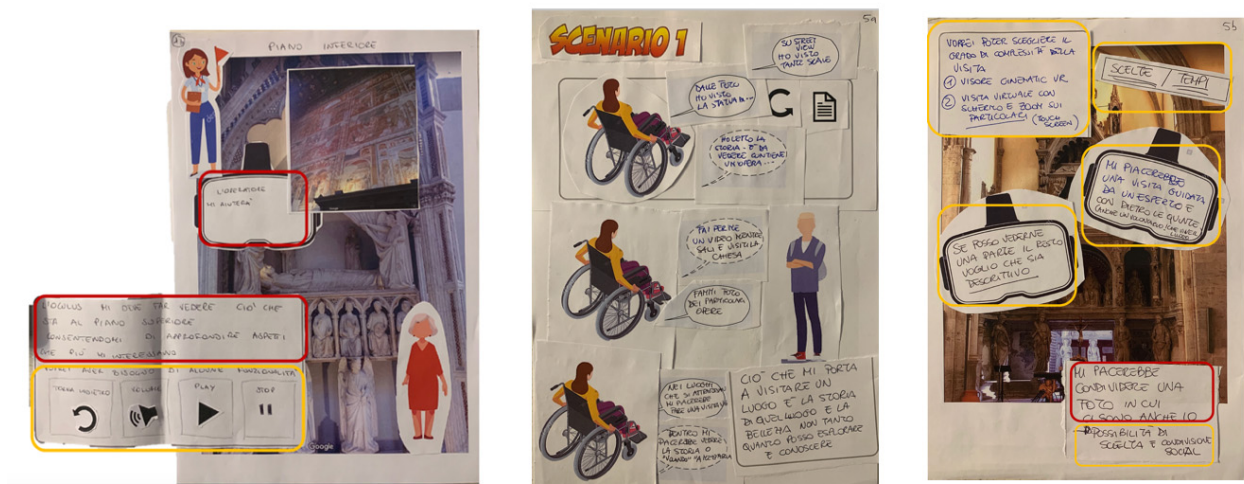


Figure 4: an excerpt of scenarios envisioned by a user during the co-design activity

Each pair chose one of the preliminary scenarios and with the collage technique designed the sequel of the story visualizing the VR experience in relation to both experiential and functional requirements (Fig. 4).

## 4 FINDINGS

The co-design activity based on SBD allowed identifying the experiential as well as the functional elements for the VR-enabled accessible visit, as detailed as follows.

Experiential requirements:

- a. Universal solution - The experiential elements should satisfy the requirements of users with different accessibility needs in line with Design for All principles. The VR solution is expected to be implemented as an innovative experience for all the potential visitors onsite rather than as a separate experience provided only to users with mobility impairments. Dedicated solutions for disabled visitors are indeed negatively perceived as marginalizing experiences.
- b. Cultural engagement - Curiosity for the cultural experience is the driving force of every imagined scenario, independently from the level of cultural motivation.
- c. Connectedness - The possibility to share the experience with users with full access possibility is a fundamental element of the cultural experience, in line with tom Dieck et al. [3] and Marasco e Balbi [11], and it is imagined as the live streaming of contents during the visit.
- d. Intellectual stimulation - The virtual experience is a trigger to arouse the interest and motivation of users to learn about the site, as also outlined by Paladini et al. [14]. It is stressed that it should be promoted during the pre-visit phase in order to engage users with different accessibility needs.
- e. Storytelling – The narrative modality emerges as an important element; the involvement through an engaging storytelling is requested by users, as previously found by tom Dieck et al. [3], especially in the initial moments of the on-site visit.
- f. Autonomy – the smartphone is preferred to head-mounted displays in relation to the possibility to use it more autonomously. For VR devices, human assistance is requested by users. If the navigation has an immersive modality, it must include at least an interactive independent part through a smartphone or a screen display.

Functional requirements:

- a. Multimodal interaction. Narrative modalities of contents are strictly linked with the devices' functional requirements: for a more immersive visit with the possibility to explore many details, mixed reality and the use of different devices is preferred;
- b. Minimal action/Minimalist design. The immersive modality is requested in the first moment of the visit, but, aim to autonomous use, the users need an interface menu useful to navigate in all directions and zoom on details of the site through basic controls;
- c. Capacity of immersion/ user participation/use of diegetic and extra-diegetic sound. In relation to the point of view, generally, it is described as very close to artifacts, even "inside" them, i.e. the visitor wants to be immersed into the scenes of a painting. In addition, they were sharing the point of view of those who are not in the same part of the site via streaming. For improved accessibility, the functional requirement should provide for details of the site, such as the presence of inaccessible areas as well as details regarding impediments and difficulties (floors, ceilings) in the areas that the user can visit. A voice-over of an expert is requested to fulfill intellectual needs;
- d. Application access. For increased autonomy, no profiling is preferred, but just an automatic link when the visitor is onsite through personal devices;

e. Ease of use. Headsets/headphones are not preferred as they could not be handled independently;  
 f. Co-located collaboration/ Remote collaboration. The designed scenario includes the sharing of the experience through various modalities: the posting of photographs after the visit through the application, during the visit via chat (i.e. in the form of call out), through a form of interaction like multi-player. The requirements obtained with SBD method can be integrated into a design framework that establishes a correspondence between experiential and functional requirements (Fig. 5), summarized in the final scenario (Fig. 6), which supports the development of the low-fi prototype of the VR solution.



Figure 5 (left): A framework integrating experiential and functional requirements of VR for accessible cultural heritage

Figure 6 (right): Extract from the final scenario to be used for low-fi prototyping

## 5 CONCLUSION AND FUTURE STEPS

Notwithstanding the increased emphasis on user engagement to address the challenge of accessibility from a design for all perspective, research on the tools and methods for the co-creation and co-design of digital solutions for user empowerment is still in its infancy. This study aims to contribute knowledge in this direction by providing insights into the process, tools, and outcomes of user experience design based on SBD. In particular, this study illustrates how Scenario-Based Design can be used to support the design of a VR solution as a compensatory experience for heritage accessibility [1] through the identification of the requirements of users with disabilities, with specific regard to mobility impairments. As the main advantage for the design process, it provides the possibility to visualize the specific context of use and the visitor interaction therein.

The early findings allowed to envision the possibilities of interaction with the VR solution on site in three different cultural heritage sites characterized by different level of accessibility.

This method provided VR designers with useful knowledge about who are the possible users to design for as well as the socio-cultural contexts of their lives that frame the VR-enabled experiences. From a design for all perspective, the value of this method was supported in enabling and empowering users to actively co-create the VR solution in relation to their needs and expectations. As a potential drawback, it needs to be integrated with an early prototyping process in order to evaluate the solutions during the all iterative cycle of design. In this sense, the SBD approach can be usefully leveraged to refine the interfaces that will be evaluated, imagining their use in a specific context. Based on the early findings, the co-design activity provided development guidelines for the VR solution that have been mostly implemented within the research project in relation to the choice of the device, content development, user interaction, and interfaces. A further step of the project will focus on a unified test of usability, accessibility, and impact of the VR solution on-site in the real scenario with end-users. In particular, the effectiveness of the solution needs to be assessed in relation to different dimensions of users' experience at cultural heritage sites, including engagement with cultural heritage, intellectual stimulation, connectedness, authenticity, also in consideration of the experiential requirements emerged in the co-design research. Further, the accessibility of the solution deserves more attention also in relation to the possibility of improving personalization features for wider users' access needs in line with recent research [13].

## **6 FUNDING**

This study is supported by the Campania Region (POR FESR CAMPANIA 2014/2020) - Project #In-Heritage - Integrated Technology System for Accessing In-Visible Heritage.

## REFERENCES

1. Agostiano, Maria. «Enjoyment of Cultural Heritage and “Compensatory Solutions” for an Alternative Accessibility». In *Accessibility as a key enabling knowledge for enhancement of cultural heritage*, di A. Arengi, I. Garofalo, e O. Sormoen, 131–47. Milano: Franco Angeli, 2016.
2. Arengi, Alberto, e Maria Agostiano. «Cultural Heritage and Disability: Can ICT Be the ‘Missing Piece’ to Face Cultural Heritage Accessibility Problems?» In *International Conference on Smart Objects and Technologies for Social Good*, 70–77. Springer, 2016.
3. Dieck, M. Claudia tom, Timothy Jung, e Eleni Michopoulou. «Experiencing virtual reality in heritage attractions: Perceptions of elderly users». In *Augmented reality and virtual reality*, 89–98. Springer, 2019.
4. Fogli, Daniela, e Alberto Arengi. «“Design for All’versus” One-Size-Fits-All’: The Case of Cultural Heritage.» In *CoPDA@ AVI*, 89–96, 2018.
5. Fogli, Daniela, Davide Sansoni, Emanuele Trivella, Alberto Arengi, e Ivana Passamani. «Advanced interaction technologies for accessible and engaging cultural heritage». In *International Conference on Smart Objects and Technologies for Social Good*, 364–373. Springer, 2017.
6. Guttentag, Daniel A. «Virtual reality: Applications and implications for tourism». *Tourism management* 31, n. 5 (2010): 637–651.
7. Hall, T., e L. Bannon. «Co-operative design of children’s interaction in museums: a case study in the Hunt Museum». *CoDesign* 1, n. 3 (2005): 187–218.
8. Han, Dai-In Danny, Jessika Weber, Marcel Bastiaansen, Ondrej Mitas, e Xander Lub. «Virtual and augmented reality technologies to enhance the visitor experience in cultural tourism». In *Augmented reality and virtual reality*, 113–128. Springer, 2019.
9. ISO, BSEN, e BRITISH STANDARD. «Ergonomics of human-system interaction», 2010.
10. Ladner, Richard E. «Design for user empowerment». *interactions* 22, n. 2 (2015): 24–29.
11. Marasco, Alessandra, e Barbara Balbi. «Designing Accessible Experiences for Heritage Visitors Through Virtual Reality.» *E-review of Tourism Research* 17, n. 3 (2019).
12. McCabe, Scott, Mike Sharples, e Clare Foster. «Stakeholder engagement in the design of scenarios of technology-enhanced tourism services». *Tourism Management Perspectives* 4 (2012): 36–44.
13. Montagud, Mario, Pilar Orero, e Anna Matamala. «Culture 4 all: accessibility-enabled cultural experiences through immersive VR360 content». *Personal and Ubiquitous Computing*, 2020, 1–19.
14. Paladini, Alice, Abhijit Dhanda, Miquel Reina Ortiz, Adam Weigert, Eslam Nofal, A. Min, M. Gyi, S. Su, Koen Van Balen, e Mario Santana Quintero. «Impact of virtual reality experience on accessibility of cultural heritage». *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 42 (2019): 929–936.
15. Radder, Laetitia, e Xiliang Han. «An examination of the museum experience based on Pine and Gilmore’s experience economy realms». *Journal of Applied Business Research (JABR)* 31, n. 2 (2015): 455–470.

16. Rosson, Mary Beth, e John M. Carroll. «Scenario based design». *Human-computer interaction. boca raton, FL*, 2009, 145–162.
17. Schulz, Trenton, e Kristin Skeide Fuglerud. «Creating personas with disabilities». In *International Conference on Computers for Handicapped Persons*, 145–152. Springer, 2012.
18. Southall, Helen, Maeve Marmion, e Andrew Davies. «Adapting Jake Knapp’s design sprint approach for AR/VR applications in digital heritage». In *Augmented reality and virtual reality*, 59–70. Springer, 2019.



# Il progetto Overtheview: schema di progettazione per l'accessibilità museale

Gavino Paddeu<sup>1</sup>, Anna Maria Marras<sup>2</sup>, Andrea Ferrero<sup>3</sup>, Francesca Mura<sup>4</sup>, Antonio Pintori<sup>5</sup>

<sup>1</sup> CRS4 – Centro di Ricerca, Sviluppo e Studi Superiori in Sardegna – gavino.paddeu [at] crs4.it

<sup>2</sup> Università di Torino – annamaria.marras [at] unito.it

<sup>3</sup> CRS4 – Centro di Ricerca, Sviluppo e Studi Superiori in Sardegna – andrea.ferrero [at] crs4.it

<sup>4</sup> CRS4 – Centro di Ricerca, Sviluppo e Studi Superiori in Sardegna – francesca.mura [at] crs4.it

<sup>5</sup> CRS4 – Centro di Ricerca, Sviluppo e Studi Superiori in Sardegna – antonio.pintori [at] crs4.it

## SINTESI

Negli ultimi anni nei musei l'uso delle tecnologie digitali per supportare l'accessibilità è aumentato considerevolmente, ma occorre fare una scelta delle tecnologie digitali “adatte” ad un determinato contesto museale. Per questo motivo è stato elaborato uno schema/scheda di supporto per la scelta delle tecnologie digitali, schema che è stato seguito nella realizzazione dei prototipi e delle applicazioni nell'ambito del progetto Overtheview. In questo contributo presentiamo i risultati della ricognizione sulle tecnologie accessibili in ambito museale, lo schema di progettazione e le applicazioni che sono state realizzate dal progetto Overtheview.

## PAROLE CHIAVE

musei, accessibilità, intelligenza artificiale, NLP, Design for All

## INTERVENTO

### 1 INTRODUZIONE AL PROGETTO

Il progetto OvertheView si pone l'obiettivo di approfondire le conoscenze nel campo delle esposizioni museali accessibili in una prospettiva di progettazione universale e con particolare riguardo alle persone con disabilità sensoriale. Il proposito del progetto è la valutazione di una modalità progettuale che utilizzi le tecnologie allo stato dell'arte, per lo sviluppo di nuove applicazioni al fine di rendere interessante, piacevole ed educativa l'esperienza di una visita museale anche per le persone con disabilità sensoriale.

Nel presente lavoro viene descritto lo schema progettuale, il suo impiego e la sperimentazione nei casi relativi ai plastici interattivi del Santuario Nuragico “Santa Vittoria” di Serri e dell'Area Archeologica “Genna Maria” di Villanovaforru.



## 2 LE TECNOLOGIE PER L'ACCESSIBILITÀ DIGITALE

Le tecnologie digitali sono degli strumenti che, se realizzate adeguatamente<sup>1</sup>, rendono le informazioni culturali più accessibili [1], aspetto questo messo in evidenza anche dal Piano Triennale per la Digitalizzazione e l'Innovazione dei Musei<sup>2</sup>. Secondo lo studio condotto dal progetto europeo LEM<sup>3</sup> emerge come sia fondamentale non solo la qualità dei contenuti pubblicati (testi, immagini) ma anche l'usabilità dei supporti. Per questo motivo sempre di più si tende a realizzare delle soluzioni che siano utilizzabili attraverso i dispositivi (smartphone, tablet etc...) dei visitatori attraverso il principio del Bring Your Own Device (BYOD), questo vale anche per i supporti di visita più diffusi come le audioguide scaricabili direttamente sul proprio dispositivo.

Una delle fasi del progetto ha previsto una ricognizione sulle applicazioni e sugli strumenti per l'accessibilità presenti nei musei italiani, da questa analisi emerge che l'aspetto della tattilità, probabilmente grazie anche alla maggiore diffusione delle stampanti 3D, riveste un ruolo importante nei percorsi museali.

Gli altri strumenti molto usati sono i video nella lingua dei segni (LIS) sia in italiano che in inglese accompagnati dai sottotitoli disponibili non solo nelle sale dei musei, ma anche come contenuti multimediali delle applicazioni realizzate. Invece tra le tecnologie che si possono definire di supporto per la fruizione di contenuti le più usate sono quelle QR Code (Quick Response Code) e RFID (Radio Frequency Identification).

## 3 SCHEMA DI PROGETTAZIONE

Nella progettazione di tecnologie e percorsi accessibili, oltre a far riferimento alle diverse linee guida disponibili<sup>4</sup>, si parte dai principi che guidano la progettazione universale, a questi si devono aggiungere alcuni elementi che possono determinare la buona realizzazione del progetto di accessibilità digitale in un dato contesto museale e culturale. Per questo motivo abbiamo realizzato un supporto progettuale sotto forma di scheda / schema (Tabella 1) pensato come guida per aiutare la scelta delle tecnologie. Lo schema proposto è pensato come uno strumento utile al direttore o al responsabile del museo per scegliere la tecnologia più adatta alle esigenze del museo; ad esempio se voglio realizzare un'applicazione è necessario che il museo abbia wi-fi per il collegamento dei dispositivi, pensare ai costi non solo di realizzazione dell'applicazione ma anche di formazione del personale, aggiornamento, ecc.

- 
- 1 Smithsonian Guidelines for Accessible Exhibition Design  
<https://www.thc.texas.gov/public/upload/publications/Smithsonian%20Guidelines%20for%20accessible%20design.pdf>
  - 2 <http://musei.beniculturali.it/wp-content/uploads/2019/08/Piano-Triennale-per-la-Digitalizzazione-e-l%E2%80%99Innovazione-dei-Musei.pdf>
  - 3 Learning in Museums and Young People - Network of European Museum Organisations.
  - 4 Guida Esposizioni e percorsi di visita accessibili, traduzione a cura di + Cultura accessibile  
<http://www.cinemanchio.it/esposizioni-e-percorsi-di-visita-accessibili/>  
Smithsonian Guidelines for Accessible Exhibition Design  
<https://www.thc.texas.gov/public/upload/publications/Smithsonian%20Guidelines%20for%20accessible%20design.pdf>  
Linee guida sull'accessibilità informatica dell'Agenzia Nazionale per l'Italia Digitale  
<https://www.agid.gov.it/it/agenzia/stampa-e-comunicazione/notizie/2019/08/09/disponibili-linee-guida-laccessibilita-degli-strumenti-informatici>

| Target utenti             | Caratteristiche socio-demografiche  |               | Caratteristiche museologiche |              | Range                             |                | Caratteristiche fisiche e psicologiche |              |
|---------------------------|-------------------------------------|---------------|------------------------------|--------------|-----------------------------------|----------------|--|--------------|
| Infrastruttura Network    | WiFi                                |               |                              |              | 4G/5G copertura                   |                |  |              |
| Spazio ICT                | Accessibilità                       |               |                              |              | Dimensioni                        |                |  |              |
| Caratteristiche del museo | Coerenza con il percorso espositivo |               |                              |              | Coerenza con la mission del museo |                |  |              |
| SW                        | Usabilità e UX                      | Licenza d'uso | Manuale                      | Formazione   | Manutenzione                      | Assistenza     | Analytics                              |              |
| HD e Devices              | Manutenzione                        | Consumi       | Formazione                   | Riusabilità  | Sostenibilità                     | Manuale utente | Insurance                              |              |
| Contenuti                 | Colori                              | Font          | Dimensioni                   | Allenamenti  | Comprensibilità                   |                | Leggibilità                            |              |
| Stakeholders              | Focus group                         |               | Monitoraggio                 |              | Analytics                         |                | Soddisfazione                          |              |
| Budget                    | Sostenibilità                       | Formazione    | Supporto                     | Manutenzione | Aggiornamento                     | Design         | Sviluppo                               | Monitoraggio |

Tabella 1: Schema/scheda di progettazione delle tecnologie

## 4 I PROGETTI E I CASI STUDIO

La scheda descritta in precedenza è stata utilizzata per la progettazione di due installazioni interattive composte da fedeli rappresentazioni 3D del complesso nuragico di “Genna Maria” di Villanovaforru e del Santuario Nuragico “Santa Vittoria” di Serri.

Si tratta di plastici in scala 1:100 e 1:200 realizzati attraverso acquisizioni aerofotogrammetriche effettuate con l’ausilio di un drone per ottenere un modello digitale ad altissima risoluzione. Il modello 3D, realizzato in MDF, prevede la predisposizione di scassi per l’inserimento delle parti architettoniche realizzate in stampa 3D.

Nelle aree di maggiore interesse storico e architettonico sono stati inseriti dei sensori che, se sfiorati, permettono all’applicazione di riconoscere quale parte del plastico è stata toccata. Il visitatore può esplorare tattilmente il modello ottenendo descrizioni accurate e puntuali e formulare delle domande per soddisfare qualunque tipo di curiosità (Figura 1).



Figura 1: Esplorazione tattile del plastico interattivo.

I sensori sono collegati a una scheda basata su microcontrollore riconosciuta dal computer come periferica generica.

Attraverso il software dell'audioguida, realizzato con apposito motore conversazionale risultato dagli studi nel campo della NLP (Natural Language Processing), e configurando un contenuto audio-testuale differente per ciascun sensore, è possibile offrire una descrizione molto dettagliata all'utente che si accinge a esplorare il plastico.

Al tocco di un sensore, nel monitor annesso, compare l'animazione del modello digitale e un testo scorrevole relativi alla zona esplorata [4].

Di seguito sono descritti gli elementi indicati nella scheda di progettazione.

Target utenti: il target era costituito da persone di tutte le età, con particolare riguardo a quelle in età scolare, vista la loro predisposizione all'utilizzo di sistemi tecnologici. L'analisi delle caratteristiche museologiche degli utenti ci ha indirizzato verso utenti assimilabili a un pubblico amatoriale ma comunque appassionato di storia antica e abituato alla frequentazione dei musei.

Infrastruttura Network e spazio ICT: una ricognizione accurata degli spazi e delle infrastrutture all'interno del Museo Archeologico Nazionale di Cagliari, ci ha confermato la presenza della rete Wifi e di un adeguato spazio accessibile anche alle persone con disabilità.

Caratteristiche del museo: l'installazione interattiva progettata è completamente integrabile, dal punto di vista dei contenuti, con il percorso museale: infatti sia il museo di Cagliari che quello di Villanovaforru espongono un gran numero di reperti etestimonianzerelativi al periodo nuragico.

HD e SW: per quanto riguarda l'Hardware e il Software, ci si è maggiormente soffermati, in fase di progettazione, sulla necessità di un sistema di facile gestione (maintenance) e di una formazione adeguata (staff training) per il personale del museo.

Contenuti: inizialmente i contenuti dei testi sono stati concepiti per essere rivolti alle persone con disabilità visiva e, in una fase successiva, sono stati adattati anche per le persone sorde intervenendo sulle modalità di presentazione dei testi relativamente al colore e al tempo di permanenza.

Stakeholder: il gruppo di lavoro è stato composto da ricercatori e tecnologi dell'ICT, esperti di comunicazione e di accessibilità, archeologi ed esperti di educazione museale.

Budget: il progetto è completamente finanziato con fondi POR FESR 2014/2020, Asse prioritario I "Ricerca scientifica, sviluppo tecnologico e innovazione".

Alla sperimentazione ha partecipato un campione formato da 115 utenti composto anche da persone con disabilità sensoriali, ognuno dei quali ha posto almeno una ventina di domande alla guida audio-testuale, ottenendo un feedback positivo che mostra una soddisfazione complessiva dell'80%.

Le principali osservazioni che ci sono state indicate vertono sulla velocità di permanenza del testo sullo schermo e sul miglioramento delle funzionalità per le persone sorde.

## **5 CONCLUSIONI**

Le tecnologie digitali sono quindi un supporto importante, ma, se non ben progettate, possono costituire una ulteriore barriera e per questo motivo è necessario seguire una attenta analisi che prenda in esame tutti gli aspetti progettuali e le indicazioni già esistenti per applicazioni accessibili. Il progetto Overtheview si pone proprio questo obiettivo unendo competenze tecnologiche di alto livello con competenze specialistiche dell'accessibilità museale, nella comunicazione accessibile, nella tifologia e nella integrazione sociale dei disabili.

## BIBLIOGRAFIA

1. Marras, Anna Maria. «L'accessibilità digitale». In *Web Strategy museale, Monitorare e progettare la comunicazione culturale nel web*. Milano: ICOM, 2019.
2. NEMO - LEM Working Group. *Learning in Museums and Young People*. Berlin: NEMO – The Network of European Museum Organisations, 2015.
3. Neumüller, Moritz, Andreas Reichinger, Florian Rist, e Christian Kern. «3D printing for cultural heritage: Preservation, accessibility, research and education». In *3D Research Challenges in Cultural Heritage*, 119–134. Springer, 2014.
4. Paddeu, Gavino, Andrea Devola, Andrea Ferrero, e Antonio Pintori. «INTERACTIVE AUDIO-TEXT GUIDE FOR MUSEUM ACCESSIBILITY». In *Proceedings of the 18th International Conference on WWW/Internet 2019*, 186–88. IADIS Press, 2019. [https://doi.org/10.33965/icwi2019\\_201913P027](https://doi.org/10.33965/icwi2019_201913P027).



# La collaborazione inclusiva: un'esperienza didattica di annotazione tramite Euporia

Federico Boschetti<sup>1</sup>, Marilena Crucitti<sup>2</sup>, Filippo Lepori<sup>3</sup>, Andrea Taddei<sup>4</sup>

<sup>1</sup> CNR-ILC e VeDPH – federico.boschetti [at] yahoo.com

<sup>1</sup> Liceo classico Garballo (Siracusa) – marichiocciol [at] gmail.com

<sup>1</sup> Università di Pisa – pippolepori [at] icloud.com

<sup>1</sup> Università di Pisa – andrea.taddei [at] unipi.it

## SINTESI

Il presente contributo descrive il progetto EuporiaEdu, che ha lo scopo di promuovere pratiche di annotazione linguistica e stilistica di testi letterari nei licei e negli istituti superiori.

## PAROLE CHIAVE

domain-specific languages, accessibilità, didattica inclusiva

## INTERVENTO

### 1 INTRODUZIONE

Il presente contributo descrive il progetto EuporiaEdu, che ha lo scopo di fornire uno strumento informatico per la didattica delle lingue antiche (greco e latino) e moderne (italiano) e di promuovere pratiche di annotazione linguistica e stilistica di testi letterari nei licei e negli istituti superiori.

Si tratta di un progetto nato dalla collaborazione tra il Laboratorio di Antropologia del Mondo Antico (LAMA) dell'Università di Pisa, il Laboratorio di Filologia Collaborativa e Cooperativa (CoPhiLab) dell'Istituto di Linguistica Computazionale "A. Zampolli" del CNR (ILC-CNR) di Pisa e due licei classici, ovvero il Liceo "Galileo Galilei" di Pisa e il Liceo "Tommaso Gargallo" di Siracusa.

EuporiaEdu è un sistema di annotazione collaborativa basata su Domain Specific Languages (DSL), vale a dire su linguaggi formali che servono a strutturare le informazioni in modo compatto e ottimizzato per ambiti specifici del sapere (come ad esempio la filologia classica e la critica letteraria). Ciò riduce il sovraccarico cognitivo degli annotatori, che possono concentrare la loro attenzione più sul contenuto dell'annotazione che sulla complessità imposta da formalismi eccessivamente verbosi, come XML-TEI.

In linea con il principio della familiarità, il sistema di annotazioni di Euporia si serve dell'uso di hash-tags, dato che gli studenti hanno l'abitudine di usarli sui social networks per evidenziare alcuni aspetti di una foto (Instagram) o di una frase (Facebook e Twitter).

Uno dei principali punti di forza della piattaforma EuporiaWeb, creata per applicare la metodologia basata su DSL, è la versatilità, sul modello degli strumenti sviluppati dall'Alpheios Project: si presta infatti a vari tipi di studio (morfosintattico, lessico-semantico, metrico, stilistico) dei testi letterari, adattandosi agli obiettivi didattici dei singoli docenti. Proprio per questa ragione è stato possibile sviluppare con Euporia due progetti diversi, distinti per obiettivi specifici e per età evolutiva dei discenti, nei due licei che hanno aderito all'iniziativa di annotazione tramite DSL.

Una classe del quinto anno del liceo classico "Tommaso Gargallo" di Siracusa ha concluso, nell'anno scolastico 2019-2020, un lavoro triennale sulla parodo e su una parte del primo episodio dei Persiani di Eschilo. Scopo principale del lavoro è stato realizzare un'analisi delle figure retoriche, dei composti omerici e dei nomi di persone, di popoli e di luoghi menzionati nella prima parte della tragedia eschilea; gli studenti, inoltre, hanno fornito una traduzione collettiva, nata dallo studio approfondito dell'originale e dal continuo confronto critico con la traduzione di Giuseppe Fraccaroli, lontana da loro per lingua e sensibilità e quindi fonte di ulteriore riflessione.

Per quanto riguarda invece il liceo classico "Galileo Galilei" di Pisa, hanno partecipato al progetto due classi terze. Di queste una era alla prima esperienza con la piattaforma Euporia, mentre l'altra, durante il precedente anno scolastico, aveva realizzato un'annotazione morfologica e sintattica dell'orazione lissiana "Per l'uccisione di Eratostene", avendo anche cura di individuare e taggare il lessico giuridico e della famiglia.

Durante l'anno scolastico 2019-2020, le due classi menzionate hanno lavorato sul testo del IX libro dell'Iliade. È stata cura degli studenti fornire una traduzione personale e annotare le caratteristiche morfologiche, sintattiche e retoriche della porzione testuale assegnata a ciascuno di loro, con particolare attenzione nell'individuazione di parole riconducibili al lessico del "dire".

In tutti i casi di lavoro sopra elencati, seppur con l'assegnazione di tasks diversi, il testo letterario scelto è stato suddiviso in modo che ciascuno studente si dedicasse a una porzione limitata di circa 20 versi o righe (cooperazione), avendo sempre la possibilità di vedere il lavoro degli altri e trarne aiuto per le proprie annotazioni (collaborazione).

## 2 CREAZIONE E USO DEI DSL

Un Domain Specific Language, come i più popolari linguaggi di programmazione e di mark-up, è definito da una Context-Free Grammar che ne descrive in modo formale la sintassi e le parole riservate. I DSL che stanno alla base di EuporiaEdu sono costituiti da due moduli principali: un modulo per il riferimento al testo da annotare e un modulo per strutturare il contenuto dell'annotazione.

Nel seguente esempio

```
☛ [14 De' Cissî ~ il balüardo] #FR: iperbato
```

la manicola segna l'inizio dell'annotazione; il riferimento al testo si trova fra quadre (notare la tilde, che indica segmenti discontinui del verso "De' Cissî hanno lasciato il balüardo"); l'hashtag #FR classifica l'annotazione come rilevazione di una figura retorica, in questo caso specificata come "iperbato".

Il modulo per strutturare il contenuto dell'annotazione viene discusso e concordato con gli studenti stessi, in modo da adottare un formalismo che sia al tempo familiare (ad esempio l'uso di hashtags per classificare le informazioni, come sui social networks), adeguatamente espressivo e privo di ambiguità. Benché il DSL produca annotazioni compatte, la quantità di informazione e la struttura risultante permette la conversione verso gli standard (ad esempio, in base alle esigenze e alle CFG create, verso XML-TEI, OWL, etc.).

### **3 PERSEGUIRE L'INCLUSIONE, POTENZIARE L'ACCESSIBILITÀ**

Il metodo Euporia, nato in ambito accademico per scopi di ricerca, nella sua declinazione a scopi didattici adotta il principio della massima inclusione, che parte proprio dalla partecipazione di tutti gli studenti anche alle fasi di definizione del linguaggio.

Nella fase iniziale di progettazione del lavoro, agli studenti viene lasciata ampia libertà: sulla base degli obiettivi didattici scelti dai docenti, sono gli studenti stessi a fare proposte sul linguaggio formale da usare nelle annotazioni. Dopo che i docenti e i membri del team Euporia hanno rivisto le proposte degli studenti e hanno reso il tagset quanto più scientifico e coerente possibile, ha inizio la fase di training, attraverso attività laboratoriali in classe, e subito dopo il lavoro di traduzione e annotazione da parte dei singoli studenti.

Non si dà dunque un formalismo calato dall'alto a cui gli allievi si devono adeguare, ma si danno molteplici livelli di strutturazione dell'informazione, che permettono a tutti gli allievi di contribuire in maniera non strutturata (nei commenti liberi), semistrutturata (commenti in linguaggio naturale con vocabolario controllato) e strutturata (annotazioni machine actionable).

È durante i cicli di revisione, cuore stesso del metodo Euporia, che gli studenti progressivamente estrapolano dalle osservazioni non strutturate informazioni che presentano patterns ricorrenti, imparando l'uno dall'altro come strutturare in modo sempre più rigoroso quanto espresso in forma discorsiva. All'inizio dell'esperienza, ad esempio, era frequente trovare osservazioni libere (OL) del tipo:

☛ [14 De' Cissî ~ il balüardo] #OL: qui il poeta usa un iperbato sintetizzato successivamente in #FR: iperbato.

La presenza di uno studente cieco in una delle classi del liceo di Pisa che ha partecipato all'iniziativa ha permesso inoltre di affrontare l'importantissimo (e troppo spesso trascurato) tema dell'accessibilità e ci ha dato l'opportunità di potenziare il DSL iniziale a vantaggio di tutti. Lo studente cieco ha svolto lo stesso lavoro assegnato agli altri (traduzione e annotazione di peculiarità morfologiche e sintattiche della porzione di testo). Ma poiché lo studente usa una tastiera braille per l'immissione dei dati, l'eccessiva presenza di segni tipografici poco familiari (quali tilde, cancelletto, etc.) usati come separatori costituiva un elemento di notevole disturbo. Lo studente stesso, nell'annotazione di analisi morfologiche richieste alle classi terze, ha proposto un sistema di annotazione leggermente più verboso ma molto più vicino al linguaggio naturale e facile da codificare in braille.



Con poche modifiche alla CFG del DSL adottato nelle terze classi del liceo di Pisa, tanto

10 κηρύκεσσι:#LEM κῆρυξ/n=araldo #MOR pl masc dat

quanto

10 κηρύκεσσι, dativo plurale maschile del sostantivo κῆρυξ: araldo

vengono rappresentati, in ultima analisi, nello stesso modo (con la preposizione articolata “dell dell” che assume funzione analoga a #LEM).

In conclusione, secondo il principio dell’inclusione, l’unica differenza risiede nel fatto che allo studente cieco è stato chiesto di usare strumenti a lui più familiari (rispetto agli hashtags), al fine di evitare un ulteriore sovraccarico cognitivo.

## 4 CONCLUSIONI

I risultati prodotti dagli studenti del Liceo “Gargallo” di Siracusa, anche se ancora in fase di revisione, sono consultabili online<sup>1</sup>. Il lavoro degli studenti di Pisa è invece ancora in corso.

In conclusione, pensiamo valga la pena menzionare che gli studenti stessi hanno riconosciuto a questa esperienza i seguenti punti di forza: (1) favorisce la collaborazione: le annotazioni possono essere viste da tutti gli studenti, che, in questo modo, possono migliorare il proprio lavoro a partire dal confronto con quello degli altri; (2) rende lo studio più coinvolgente; (3) stimola il pensiero critico e l’apprendimento “dal particolare al generale” (metodo induttivo); (4) permette lo sviluppo di competenze informatiche a partire dall’uso di Euporia; (5) favorisce il progresso nella precisione nel linguaggio scientifico grazie all’uso degli hashtags.

Fra i punti deboli gli studenti hanno giustamente segnalato il fatto che la piattaforma è ancora un prototipo con numerosi difetti tecnici da aggiustare.

Rispetto al tema dell’inclusione, è utile sottolineare nuovamente l’importanza del concetto di familiarità nella progettazione e nell’utilizzo di Euporia. Per quanto più verboso, infatti, il formalismo proposto dallo studente cieco è più vicino alle sue abitudini di lavoro e l’annotazione viene convertita automaticamente nello stesso formato dati usato dai compagni di classe. Inoltre, vale la pena aggiungere che anche un gruppo di altri studenti, più familiari con i libri a stampa che con i social networks, ha ritenuto il formalismo adottato dallo studente cieco più efficace di quello basato sull’uso di hashtags. Ne consegue che la partecipazione di uno studente cieco al progetto EuporiaEdu ci ha permesso di ampliare il campo di possibili tipi di annotazione, adattandoli alle peculiarità e alle esigenze di singoli studenti o gruppi di studenti.

---

1 <http://himeros.eu/persiani>

## 5 RINGRAZIAMENTI

La realizzazione del progetto EuporiaEdu è stata resa possibile dalla collaborazione e dalla disponibilità di docenti liceali e dirigenti scolastici. È dunque doveroso ringraziare la professoressa Michela Benedetti, la professoressa Antonella Soldani e il professor Vincenzo Muggittu e, allo stesso tempo, la professoressa Roberta Mirandola e la professoressa Greta Maneschi, che hanno seguito da vicino il lavoro dello studente cieco al liceo classico “Galileo Galilei” di Pisa.

Un sentito ringraziamento deve essere poi espresso per la dirigente scolastica del Liceo classico “Tommaso Gargallo” di Siracusa, Maria Grazia Ficara. Per la realizzazione dei manualetti di istruzioni all’uso di Euporia, infine, devono essere ringraziate la dottoressa Gloria Mugelli e la dottoressa Giulia Re.

## BIBLIOGRAFIA

1. Berti, Monica. *Digital Classical Philology: Ancient Greek and Latin in the Digital Revolution*. De Gruyter Saur, 2019. <https://www.degruyter.com/view/title/537705>.
2. Crane, Gregory, Stella Rose Dee, e Anna Krohn. «Participatory Philology: Computational Linguistics and the Future of Historical Language Education». *Human Computation* 1, n. 2 (20 dicembre 2014). <https://doi.org/10.15346/hc.v1i2.8>.
3. German Archaeological Institute, e Francesco Mambrini. «The Ancient Greek Dependency Treebank: Linguistic Annotation in a Teaching Environment». In *Digital Classics Outside the Echo-Chamber: Teaching, Knowledge Exchange & Public Engagement*, a cura di University of London, Gabriel Bodard, e Matteo Romanello, 83–99. Ubiquity Press, 2016. <https://doi.org/10.5334/bat.f>.
4. Mugelli, Gloria, Federico Boschetti, Riccardo Del Gratta, Angelo Mario Del Grosso, Fahad Khan, e Andrea Taddei. «A User-Centred Design to Annotate Ritual Facts in Ancient Greek Tragedies». *Bulletin of the Institute of Classical Studies* 59, n. 2 (2016): 103–20. <https://doi.org/https://doi.org/10.1111/j.2041-5370.2016.12041.x>.
5. Mugelli, Gloria, Giulia Re, e Andrea Taddei. «Lingue antiche e Digital Humanities: annotazione digitale dei testi tra ricerca e didattica». *Umanistica digitale* 9 (2020).

# Thinking inclusion through making

Patrizia Marti<sup>1</sup>, Annamaria Recupero<sup>2</sup>

<sup>2</sup> Università di Siena – marti [at] unisi.it

<sup>2</sup> Università di Siena – annamaria.recupero [at] unisi.it

## ABSTRACT

Today we are witnessing the growth of bottom-up initiatives, innovative projects, and wide networks of people and labs engaged in a collective response to the environmental barriers and accessibility issues pervading our society, notwithstanding the huge effort in defining standards, rules, and recommendations for human-centric inclusive design. In this scenario, Fab Labs (fabrication laboratories) and maker spaces are playing a role as innovation labs both for their skills in rapid prototyping and as participatory design spaces involving people who can bring various skills and expertise throughout the development of human-centric solutions. This paper presents two design cases developed at the Fab Lab of the University of Siena, Italy, and offers a reflection about engaging people with a disability in hands-on activities grounded on the making practice. In the cases, we adopted a particular stance on making based on co-creation and joint reflection. This approach, known in the design literature as “thinking-through-making,” shows that cross-competence, collaborative teamwork, inspiring design contexts, rapid iterative prototyping activities, and joint reflection on intermediate prototypes are quintessential to scaffold participants’ knowledge and creativity and to design innovative and inclusive solutions.

## KEYWORDS

thinking-through-making, accessibility, disability, inclusive design, digital fabrication.

## TALK

### 1 INTRODUCTION

In the past two decades, the diffusion of Do-It-Yourself (DIY) practice and digital manufacturing has emerged as a heterogeneous phenomenon with diverse actors (from amateurs to professional scientists, from local fabrication laboratories to wide online communities), methods and objectives that promote extended participation in science and innovation [2]. Such innovation in manufacturing enriches the traditional craftsmanship process offering new tools to express creativity and better compete in the market. It also makes the production process accessible for non-technical people, non-profit organizations, and educational institutions, who can be involved in distributed design networks and carry out innovative projects [3].

This phenomenon is centred in numerous fabrication laboratories (Fab Labs) existing all around the world, located in more than 100 countries, open to different stakeholders to drive collaborative bottom-up innovation<sup>1</sup>.

This is an extremely variegated and articulated network of people, skills, knowledge, and technology dealing with several challenges from health, art, education, sustainability, employability, and inclusion, to cite an incomplete list.

Accessibility is a particularly promising area of application of digital fabrication technologies since they allow personalization of assistive devices, cost reduction, and creativity of solutions. MAKEtoCARE<sup>2</sup> is a contest launched in 2016 to support the creation and growth of innovative solutions to meet the needs of people with disabilities. This initiative is promoted by Sanofi Genzyme in collaboration with Fab Labs and educational institutions. Maffei and colleagues [8] promoted a research initiative to map and investigate the numerous experiences of healthcare innovation through collaborative making practices.

For example, Makers Making Change<sup>3</sup> gathers together makers, therapists, and people with disabilities to develop affordable Open Source Assistive Technologies. Tikkun Olam Makers<sup>4</sup> is a global movement of communities, launched in 2014 to create and disseminate affordable solutions to neglected challenges of people living with disabilities, the elderly, and the poor. In 2017, it launched a hackathon for the design of accessible solutions.

e-NABLE<sup>5</sup>, Open Prosthetics Project<sup>6</sup>, Hackability<sup>7</sup>, and the other networks offer resources, skills, peer support, and crowdfunding opportunities to boost the innovation in the design for health and disability support. They also offer the opportunity to initiate a distributed production with laboratories around the world, supporting bottom-up models to innovate the healthcare system and the design of assistive devices.

In what follows, we present a research carried out at the University of Siena (Italy), where the Fab Lab facility was used as a co-design space with disabled people to run a series of different projects. Involving vulnerable people in the design process is challenging since, besides the definition of an ergonomic and functional solution, the designer has to understand the sociocultural context of use as well as the methods and spaces to guarantee an inclusive and respectful mode of participation. In the design cases presented in this paper, the Fab Lab facility was used as an inclusive co-design space to engage with the local community and allow people with diverse skills to join and contribute as equal partners in the design process. Throughout the activity, thinking-through-making proved to be a particularly successful approach in achieving equality, stimulating creativity, and joint reflection.

---

1 <https://fabfoundation.org/global-community/>

2 <https://www.maketocare.it/>

3 <https://www.makersmakingchange.com/>

4 <https://tomglobal.org/>

5 <http://enablingthefuture.org/>

6 <https://openprosthetics.org/>

7 <http://www.hackability.it/>

The design cases offer a methodological reflection about accessibility and inclusion: how can we generate knowledge that enables people with disabilities to express their expertise and creativity to play a more strategic role in service and product innovation for an inclusive society?

## 2 DESIGN CASES

Thinking-through-making is a process where the experience and skills of people are exploited to drive embodied explorations of a future artefact. In his famous book “Making: Anthropology, Archaeology, Art and Architecture,” Tim Ingold [5] explores the idea that the envisioning processes of makers and the materials they use are in a continuous correspondence coming through one another. This allows creativity to emerge from within an ongoing, improvisational process between makers, materials, tools, and the physical environment.

Materials, tools, and the physical environment play an active role in influencing the thought processes of the maker and vice-versa. The creation of objects flourishes from within complex and reciprocal relations between all these forces. This means that new artefacts are not produced from a ready-made decision taken onto the material, form, and function that the object is supposed to serve.

Rather a new artefact is generated, on the one hand, in the flows and transformations of materials and, on the other hand, in the imagination and the sensory awareness of the maker. Therefore, the act of creation is a moment where an interconnected network of relations is materialized.

Digital prototyping tools and spaces are ideal contexts to make thinking-through-making happen in a creative and participated mode. The design cases briefly described below are a testimony of such processes.

### 2.1 Smart jewels for deaf people

Quietude is an EU-funded project which provides solutions for deaf and hard of hearing people, addressing not only functional but also experiential issues like self-confidence, aesthetics, and personal sense of style in wearing accessories, including hearing aids [11]. Eight Deaf people, together with makers/engineers, designers, and Italian sign language interpreters, were engaged in a thinking-through-making process realized through various workshop sessions, where the experience of deaf participants was exploited to drive embodied explorations of future hearing aids (Figure 1). Twelve people with mild, severe, and profound deafness were then involved in a session to evaluate and critically analyse the project solutions [10]. After some iterations, the jewellery system was presented in an open exhibition targeted to the Deaf community in Siena. Around 50 people participated and tried out the jewellery, sharing comments and experiences. The system was later exhibited in public events, including Florence Biennale of Contemporary Art (Florence, 2017), Future Fest (London, 2018), Ars Electronica (Linz, 2018), Dutch Design Week (Eindhoven, 2018), and is currently displayed at the Ulm Museum at the exhibition “Transhuman From Prosthetics to Cyborg.”

The project developed aesthetically rich jewellery for deaf people, which are able to filter and recognize environmental sounds of personal interest (e.g., my doorbell, my dog) and expressively notify the wearer of their occurrence through light, vibration, and shape change. The project output is a modular system that can be personalized in terms of input (sound to recognize), sound notification (light, vibration, or shape change), and appearance of the jewellery [9]. This is to accommodate diverse users’ needs and preferences, as well as to adapt the device to specific usage situations (e.g., quiet or noisy environments).

The pieces of jewellery are produced with different materials (e.g., leather, textile), using a 3D printer and laser cutter, with custom-made printed circuit boards and electronic components (Figures 2-3).



Figure 1: thinking-through-making with deaf people

Figures 2 – 3: fabrication phases

These accessories are designed to go beyond the functional goal of supporting hearing. Rather they fulfil emotional and sociocultural needs such as aesthetics, self-expression, and personal identity. This is one of the main requirements that emerged during co-design sessions hosted at the Fab Lab.

Low-fidelity prototyping sessions were conducted using paper, cardboard, and simple fabrics to enable participants to explore possibilities without being constrained by technical limitations. The emergent expertise enabled the design team to map out the accessory space. For example, while creating probes of jewellery, a deaf participant highlighted that hearing aids are usually designed to be as discreet as possible, with some sitting deep inside the ear canal, virtually invisible. This invisibility of deafness can lead to challenges in interactions. In social situations, hearing people can forget to attend to deaf people's needs and inadvertently exclude them. This reflection suggested adopting a counter-trend design compared to that of modern hearing aids, developing very visible jewels that can communicate deafness in an elegant and sophisticated way.

## 2.2. Inclusive hand weaving loom

Re/CREATE is a project for community engagement that aims at experimenting with practices of inclusive design and creative craftsmanship. In particular, the project designed, developed, and tested a hand weaving loom tailored to people with cognitive and sensorial disabilities. Our aim in the project was not to “compensate” for disability. We rather wanted to explore new possibilities for inclusion, selfhood, and community engagement to turn disabled people into competent design partners. Through a series of weaving workshops, a commercial loom was modified and enriched with semi-automatic functionality. Thanks to rapid prototyping and the involvement of disabled participants in the iterative development, the loom was augmented with additional features (e.g., flashing LED lights used as feed-forward mechanism) to enable disabled people to creatively weave on their own without the need for support. In fact, during the weaving process, the weaver has to remember which levers have to be raised or lowered. This task is complex and frustrating for people with cognitive impairments. In order to guide and make this task easier, a band of LEDs was located under each lever to indicate the levers to lower down (Figures 4-5).



Figure 4: loom with flashing LEDs

Figures 5 – 6: a person with cognitive disability (left) and blind women weaving (right)

Moreover, a sensor was embedded to detect when the reed is beaten so that a servo motor raises all the levers and prepares the loom to repeat the task. This step proved to be necessary since disabled people often forget to raise up the levers after beating the reed and get stuck. The new automatic process of lifting the levers up was associated with augmented auditory feedback to indicate to blind users when the sequence was completed (Figure 6). In addition to the flashing LEDs, the prototype included 3D printed gates located near each lever to forbid the use of wrong levers.

The project developed through an intense phase of thinking-through-making activities resulting in the hacking of a commercial loom enhanced and transformed using digital fabrication technology. Six people with various types of disabilities, their caregivers, and weaving experts were engaged in a series of participatory design workshops focused on eliciting requirements and desires, inspiring the design process, developing and testing in highly iterative processes.

As an example of reflection emerged during a co-design session, a woman with a cognitive disability was particularly attracted by the flashing LEDs, which made her confident on when and how to perform the next step with the loom (Figure 5). Comparisons with another loom previously used emerged spontaneously. The woman remembered the embarrassment and frustration experienced when she did not remember the various steps, and this was a reason to give up.

### 3 REFLECTIONS ON MAKING FOR INCLUSION

The design cases presented above offer concrete experiences to reflect upon the importance of supporting and stimulating self-expression and creativity in people with special needs by engaging them in participatory design processes focused on thinking-through-making. We used the making practices as the basis for enabling collaborations between designers and other participants, adapting the activity to their skills and expectations. The combined use of various materials, including textiles, paper, cardboard, tape, pens, glue, and hooks and the use of digital fabrication technologies fostered the definition and transformation of prototypes, from low-fidelity probes to fully working systems. Participants were encouraged to fabricate their own personal accessory and suggest a specific solution. Ideas that emerged during the making process were materialized in the next prototyping iteration, moving from abstract and raw to concrete and structured. All participants appreciated to see their creation shaping as real devices along the process, and this encouraged them to take an active role and feel committed. It is worth mentioning that sometimes custom-made devices fabricated in Fab Labs are questioned because of their rudimentary and unpolished appearance [5]. In the cases reported above, the continuous fine-tuning of the solutions supported by product designers and the collection of participants' feedbacks throughout the process allowed to reach a good balance between functionality and aesthetics.

One of the strengths of the making process carried out in the Fab Lab was the possibility to ideate and customize the devices to meet emergent needs, as well as to support designers in adopting a holistic perspective on the user experience. The personalization of the product's form and features is particularly important in the context of a disability, considering that needs and preferences are dynamic, depend on specific situations of use and change over time.

### **3.1 Iterative development**

The making practice is iterative by nature, with several cycles of exploration, ideation, prototyping, and evaluation. Unlike a linear production process, iterative design can quickly create diverse design alternatives, experimenting with different materials and levels of fidelity, to test and refine. In thinking-through-making with disabled participants, making and thinking alternate back and forth all the time, in rapid iterations. This supports the emergence of "frugal innovation", where small and low-cost adaptations are developed in a quick and easy way. In these quick iterations, the making or designing often takes place intuitively, and the value of any small adaptation relies on reflecting on what has been made to generate new knowledge and insights.

### **3.2 Cross-competence and co-design**

The design cases presented above were carried out by multidisciplinary teams bringing diverse backgrounds, skills and points of view into the design activity. Of course, engaging people with different skills implies a suitable arrangement of the working space that hosts the co-design activities and a careful adaptation of the co-design methods.

For example, in Quietude, we arranged the workshop space ensuring proper sensory reach (visual and tactile) in 360 degrees to extending deaf people's awareness and making spatial orientation easier; arranging space and proximity to allow signers maintain enough distance to accommodate each other's signing space when conversing; maintaining clear visual communication while walking and conversing; avoiding shadow patterns and backlighting that disrupt visual communication; preventing acoustic reverberation that can be only distracting and also painful to deaf people use hearing aids or cochlear implants. All these architectural elements were accommodated to make the co-design space comfortable for deaf people. Furthermore, we experimented with different methods for engaging deaf people in the design process. However, not all of them were equally successful. For example, activities involving abstract thinking, like the one focused on reflection on ways of accessorizing the body performed in the first workshop of the Quietude project, were less engaging, while all making activities were extremely engaging, fun, and fruitful.

## **4 CONCLUSIONS**

The research described in this paper aims to answer the following question: how can we generate knowledge that enables people with disabilities to express their expertise and creativity to play a more strategic role in service and product innovation for an inclusive society? We explored this question by starting a series of projects exploiting making as a central drive of doing research and design with disabled people. Taking advantage of the expertise and technology available at our university Fab Lab, we explored thinking-through-making as a primary source for inspiration and reflection on people's experiences. People with different kinds of disabilities, from sensory to cognitive impairment, were engaged in a series of workshops holding "making" in their core.



Participants were encouraged to reflect and share their everyday experience, materialize and transform insights into design probes, and contribute with their designed artefacts to the discussion within the design team, inviting all participants to be kinaesthetically engaged and to reflect together.

The vision of the approach is closely connected to reflective practice [12] and craftsmanship, where knowledge is acquired through the exercise of craft-inspired practices throughout design and fabrication processes [13].

One of the lessons learned from this approach is that engaging disabled people in the design process through making was a response to unmet needs. By listening and interpreting people's requirements emerged during making together, we could transform them into products that can be suitable and effective and at the same time desirable. This led us to design jewellery for deaf people and a hand-weaving loom to make the creativity of disabled people flourish during the weaving process. In both projects, experiential needs like aesthetics, playfulness, gendered body design, and self-expression drove the design process from the very beginning of the design process. Both projects are still ongoing: while Quietude is more mature in terms of the number of deaf people engaged, in both cases, the participatory making practice proved to be successful, stimulating reflections and inspiring the process. The practice of rapid prototyping to frame and address the problem space, inherent in the making practice, is a valuable, constructive research method for the design science [4, 6]. Indeed, it balances creativity and serendipity that are essential for innovation, with the rigour of the design methods for experimentation and evaluation.

Adopting a user-centered approach [1], we also learned that the solution to problems lies in an open dialogue with people with disabilities before any design work has been done. By working from "the niche," more universal designs can be developed to bring accessibility to the wider market. The artefacts we managed to design appear "ordinary" to the everyday eye, but in fact, they put the needs of a disabled group of people first. Our jewels are objects of desire for everyone, and the augmented loom proved to be useful and playful also for novice weaver.

Our hope is that this approach and solutions can be brought into the mainstream, too, as helpful, beautiful, and useful for all.

## REFERENCES

1. Abras, Chadia, Diane Maloney-Krichmar, e Jenny Preece. «User-centered design». *Bainbridge, W. Encyclopedia of Human-Computer Interaction. Thousand Oaks: Sage Publications* 37, n. 4 (2004): 445–456.
2. Ferretti, Federico. «Mapping Do-It-Yourself Science». *Life Sciences, Society and Policy* 15, n. 1 (2019): 1. <https://doi.org/10.1186/s40504-018-0090-1>.
3. Gershenfeld, Neil, e Joel Cutcher-Gershenfeld. *Designing reality: how to survive and thrive in the third digital revolution*. Basic Books, 2017.
4. Hevner, Alan R. «A three cycle view of design science research». *Scandinavian journal of information systems* 19, n. 2 (2007): 4.
5. Hook, Jonathan, Sanne Verbaan, Abigail Durrant, Patrick Olivier, e Peter Wright. «A Study of the Challenges Related to DIY Assistive Technology in the Context of Children with Disabilities». In *Proceedings of the 2014 Conference on Designing Interactive Systems*, 597–606. Vancouver BC Canada: ACM, 2014. <https://doi.org/10.1145/2598510.2598530>.
6. Iivari, Juhani. «Information systems as a design science». In *Information Systems Development*, 15–27. Springer, 2005.
7. Ingold, Tim. *Making: Anthropology, archaeology, art and architecture*. Routledge, 2013.
8. Maffei, Stefano, Massimo Bianchini, Barbara Parini, e Zotti Delli. «MakeToCare. An ecosystem of actors and user-centred solutions for innovation in the healthcare sector», 2018.
9. Marti, Patrizia, Iolanda Iacono, e Michele Tittarelli. «Experiencing sound through interactive jewellery and fashion accessories». In *Congress of the International Ergonomics Association*, 1382–1391. Springer, 2018.
10. Marti, Patrizia, e Annamaria Recupero. «Is Deafness a Disability?: Designing Hearing Aids Beyond Functionality». In *Proceedings of the 2019 on Creativity and Cognition*, 133–43. San Diego CA USA: ACM, 2019. <https://doi.org/10.1145/3325480.3325491>.
11. ———. «Is Deafness a Disability?: Designing Hearing Aids Beyond Functionality». In *Proceedings of the 2019 on Creativity and Cognition*, 133–43. San Diego CA USA: ACM, 2019. <https://doi.org/10.1145/3325480.3325491>.
12. Marti, Patrizia, Michele Tittarelli, Matteo Sirizzotti, Iolanda Iacono, e Riccardo Zambon. «From Stigma to Objects of Desire: Participatory Design of Interactive Jewellery for Deaf Women». In *Interactivity, Game Creation, Design, Learning, and Innovation*, a cura di Anthony L. Brooks, Eva Brooks, e Cristina Sylla, 265:429–38. Cham: Springer International Publishing, 2019. [https://doi.org/10.1007/978-3-030-06134-0\\_46](https://doi.org/10.1007/978-3-030-06134-0_46).
13. Schön, Donald. «The reflective practitioner». *New York* 1083 (1938).
14. Sennett, Richard. *The Craftsman*. Londra: Penguin Books, 2008.



# Alessandria, un ospedale dove “la medicina è la più umana delle scienze, la più empirica delle arti e la più scientifica delle humanities”

Antonio Maconi<sup>1</sup>, Mariateresa Dacquino<sup>2</sup>, Mariasilvia Como<sup>3</sup>, Federica Viazzi<sup>4</sup>, Patrizia Santinon<sup>5</sup>

<sup>1</sup> Azienda Ospedaliera SS. Antonio e Biagio e Cesare Arrigo (Alessandria)  
amaconi [at] ospedale.al.it

<sup>2</sup> Azienda Ospedaliera SS. Antonio e Biagio e Cesare Arrigo (Alessandria)  
mdacquino [at] ospedale.al.it

<sup>3</sup> Azienda Ospedaliera SS. Antonio e Biagio e Cesare Arrigo (Alessandria)  
maria.como [at] esterni.ospedale.al.it

<sup>4</sup> Azienda Ospedaliera SS. Antonio e Biagio e Cesare Arrigo (Alessandria)  
federica.viazzi [at] esterni.ospedale.al.it

<sup>5</sup> Azienda Ospedaliera SS. Antonio e Biagio e Cesare Arrigo (Alessandria)  
psantinon [at] ospedale.al.it

## SINTESI

L’Azienda Ospedaliera SS. Antonio e Biagio e Cesare Arrigo ha una storia lunga e profondamente intrecciata a quella della città di Alessandria. Il primo documento nel quale viene citato l’istituto di cura risale al 1493, ma è nel 1579 che entra in funzione lo “Spedale Grande” dei Santi Antonio e Biagio. Da allora e sino al XXI secolo l’Azienda Ospedaliera ha perseguito l’obiettivo di offrire a tutti i cittadini cure ad alto livello di specializzazione e all’avanguardia.

In quest’ottica di assistenza a tutto tondo si colloca il Centro Studi Cura e Comunità per le Medical Humanities, costituito a maggio del 2019 con la formalizzazione di un project group multidisciplinare dedicato.

La compresenza negli stessi ambienti di Ufficio stampa e comunicazione e della Biblioteca ha creato una condivisione di saperi e una fruttuosa, reciproca, contaminazione di buone pratiche che hanno aiutato il gruppo di lavoro a strutturare progetti e attività volti al benessere di pazienti e cittadini.

Il paper ha l’obiettivo di illustrare le attività connesse alle discipline proprie delle Medical Humanities messe in atto dal Centro Studi e dall’Azienda Ospedaliera.

## PAROLE CHIAVE

medical humanities, public engagement, narrative medicine

# INTERVENTO

## 1 INTRODUZIONE

Secondo l'OMS la medicina moderna non ha solo il compito di curare la malattia, ma anche quello di conservare la salute, intesa come «stato dinamico di completo benessere fisico, mentale, sociale e spirituale, non mera assenza di malattia».

L'Azienda Ospedaliera SS. Antonio e Biagio e Cesare Arrigo ha recepito le indicazioni dell'Organizzazione Mondiale della Sanità e le ha trasposte nella pratica attraverso la costituzione di un Centro Studi dedicato alle Medical Humanities. Le Medical Humanities, si legge sulla pagina descrittiva de "L'Arco di Giano", rivista specializzata sull'argomento, hanno come mission "quella di cambiare l'immagine stessa della medicina mediante la mobilitazione di tanti e diversi saperi. La disciplina delle Medical Humanities nasce [...] dall'esigenza di arricchire gli studi nelle scienze mediche con le discipline umanistiche" e "riconduzione la pratica della sanità alle sue finalità originarie: essere medicina per l'uomo".

Il Centro Studi, la cui costituzione è stata formalizzata nel 2019, ha raccolto l'eredità di alcune esperienze pregresse già attive all'interno dell'Ospedale, come la stanza della scrittura nata nel 2016 [5], e ha sviluppato gli ambiti disciplinari delle Medical Humanities [6] incardinandone i progetti all'interno delle attività di valorizzazione del patrimonio aziendale, materiale e immateriale: fotografia, storia dell'arte e del libro, medicina narrativa, musica, diritto, architettura e storia della medicina.

La progettualità in seno al Centro Studi Medical Humanities si inserisce all'interno del contesto più ampio di una mission aziendale finalizzata a sostenere la candidatura dell'Ospedale a primo Istituto di Ricovero e Cura a Carattere Scientifico (IRCCS) per le Patologie Ambientali. Al centro di questa strategia di sviluppo che vede tra i suoi valori fondanti la partecipazione, la collegialità nell'assunzione delle decisioni che riguardano le tipologie e le modalità di erogazione delle prestazioni sanitarie, ci sono i pazienti e i loro familiari, ognuno con i suoi bisogni.

In quest'ottica di "umanizzazione" dei percorsi di cura, degli ambienti ospedalieri e della qualità della vita dei pazienti, dei loro familiari e degli operatori sanitari, è emerso con prepotenza che la cultura, attività che come poche tocca le corde più profonde dell'animo umano [1], poteva essere l'elemento che faceva la differenza per distinguere la guarigione del "corpo" da quella della "persona". Infatti, già l'essere consapevoli della storia che permea gli spazi che ci circondano, o quella delle persone che li hanno vissuti prima di noi, contribuisce a creare e rafforzare il necessario legame tra medico e paziente e, di conseguenza, allargando la prospettiva di riflessione, tra società e struttura ospedaliera.

Anche nel piano strategico aziendale 2018-2023 che prende il nome dall'acronimo OPERA (Operazione Eccellenza, Perseguire L'efficienza, Essere con il Territorio, Ricerca Come Missione), alla voce Accoglienza e Ben-Essere, tra gli obiettivi si legge quello di "Garantire ospitalità intesa come benessere, vicinanza, ascolto, accoglienza e comfort ai pazienti", e nel dettaglio della metamorfosi della cura in "ben-essere", un ruolo importante lo ricoprono la trasformazione, la valorizzazione degli spazi e il consolidamento del senso di appartenenza, tutti indicatori di qualità della vita.

Questo graduale processo di umanizzazione dell'ospedale sta prendendo forma attraverso una serie di progettualità, tutte raccolte sotto al cappello delle Medical Humanities che verranno descritte qui di seguito e che hanno portato, come primo traguardo, al festival che si è svolto online nel settembre del 2020.

La prima edizione del Festival a tema Medical Humanities intitolato “Iconografia della Salute” è stata una quattro giorni di incontri e tavole rotonde online nella quale esperti di varie discipline si sono confrontati sui concetti di cura, diagnosi e terapia nella medicina dalle origini all’era del COVID-19.

## **2 ALESSANDRIA, UN OSPEDALE DOVE “LA MEDICINA È LA PIÙ UMANA DELLE SCIENZE, LA PIÙ EMPIRICA DELLE ARTI E LA PIÙ SCIENTIFICA DELLE HUMANITIES”.**

### **MEDICINA NARRATIVA:**

Il primo progetto riconducibile all’ambito delle Medical Humanities attivato all’Ospedale di Alessandria è la stanza della scrittura che nasce nel 2016 raccogliendo l’eredità di un’attività legata alla narrazione dell’esperienza onco-ematologica del 2014 come conseguenza delle iniziative formative all’interno del progetto “Ospedale Senza Dolore”.

La stanza della scrittura si trovava all’interno del reparto di neurologia, pensata e realizzata dal clinico responsabile, dott. Luigi Ruiz e dalla sua assistente dott.ssa Delfina Ferrandi, con l’obiettivo di offrire uno spazio, intimo e personale, dove i pazienti e i loro familiari potessero elaborare ed esprimere le emozioni legate all’esperienza ospedaliera. Era un ambiente stimolante, decorato con fogli colorati e bandierine di carta manoscritte che fungevano da piccole guide, metafore e suggerimenti utili a stimolare la creatività, ma allo stesso tempo protetto, poiché aveva una sua ubicazione specifica e dedicata.

L’esito positivo della stanza della scrittura e l’ampliarsi del numero di evidenze scientifiche [3] a supporto del suo valore terapeutico [4] ha portato alla creazione di un vero e proprio concorso di medicina narrativa intitolato “Racconto la mia cura”. Nel 2019, con la terza edizione si è deciso di ampliare le forme di espressione anche al mondo delle immagini accogliendo fotografie, disegni e dipinti, oltre ai racconti e alle poesie.

Gianfranco Mogliotti, uno dei precedenti vincitori del concorso di medicina narrativa con un’opera autobiografica sulla sclerosi multipla, è stato tra i protagonisti del recente festival “Iconografia della Salute” dialogando con il maestro della Medicina Narrativa dr. Sandro Spinsanti.

Su questo tema si sono sviluppate linee di ricerca qualitativa che utilizzano il metodo dell’autoetnografia [2] analitica (nella fattispecie della versione collaborativa), primo esempio in Italia di tale metodologia, con la volontà di mettere al centro del contributo le soggettività, la posizione dei ricercatori e le loro rispettive esperienze soggettive e intersoggettive adottando nella scrittura una forma dialettica e multivocale. Con il termine autoetnografia si intende, sposando la teoria di Carolyn Ellis,: «[u]n approccio di ricerca e un modo di scrittura che intendono descrivere e analizzare sistematicamente (graphy) le esperienze personali (auto) allo scopo di [contribuire alla comprensione delle] esperienze culturali (ethno) [in cui sono collocate] [2].

### **ARCHITETTURA E FOTOGRAFIA:**

L’Azienda Ospedaliera è proprietaria di alcune tra le più importanti testimonianze dell’architettura razionalista in Italia. All’interno di un percorso di valorizzazione, promozione e conoscenza del patrimonio storico artistico del territorio, l’Azienda ha attivato una collaborazione con il Fondo Ambiente Italiano (FAI) per la creazione di percorsi a tema e organizza delle visite guidate nelle quali gli ambienti ospedalieri si aprono alla cittadinanza.

Gli spazi ospedalieri sono protagonisti anche di un progetto fotografico denominato “Hospitalia”, pensato e realizzato dall’architetto e fotografa Elena Franco: nasce come un lavoro di “documentazione che parte dalle architetture degli antichi ospedali europei - vere e proprie città nella città - per aprirsi al territorio, attraverso le donazioni che, nel corso dei secoli, i benefattori hanno devoluto per la cura di questi centri dell’“accoglienza” che attualmente ha visto realizzate due mostre, una pubblicazione e incontri all’interno delle architetture Gardelliane.

L’architetto Franco, durante il festival delle Medical Humanities, ha dialogato con Paolo Galimberti, dirigente per i beni culturali della Fondazione Cà Granda di Milano, sul valore storico e l’importanza della beneficenza ospedaliera in una sessione intitolata “L’iconografia della beneficenza. L’Ospedale Maggiore Ca’ Granda di Milano e l’Ospedale dei SS. Antonio, Biagio e Cesare Arrigo attraverso i ritratti dei benefattori”.

#### PROMOZIONE DELLA LETTURA:

L’umanizzazione delle cure passa anche attraverso la collaborazione della Biblioteca Biomedica e dell’Ufficio Comunicazione con le altre realtà cittadine, come la Biblioteca Civica, con la quale è attivo un progetto di promozione della lettura chiamato “Cultura è Cura” che, sino all’inizio del 2020 vedeva i volontari del Servizio Civile Nazionale farsi interpreti di letture a voce alta e prestito itinerante nei reparti, e che attualmente si declina con una serie di “pillole letterarie” online, liberamente accessibili sulle pagine social dell’Azienda. I ragazzi, prendendo spunto da ricorrenze o giornate nazionali, costruiscono brevi recensioni sui libri a tema medico/ospedaliero oltre a creare brevi video di “alfabetizzazione sanitaria” nei quali spiegano con parole semplici, da fonti autorevoli, concetti medici altrimenti complessi.

#### MUSICA:

La sinergia tra il Centro Studi per le Medical Humanities e il territorio si è sviluppata anche con il mondo musicale grazie alla collaborazione con il conservatorio “Vivaldi”, in particolare nell’ambito di laboratori di musicalità in Ospedale, e di interventi di musicoterapia negli spazi dell’ex Ospedale psichiatrico San Giacomo, nonché di insegnamento di pedagogia della musica per i musicisti del biennio di musicoterapia a cura dei professionisti aziendali. Si è inoltre rafforzato il legame con i promotori del Concorso di Chitarra Classica “Michele Pittaluga”, che con il progetto “Chitarre in Corsia” hanno visto gli artisti del “Pittaluga” suonare per i pazienti dell’Ospedale cittadino e delle case di riposo per anziani.

#### STORIA DELLA MEDICINA:

La Biblioteca Biomedica - Centro di Documentazione, oltre al patrimonio documentale tipico di una struttura ospedaliera con abbonamenti a banche dati e riviste scientifiche, possiede un importante fondo storico composto da quasi 1.500 volumi dei quali oltre 700 antichi, tutti catalogati in SBN nonché oggetto di studio e valorizzazione attraverso mostre, come ad esempio l’ultima allestita sul tema delle patologie ambientali che ha proposto un *excursus* storico sul legame inscindibile tra salute e ambiente, anche con pubblicazioni dedicate.

#### DIVULGAZIONE SCIENTIFICA:

La conoscenza e l'approfondimento delle Medical Humanities passa anche attraverso i canali accademici. È stato infatti organizzato, e in parte realizzato, durante la primavera 2020 un ciclo di incontri in collaborazione con l'Università del Piemonte Orientale (UPO) tenuto dalla prof.ssa Roberta Lombardi, docente di Diritto Amministrativo all'UPO e dal dott. Antonio Maconi, direttore del Centro Studi Cura e Comunità, nel quale si è parlato dello sviluppo storico e filosofico della medicina, delle sue ricadute sul diritto, sui percorsi di qualità delle cure, sulla relazione con il paziente, nonché sulla gestione e programmazione sanitaria.

A completamento delle attività che vedono il Centro Studi parte attiva e partner in progetti legati al *welfare* cittadino, la Biblioteca Biomedica - Centro di Documentazione dell'Azienda Ospedaliera di Alessandria diffonde tutte le domeniche una rassegna delle "migliori evidenze" legate al tema delle Medical Humanities. Il Centro di Documentazione, infatti, tutti i giorni invia ai professionisti aziendali una rassegna di articoli scientifici, prima legati al tema Covid, poi, rientrato parzialmente l'allarme, su argomenti d'interesse pubblico e ora (dicembre 2020) di nuovo a tema Covid.

#### FESTIVAL DELLE MEDICAL HUMANITIES:

L'organizzazione del festival delle Medical Humanities "Iconografia della Salute" è stato un modo per segnare un percorso appena iniziato ma, come si è visto, già ricco di contenuti. In quattro giorni, infatti, dal 21 al 24 settembre, sono stati discussi argomenti che integrano al loro interno la medicina per un migliore approccio al paziente e alla malattia e affrontati i temi della malattia e della cura sotto molti punti di vista (il primo giorno è stato dedicato alla rappresentazione della salute, il secondo al tema della narrazione della salute, la terza giornata ha visto l'architettura e l'arte al centro dei percorsi della salute, mentre scienza medica, diritto e salute hanno chiuso il festival).

### 3 CONCLUSIONI

Il progetto di inserimento e valorizzazione della cultura come parte integrante del processo terapeutico e di trasformazione dell'ospedale da luogo di cura del corpo a luogo di cura del corpo e dello spirito, sta prendendo forma, consci che i progetti legati alle Medical Humanities sono tra quelli che meglio favoriscono la creazione di valore all'interno del processo terapeutico. Tale valore va inteso anzitutto come un miglioramento del rapporto medico-paziente volto ad avviare una co-costruzione dei percorsi di cura che aiuti ad aumentare, sul medio-lungo periodo, il grado di alfabetizzazione sanitaria dei pazienti stessi, indicatore fondamentale del grado di benessere di una società.

Proprio in questa prospettiva si sta anche lavorando per l'organizzazione di un corso di formazione online finalizzato agli operatori di cura per l'aumento dell'empatia e la sensibilizzazione sull'importanza della cultura.

Se da una parte il Centro Studi Medical Humanities non è ancora in grado di fare bilanci su quale siano il concreto valore aggiunto e l'impatto positivo delle strategie avviate, dall'altra i feedback positivi provenienti dalle pagine social dell'Azienda Ospedaliera, insieme con l'interesse a instaurare dei rapporti di collaborazione da parte di realtà nazionali che esulano dal mondo sanitario, indicano che la strada intrapresa è quella giusta.

Parte di questa aleatorietà dei primi risultati ottenuti è forse da imputare al Covid-19.



La pandemia che identificherà questo 2020 nei libri di storia, ha obbligato tutti noi a ripensare a routine di lavoro e progettualità, il Centro Studi Medical Humanities non fa eccezione, e il gruppo di lavoro, fortemente orientato all'interazione personale e alla costruzione di un meccanismo empatico "in presenza" ha dovuto velocemente rimodulare l'offerta e ripensarsi in chiave virtuale. I primi frutti di questo cambiamento sono le già citate "pillole" di cultura letteraria e il Festival delle Medical Humanities online, in prospettiva, quando l'allarme sarà completamente rientrato, il mondo virtuale non verrà abbandonato ma diventerà complementare a quello reale, attraverso, ad esempio, la creazione di contenuti dedicati a specifiche categorie svantaggiate, o la possibilità di creare dei tour virtuali alle mostre così da raggiungere un pubblico potenzialmente infinito.

## BIBLIOGRAFIA

1. Bolaki, Stella. «Introduction: Artists' Books and Medical Humanities». *Journal of Medical Humanities* 41, n. 1 (2020): 1–5.
2. Ellis, Carolyn, Tony E. Adams, e Arthur P. Bochner. «Autoethnography: an overview». *Historical social research/Historische sozialforschung* 36 4, n. 138 (2011): 273–290.
3. Ofri, Danielle. *Cosa dice il malato, cosa sente il medico*. Roma: Il pensiero scientifico, 2018.
4. Reilly, Jo Marie. «Storytelling as a vehicle of healing». *Families, Systems, & Health* 38, n. 3 (2020): 330–33. <https://doi.org/doiLanding?doi=10.1037/fsh0000502>.
5. Ruiz, Luigi, Delfina Ferrandi, T. Bolgeo, G. Montobbio, A. Pepoli, e A. Barbierato. «La stanza della scrittura: scrivere aiuta la guarigione». *Working Paper of Public Health* 14 (2016).
6. Wigand, Moritz E., Thomas Becker, e Florian Steger. «Psychosocial Reactions to Plagues in the Cultural History of Medicine: A Medical Humanities Approach». *Journal of Nervous & Mental Disease* 208, n. 6 (2020): 443–44. <https://doi.org/10.1097/NMD.0000000000001200>.

## ULTERIORI RIFERIMENTI

Concorso di chitarra classica “Michele Pittaluga”

<https://www.pittaluga.org/concorso-internazionale-di-chitarra-classica-storia.html>

Concorso di medicina narrativa “Racconto la mia cura”

<https://www.ospedale.al.it/concorso-medicina-narrativa-racconto-la-mia-cura/>

Hospitalia

<http://elenafranco.it/home/hospitalia/>

Iconografia della salute: festival delle medical humanities (21-24 settembre)

<https://www.ospedale.al.it/iconografia-della-salute-festival-delle-medical-humanities-dal-21-al-24-settembre/>

Iconografia della Salute: quattro giorni di riflessione sulle Medical Humanities

<http://curaecomunita.it/2020/09/27/iconografia-della-salute-quattro-giorni-di-riflessione-sulle-medical-humanities/>

Le medical humanities

<http://www.arcodigiano.org/medical-humanities>

Organizzazione Mondiale della Sanità

<http://www.salute.gov.it/portale/rapportiInternazionali/dettaglioContenutiRapportiInternazionali.jsp?lingua=italiano&id=1784&area=rapporti&menu=mondiale>

Salute (Wikipedia)

<https://it.wikipedia.org/wiki/Salute>

Seminario sulle Medical Humanities

<https://www.digspes.uniupo.it/eventi/seminario-sulle-medical-humanities>



# **Sessione III**

## **Rita Levi Montalcini**



# Analyzing the Sensor Data Stream for Monitoring and Visualization of Early Autism Signs (MoVEAS)

Mariasole Bondioli<sup>1</sup>, Michele Zoncheddu<sup>1</sup>, Stefano Chessa<sup>1</sup>, Antonio Narzisi<sup>2</sup>, Susanna Pelagatti<sup>3</sup>

<sup>1</sup> Università di Pisa

<sup>2</sup> IRCCS Stella Maris (Calambrone)

<sup>3</sup> Università di Pisa – susanna.pelagatti [at] unipi.it

## ABSTRACT

Observing children affected by autistic spectrum disorders while they play is an important and widely used diagnostic method that is conducted in a clinic under the supervision of a specialist. In this work, we present a prototype of a sensorized toy that can classify its own movements (as they are applied by children), to allow observation of the play of the children even without the direct supervision of the specialist. This would allow to perform such tests even in environments more familiar to the children (at home, at school), in a non-invasive way.

## KEYWORDS

e-health, internet of things, sensors, autistic spectrum disorders, smart toys

## TALK

### 1 INTRODUCTION

MoVEAS (acronym of Monitoring and Visualization of Early Autism Signs) is a project that aims to detect, in a non-invasive way, early signs of autistic spectrum disorders (ASD), by monitoring play activities in young children through sensorized toys and classifying the activities through a neural network [4, 12]. Observing children while playing is a widespread technique for diagnosing ASD, especially in early age [16], and videotaped play session of groups of children with typical development against groups of children with ASD, have been proven to be useful for the diagnosis [3, 16, 20]. Children with ASD use the toys in atypical and restricted ways (e.g., spinning), play with repetitive behaviors, and visually explore objects in unusual ways [6, 16]. Diagnosing ASD can be difficult since there are no simple medical tests to diagnose the disorders, and the spectrum width makes it even harder.

This project's aim is to simplify the specialists' diagnosis process by identifying ASD-related behaviors in non-clinical settings (home, school...) and allowing to indirectly monitor a play session.

In particular, we present the MoVEAS project and discuss its activity recognition component based on machine learning that is aimed at classifying the movements applied to the toy by the children. We discuss the selection of the relevant data to classify the movements, the analysis of the consistency of the data obtained, and the performance of the final model in a real scenario simulation.

## **2 THE MOVEAS PROJECT**

The observation of children while they play is commonly carried out at the clinic with a standard set of toys. This setting for the observation is not ideal in that the environment and toys used are totally new for the child. The idea of MoVEAS is to embed sensors in the toys used in standard evaluation sessions in order to let the clinicians collect data more freely, even at home or at school. In this way, they can monitor the play changes when toys become more familiar to the child.

In this perspective, the system reference scenario is based on the interaction steps between the toys moved by children, the server collecting and elaborating the data of the children's actions, and the user interface, where the server gives the data back to the user (usually the clinician). The steps are the following: (1) The user accesses his/her private interface with a normal browser; (2) To start a new data collection session, the user selects a patient and the related toys and presses the "start" button; (3) The corresponding devices automatically acquire from the sensors (and send to the server) the data of the children's movement during the play session; (4) The sensor detection module detects the shift of the toy, elaborates the data and recognizes a specific movement; (5) The server communicates the information to the application web pages, where the data are visualized and analyzed for the diagnosis goals. During this process, the server provides for punctual data storage in its database, constantly receptive to the entry information from the software of the motion capture devices. The server also returns the processed information to the user through a web user interface. Through this interface, a user can also perform all management operations like create/remove a new patient in/from the database, connect a patient with a toy, start/end a new capture session, etc. This recruitment of children for this study was conducted according to the Helsinki Declaration.

## **3 THE ACTIVITY RECOGNITION COMPONENT**

The activity recognition component operates in two steps. The first step is a data fusion algorithm (Fig. 1) that extracts from the 3-axis accelerometer, gyroscope, and magnetometer time series two information for the user, which are the estimation of the orientation of the and the estimate of the intensity and direction of the movements of the toy. The two information in output are produced for each new tuple of data samples in input and hence form two time series. These time series, along with the raw data time series, are in turn input to the second step based on neural networks that classify the time series as recognized movement of the toy.

For what concerns the data fusion algorithm, note that a correct estimation of the orientation and movement is a rather complex task. Each sensor took individually (either accelerometer, gyroscope, or magnetometer) is insufficient to compute either the orientation or the movement of the toy for several reasons.

The 3-axis accelerometer measurements have limitations in the detection of position in dynamic conditions, and it is also impossible to estimate the orientation of the horizontal plane of the azimuth with reference to Earth's gravity (to which this sensor is subject). The estimate of the orientation can be improved by using the gyroscope, which, however, suffers from drifts in its measurements. To overcome the limits of the gyroscope, we thus also use a magnetometer (whose measurements refer to a global reference system given by the Earth's magnetic field). The estimate of the toy orientation (in quaternion representation) is performed by fusing all these data by means of the Madgwick algorithm [13]. The estimation of the movement is obtained from the accelerometer data, from which is detracted of the contribution of the Earth's gravity along each axis. In turn, this contribution is estimated on the basis of the orientation of the toy computed by the Madgwick algorithm. To make this estimation more stable, the accelerometer data is pre-processed with a standard Kalman filter. Both data of orientation and movement estimation are stored in a JSON record structured in two parts. The first contains the quaternion representation of the orientation, the raw acceleration along the three axes, and the estimation of the power of the acceleration signal (this is used in the 3D real-time visualization), and the second part contains the orientation of the toy expressed in the Euclidean angles of roll, pitch and yaw, and the movement estimation in terms of magnitude and angles of direction (this is used for the storage and for the visualization of the movement diagrams).

The second step of classification of the movements of the toy analyses the time series of raw and pre-processed data by finding specific patterns. Not all patterns need the same recording time: for example, the duration of the movement of the toy pushed forward is much smaller than the one where the toy is thrown against a pillow – where the toy is thrown by the hand first, then flies in the air and finally hits the pillow. Our goal is to be able to recognize many different variable-length sequences. This accomplished using a neural network.

In our work, we analysed two solutions. The first one is a time-delay neural network (TDNN), chosen for its simplicity (the training sequences are short, 14 samples, but a complex model might easily go to overfitting); the second one is a recurrent neural network (RNN), chosen for its ability to recognize short sequences. For our purposes, both networks have been trained with the same dataset of fixed-length sequences. With TDNN, the classification is simply made by a window of size 7 that slides over the entire session's data. For model training and validation, the dataset was composed of: 100 patterns with the toy going forward and 100 with the toy going backward; 200 patterns of simulated flight; 200 patterns with the still toy; 200 patterns of the toy carried while walking; 200 patterns of the toy thrown against a pillow. The training and validation portions were respectively 60% and 40% of the total, and the entire dataset is shuffled before every training. The test set was composed of 240 samples, disjoint from the training and validation ones. For what concerns the overfitting detection, this was based on the accuracy on the validation set: when the increment of accuracy on training did not match an increment on validation, the model was discarded.



For both the TDNN and RNN has been chosen the model with the highest validation accuracy: more precisely, the model selection for the TDNN has been performed with a grid search on the number of filters and size of the kernel, while for the RNN, the main hyper-parameter was the number of units of the single recurrent layer.

In TDNN, model selection started from the simplest topology: an input layer with data normalization, a hidden convolutional layer with a ReLU activation function that performed better than that, a max-pooling layer that showed better results than the average-pooling one, and finally, an output layer for the pattern classification, with a softmax function, to obtain values between 0 and 1, representing the probability of the classified pattern to belong to each possible class.

Only six features were considered relevant for the training: the gravity-free accelerations values in the three axes and the raw gyroscope data – the gyroscope data is used to discriminate some cases that are not well distinguished only by the acceleration. Some tests showed quickly that adding more features, like orientation, velocity data, or raw sensors data, made the network generally harder to train and most often gave no advantage at all in the overall accuracy.

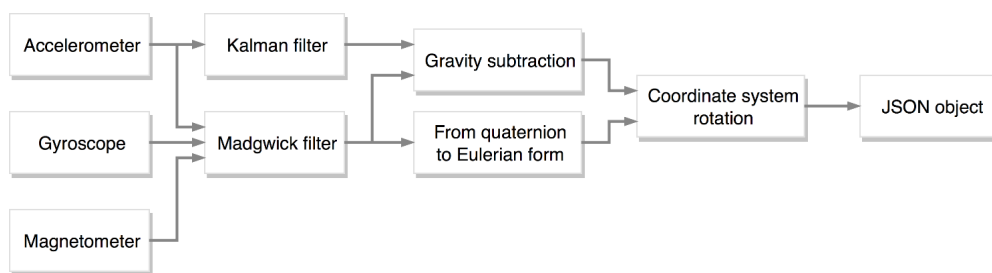


Figure 1: The data fusion schema

The search for the two most important hyperparameters, the kernel size and the number of filters, has been performed through a grid search. For each table's entry, the network has been tested ten times, with shuffled data between training and validation set for each run. Outliers have been omitted from the average. The values 9 and 7 were optimal respectively for the number of filters and the kernel size: they gave the best result in the validation set, and in particular:

- a number of filters up to 7 and a kernel size up to 5, made the network able to classify correctly only after a long and unstable training;
- from 7 filters upwards, the training curve was stable, and the overfitting occurred only from 15;
- kernel sizes from 9 upwards tended to overfit the network: the patterns were initially made of 22 samples, and the sliding window was too big;
- with filters bigger than 20, the overfitting was mitigable only with very small kernel sizes, but in that case, the network was hard to train and very easy to underfit.

More complex topologies were tried, but adding hidden layers only led the neural network to the overfitting. The best training curve was obtained using a 16-patterns long mini-batch.

The next step was to achieve maximum accuracy with less possible data. The network was trained several times with different patterns' lengths, and the best choice for stable and accurate training has proved to be 14 samples per pattern, that with the 22 Hertz sampling means 636 milliseconds long patterns. Up to 11-13 samples, the network was not able to achieve good performance in training, being hard to train and underfitting, while from 17 samples and more, the network fitted the noise, so overfitted. The convolutional layer's weights are initialized randomly in the uniform range  $[-0.000001, 0.000001]$ .

To keep under control the network complexity, a weight decay approach has been used, with tested values in the range  $[0.001, 0.05]$ . The optimization algorithm chosen for the learning was Adam because of its efficiency [11]. As done with the TDNN, the input passes through a normalization layer before going to the recurrent layer, and the activation function is ReLU for all layers, except for the last one, which uses the softmax function, particularly suitable for classification.

The best topology that ensured a good fitting without instabilities during the training is shown in the following schema: one recurrent layer and two densely connected layers, with two dropout layers before them.

In particular, the number of nodes in the recurrent layer was crucial for the model selection:

- up to 5 units, the network was able neither to learn the training samples nor generalize;
- up to 10 units, the validation accuracy reached 90%, but when the training continued, it overfitted;
- up to 12 units, the network was hard to train, and with 14 units, the network was trained smoothly, and the validation accuracy grew up to 93.5%;
- from 16 upwards, stricter regularization was needed, but the dropout layers made the training stable and avoided the overfitting. In the end, the best number of units was discovered to be 20.

The explored range of dropout for both layers was  $[0, 0.5]$ . The network was trained with a mini-batch size of 16 patterns.

Adding features besides the gravity-free acceleration and the gyroscope data gave most often no advantage at all, and sometimes made the training harder and unstable, also with more layers and more recurrent units. In this model, reducing the number of samples has not improved the validation accuracy as for the TDNN, but the performance significantly decreased only for less than 14 samples per pattern, according to the results achieved with the TDNN.

## 4 EXPERIMENTAL RESULTS

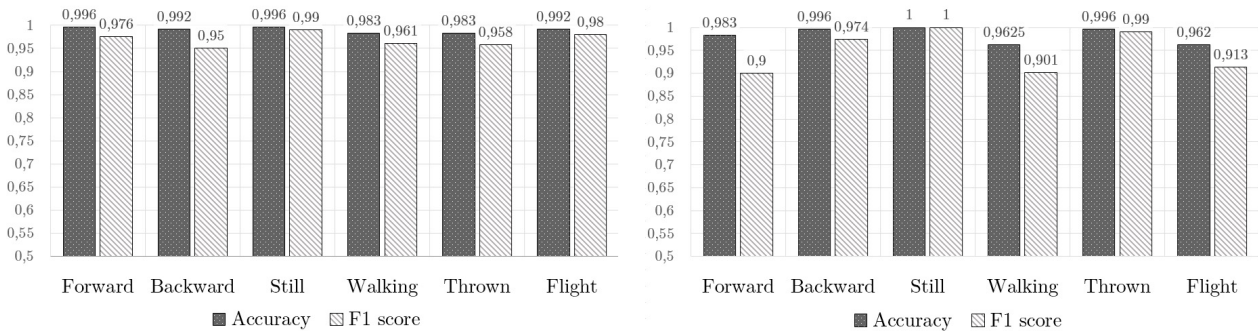


Figure 2 – 3: accuracy and F1 score for a) TDNN b) RNN

The results obtained on the dataset are shown in Figures 2 and 3. We observe that TDNN provides a better validation accuracy and a smaller loss value. Moreover, almost every test showed smoother training and validation curves, and it was easier to train effectively.

The high accuracy obtained in validation and testing might not reflect the real network’s performances in analysing real play sessions: in the first two cases, the patterns have a well-defined start and end, while in a play session, there is a continuous stream of data; for the classification, the network is given a fixed length subsection of that stream for each 22nd of a second, and the subsection’s bounds slide forward, like a sliding window, through the whole stream. When subsection contains only a single pattern, the networks behave exactly like in the validation/testing scenario, but when there is an overlapping between two patterns (e.g., the toy stops flying and stays still), the classification is harder because that’s not the case that the network was trained to deal with. For this reason, we performed a further experiment with data from a 46 seconds-long session of simulated play. The network was able to accurately recognize the still, forward, and backward patterns while in flight and especially walking, so the overall classification is correct, with some very short misclassifications. We plan to investigate this issue in detail with real play session recordings.

## 5 RELATED WORK

Previous research has established the connection between the way in which children interact with objects and the potential early identification of children with ASD. Those findings motivate our own work to develop “smart toys,” objects embedded with wireless sensors that are safe and enjoyable for very small children, that allow detailed interaction data to be easily recorded. The ways in which infants and toddlers play with objects can be indicative of their developmental progress. Depending on their age, a child’s object play activities can display simple physical milestones, such as placing objects in their mouth to sophisticated cognitive tasks such as symbolically using a banana as a telephone receiver. In fact, observing the subtleties surrounding the way in which infants play with toys may highlight early indicators for developmental delays, such as autism. To better quantify how play may serve as an early indicator for autism, researchers have conducted studies examining the differences in object play behaviours among infants.

The work in [14] demonstrated that lateral glances toward objects were significantly more common in a group of children with autism (mean age 44 months) than in children with typical development. [5] found that children with autism (mean age 33 months) who spent a large proportion of time in restricted object use showed poorer joint attention, imitation, and social engagement abilities. Finally, [19] reported group differences in repetitive movements with objects (defined as ‘at least three consecutive’). From a technological point of view, there has been a growing interest in using technology to ease several aspects of diagnosis and treatment of ASD people [7]. Automatic detection of stereotypical movements in children with autism using sensors has been investigated by Goodwin et al. [9]. Their experiment involved six children, and each child wore three wireless accelerometers, one on the left wrist, one on the right wrist, and one around the chest, with no restriction in movement. Sessions were taken at school or during therapy in situations that usually triggered stereotypical behaviour. Each session was video recorded, and specialists annotated them both online and offline using annotation software. A classifier algorithm was trained to automatically detect repetitive behaviours from the accelerometer readings with good results. Plotz et al. [17] used a commercially available device with a 3D accelerometer and a microcontroller combined with machine learning techniques to detect and classify anomalous behaviours. Other approaches [8, 10] used Kinect to recognize stereotypical movements. In [10], the Kinect camera was used to film 12 actors performing three separate stereotypical motor movements each.

Then, software recognizing these gestures was developed and instructed using a machine learning approach. Manual grading was used to confirm the validity and reliability of the software.

Our approach differs from the above ones since we aim at designing a device that is not invasive (does not require wearable devices or cameras) and can be used even in private houses without the supervision of specialists. To this purpose, our design exploits state of the art IoT architecture, which is not present in the above works. In this sense, we can highlight how several combinations of the three sensors introduced in our work have already been largely exploited in IoT literature regarding Human Activity Recognition. In this field, in fact, we can find work as that of Najafi et al.s [18], in which the authors presents a new method of physical activity monitoring based on a 2D accelerometer and a gyroscope fused in sensor systems to be put on the front of the chest, which is able to detect body postures (sitting, standing, and lying) and periods of walking in elderly persons using only one kinematic sensor attached to the chest.

Moreover, we highlight a wide number of human Activity recognition researches based on a single accelerometer system, as those in which the researchers attempted to distinguish walking on level ground from walking on a stairway using waist acceleration signals [15] and the studies focused on the design of an algorithm for analysing and classifying human activity (as standing, sitting down with lowering subjects head, sitting down and leaning against, lying down straight, lying upside down, walking, going up/downstairs, running) using a body-fixed triaxial accelerometer on the back [2]. The accelerometer is, in fact, in literature defined as a sensor that shows promise in providing an inexpensive but effective means of long-term ambulatory monitoring (in most of the case treating elderly patients) [1].

## 6 CONCLUSIONS

We discussed the MoVEAS project focusing on its new component of activity recognition. In particular, we discussed the choice of the relevant data to classify the movements, the consistency of that data over time, the effective training and comparison between two neural network models, and the performance of the deployed model in a real scenario simulation. Despite the good results achieved in this experimental phase, only working with children with ASD can reveal the actual project's usefulness. In this perspective, the next experimental phase is already planned, with a sample of already diagnosed ASD children and a control group of neurotypical children. This phase will also have the objective of devising and assessing a suitable diagnostic protocol that may make use of MoVEAS.

In the future, we will also address some current limitations of the software. For the movement classification, it might be useful to “flatten” the small spikes that affect the continuity of classification to obtain compact chunks to highlight with different colours in the session's scrollbar. Another path to be explored is to train the recurrent neural network with a variable-length patterns training set and to find the best way to choose the length of the session's subsequence dynamically to feed the network with, in order to exploit the full network's capabilities.

## REFERENCES

1. Abran, Alain, Adel Khelifi, Witold Suryn, and Ahmed Seffah. 'Usability Meanings and Interpretations in ISO Standards'. *Software Quality Journal* 11, no. 4 (2003): 325–338.
2. Albinali, Fahd, Matthew S. Goodwin, and Stephen S. Intille. 'Recognizing Stereotypical Motor Movements in the Laboratory and Classroom: A Case Study with Children on the Autism Spectrum'. In *Proceedings of the 11th International Conference on Ubiquitous Computing*, 71–80, 2009.
3. Baranek, Grace T., Carolyn R. Barnett, Erin M. Adams, Nancy A. Wolcott, Linda R. Watson, and Elizabeth R. Crais. 'Object Play in Infants with Autism: Methodological Issues in Retrospective Video Analysis'. *American Journal of Occupational Therapy* 59, no. 1 (2005): 20–30.
4. Bondioli, Mariasole, Stefano Chessa, Antonio Narzisi, Susanna Pelagatti, and Dario Piotrowicz. 'Capturing Play Activities of Young Children to Detect Autism Red Flags'. In *Ambient Intelligence – Software and Applications – 10th International Symposium on Ambient Intelligence*, edited by Paulo Novais, Jaime Lloret, Pablo Chamoso, Davide Carneiro, Elena Navarro, and Sigeru Omatu, 1006:71–79. Cham: Springer International Publishing, 2020. [https://doi.org/10.1007/978-3-030-24097-4\\_9](https://doi.org/10.1007/978-3-030-24097-4_9).
5. Britto, Talita, and E. Pizzolato. 'Towards Web Accessibility Guidelines of Interaction and Interface Design for People with Autism Spectrum Disorder'. In *ACHI 2016: The Ninth International Conference on Advances in Computer-Human Interactions*, 1–7, 2016.
6. Bruckner, Cornelia Taylor, and Paul Yoder. 'Restricted Object Use in Young Children with Autism: Definition and Construct Validity'. *Autism* 11, no. 2 (2007): 161–171.
7. Buzzi, Maria Claudia, Marina Buzzi, Beatrice Rapisarda, Caterina Senette, and Maurizio Tesconi. 'Teaching Low-Functioning Autistic Children: ABCD SW'. In *European Conference on Technology Enhanced Learning*, 43–56. Springer, 2013.
8. Gonçalves, Nuno, Sandra Costa, José Rodrigues, and Filomena Soares. 'Detection of Stereotyped Hand Flapping Movements in Autistic Children Using the Kinect Sensor: A Case Study'. In *2014 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC)*, 212–216. IEEE, 2014.
9. Gonçalves, Nuno, José L. Rodrigues, Sandra Costa, and Filomena Soares. 'Automatic Detection of Stereotyped Hand Flapping Movements: Two Different Approaches'. In *2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication*, 392–397. IEEE, 2012.
10. Josman, Naomi, Hadass Milika Ben-Chaim, Shula Friedrich, and Patrice L. Tamar Weiss. 'Effectiveness of Virtual Reality for Teaching Street-Crossing Skills to Children and Adolescents with Autism'. *International Journal on Disability and Human Development*, 2008.
11. Kingma, Diederik P., and Jimmy Ba. 'Adam: A Method for Stochastic Optimization'. *ArXiv Preprint ArXiv:1412.6980*, 2014.
12. Lanini, Marco, Mariasole Bondioli, Antonio Narzisi, Susanna Pelagatti, and Stefano Chessa. 'Sensorized Toys to Identify the Early “Red Flags” of Autistic Spectrum Disorders in Preschoolers'. In *International Symposium on Ambient Intelligence*, 190–198. Springer, 2018.

13. Madgwick, Sebastian OH, Andrew JL Harrison, and Ravi Vaidyanathan. 'Estimation of IMU and MARG Orientation Using a Gradient Descent Algorithm'. In *2011 IEEE International Conference on Rehabilitation Robotics*, 1–7. IEEE, 2011.
14. Mitchell, Peter, Sarah Parsons, and Anne Leonard. 'Using Virtual Environments for Teaching Social Understanding to 6 Adolescents with Autistic Spectrum Disorders'. *Journal of Autism and Developmental Disorders* 37, no. 3 (2007): 589–600.
15. Monibi, Mohamad, and Gillian R. Hayes. 'Mocotos: Mobile Communications Tools for Children with Special Needs'. In *Proceedings of the 7th International Conference on Interaction Design and Children*, 121–124, 2008.
16. Ozonoff, Sally, Suzanne Macari, Gregory S. Young, Stacy Goldring, Meagan Thompson, and Sally J. Rogers. 'Atypical Object Exploration at 12 Months of Age Is Associated with Autism in a Prospective Sample'. *Autism* 12, no. 5 (2008): 457–472.
17. Pilebro, Carin, and Birgitta Bäckman. 'Teaching Oral Hygiene to Children with Autism'. *International Journal of Paediatric Dentistry* 15, no. 1 (2005): 1–9.
18. Sekine, Masaki, Toshiyo Tamura, Tatsuo Togawa, and Yasuhiro Fukui. 'Classification of Waist-Acceleration Signals in a Continuous Walking Record'. *Medical Engineering & Physics* 22, no. 4 (2000): 285–291.
19. Warren, Zachary E., Zhi Zheng, Amy R. Swanson, Esubalew Bekele, Lian Zhang, Julie A. Crittendon, Amy F. Weitlauf, and Nilanjan Sarkar. 'Can Robotic Interaction Improve Joint Attention Skills?' *Journal of Autism and Developmental Disorders* 45, no. 11 (2015): 3726–3734.
20. Wetherby, Amy M., Juliann Woods, Lori Allen, Julie Cleary, Holly Dickinson, and Catherine Lord. 'Early Indicators of Autism Spectrum Disorders in the Second Year of Life'. *Journal of Autism and Developmental Disorders* 34, no. 5 (2004): 473–493. <https://doi.org/10.1007/s10803-004-2544-y>.

# Towards the early detection of the red flags of dysorthography in non-diagnostic settings

Chiara Pugliese<sup>1</sup>, Susanna Pelagatti<sup>2</sup>

<sup>1</sup> Università di Pisa - c.pugliese1 [at] studenti.unipi.it

<sup>2</sup> Università di Pisa - susanna.pelagatti [at] unipi.it

## ABSTRACT

Dysorthography is a specific disorder of spelling written words, for example, confusing similar letters and putting letters or syllables in the wrong order. In Italy, recent studies estimate on average 30.711 children with dysorthography. This is a very large number, and an early detection becomes crucial to allow teachers and family to be aware of the difficulty encountered by a child in daily school activities and to start to address the problem. The goal of our project is to sustain tutors (teachers and parents) in this difficult path. This is achieved by giving them clear suggestions on the need of a professional evaluation through the analysis of some written text. We encourage them to pursue this path with a simple, friendly app which can be used in a natural environment, such as home and school. In this paper, we present the preliminary work for the participatory design of such an application. First, we identified a focus group to understand the most important red flags of dysorthography. The group included parents and teachers recruited in specific Facebook groups and professionals from Specific Learning Disabilities Centres. Then, we built a dataset from real handwritten texts by children with dysorthography. The paper details both phases outlining the lesson learned and the strategy chosen for the main problems encountered.

## KEYWORDS

dysorthography, deep learning, handwriting recognition

## TALK

### 1 INTRODUCTION

Dysorthography is one of the Specific Learning Disabilities (SLD) and concerns errors in spelling level of written words, such as confusion between similar letters and with the order of letters or syllables. In Italy, recent studies estimate a prevalence of 20.579 children with dysorthography aged 7-10, 36.177 children aged 11-13 and 35.378 children aged 14-18 [11].



Early detection of dysorthography is very important to allow the school and the family to be aware of the difficulty encountered in daily school activities and to start to address the problem. Nevertheless, many children with dysorthography and other SLD are not recognized and diagnosed until late primary school.

This is due to the psychological and practical problems of spotting early red flags at home and in the classroom. When this happens, the children have a poor outcome, but they are considered lazy. Very often, the teachers say something like “he/she is smart but does not work enough...” which worsens the situation since it adds some serious self-confidence issues to the SLD problem [5].

Sometimes the parents and teachers have some doubts but are simply worried to go through a proper process of diagnosis until they are really convinced that the problem might be real. A process of official diagnosis is often perceived as an admission of diversity which can be dangerous for the future child life [2]. Although the number of diagnoses is rising [11], this work [2] shows that many students with SLD remain without a diagnosis. The study was conducted on a group of 9.964 Italian children aged between 8 and 10, of which only 126 (1.3% of the total) had already a formal diagnosis of SLD. During the study, several standardized tests were administered to the group. After the survey, the percentage of children with SLD had risen to 3.5%.

The goal of our project is to sustain families and teachers in this difficult path, trying to give them clear suggestions on the need of a professional evaluation through the simple analysis of some written text. Though the real diagnosis can only be made by a professional team, our goal is to encourage family and teachers to pursue this path. This is possible using a simple, friendly app which can be used in a natural environment, such as home and school. We will discuss some related work in Section 2; however, to the best of our knowledge, this is the first attempt to address this specific problem. Our project follows the three-phase model proposed by Hevner [8]: an iterative process actively involving the stakeholder and the SLD specialists. We try to understand which are the red flags which need to be taken into account and how to recognize them. In this paper, we will discuss the first stage of our project discussing the work done with a focus group, including SLD people, parents, teachers, and doctors. The aim of this work has been to study people’s needs, to explore the experiences of those who received a late diagnosis, to understand how parents and teachers noticed and discovered their children’s disorders, and to understand the best way to sustain them in this difficult path. This has been carried on through the usual methods, including questionnaires, interviews, and dataset analysis. From the results obtained, we start to outline the needs and the strategies to be included in the future prototype app.

The paper is organized as follows. Section 2. gives a brief overview of state of the art. Section 3. describes our research method. Section 4. outlines the main ideas for the future app and gives details of our focus groups and of the results obtained. Section 5. concludes.

## **2 STATE OF THE ART**

In the literature, some applications support diagnosis of dysgraphia, another SLD concerning written level. These works are software created to support doctors in making diagnoses and to monitor patients with this SLD. Their aims are to reduce costs in term of time and to monitor written texts in an objective manner. For example, *TestGraphia* [6] is an Italian software system based on known document analysis algorithms.

Starting from a digital image of written text, this software carries out *offline* handwriting detection. It extracts some features, called *static*, following the BHK (Concise Evaluation Scale for Children's Handwriting) Evaluation Protocol<sup>1</sup>, such as left margin text alignment, writing size, etc.

The offline approach starts from a digital image, processes it, and detects text area [12]. After feature extraction, it computes a score and returns a report. In this way, doctors can read the report and modify it when the results are not correct. In addition, the meantime to execute a diagnosis is nearly ten times faster with this kind of tool than without. In this way, doctors can monitor more than one patient in a short time.

Papers [1], [14] and [10] are based on *online* handwriting detection that requires technological support, such as a consumer tablet [12]. These works add the possibility to extract more than *static* features. With the online approach, it is possible to extract *dynamic* features such as pressure and tilt of the pen. Dynamic and static features together help doctors to make a more accurate diagnosis and are important to identify other disorders such as dyspraxia.

All these works are created to support diagnosis, so the disorder was already identified by parents or teachers. Instead, in this paper, we present a tool able to detect the red flags of dysorthography when this disorder is not already known. So, the aim of our project is not to help doctors and support the diagnosis. It is useful to make known SLD as much as possible and to encourage early diagnosis. In the end - while to detect the dysgraphia it is not necessary to recognize text, but only monitor its dimensions, alignment, etc. - for our work the handwriting recognition is the key for the red flags detection.

### 3 RESEARCH METHOD

First, we identified a focus group - with the help of Facebook groups - to understand the red flags of dysorthography. The focus group is composed of 230 people, including 12 people with SLD, 19 parents of children without SLD and 199 tutors of children with SLD. We have designed an interview with which we have investigated the kind of red flags and at what age they have arisen. Results confirmed the scientific data: the most common errors are confusion between similar phonemes (such as p/b, f/v, t/d), omission, insertion or inversion of letters or syllables (*palora* instead of *parola*) and omission of double letters or <h> in verb (*mama* instead of *mamma*) [7]. Instead, about the age, 24.78% of participants identified the red flags from age 8 and up (11,74% after age 10). This means that in 57 cases out of 230, the diagnosis was made later than the recommended age (during the third grade [5]).

The next step was to contact the clinicians of the IRCCS Stella Maris Hospital and the educators of the DAB-LAB Centre for SLD to collect data of children with dysorthography.

---

1 A standardized method to detect dysgraphia.

We managed to collect 181 texts based on standardized tests and written by 60 authors (Table 1).

| Grade level                  | Number of authors | Grade level                         | Number of authors |
|------------------------------|-------------------|-------------------------------------|-------------------|
| 1 <sup>st</sup> grade school | 0                 | 1 <sup>st</sup> secondary school    | 8                 |
| 2 <sup>nd</sup> grade school | 3                 | 2 <sup>nd</sup> secondary school    | 1                 |
| 3 <sup>rd</sup> grade school | 17                | 3 <sup>rd</sup> secondary school    | 4                 |
| 4 <sup>th</sup> grade school | 9                 | 1 <sup>st</sup> high school         | 2                 |
| 5 <sup>th</sup> grade school | 13                | from 2 <sup>nd</sup> high school on | 3                 |

Table 1: classification of handwritten text authors based on grade level.

Once collected data and answers to the interview, it was possible to define problems to reach our aim. We can define three problems to be faced. The first one is to detect handwriting text starting from digital images and to segment these images in characters. Before proceeding with segmentation, it was helpful to preprocess images with some actions such as noise reduction, binarization, etc. For preprocessing and word segmentation, we used built-in functions of OpenCV<sup>2</sup>. Instead, for character segmentation, we implemented algorithms following [3] and [13] because built-in functions cannot detect and segment every character in words. With these algorithms, we detected pixel intensities of each column of images, white spaces between characters (in uppercase) and ligatures in cursive texts (examples in Figure 1).

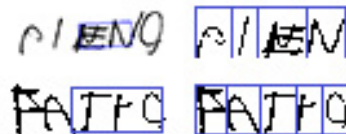


Figure 1: Left: image segmentation with built-in functions; right: image segmentation with our implemented algorithms. Blue rectangles show detected area.

The second problem is to proceed with the offline handwriting recognition of segmented characters. The last one is to detect typical errors in recognized texts. For both handwriting recognition and red flags detection, we adopt a machine learning approach, such as most works in the literature [9]. We are experimenting with different deep learning models such as Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN), in particular Bidirectional Long-Short Term Memory (Bi-LSTM). Handwriting recognition models are trained with the EMNIST dataset [4], a set of 814.255 handwritten characters - in lower and upper case - and digits, written by 500 writers. Instead, red flags detection models are trained with our collected data to which we are adding texts without errors and other texts containing errors generated automatically, in order to obtain a larger dataset.

2 <https://opencv.org/>

## 4 OUTLINING A DYSORTHOGRAPHY APP

Clinicians had some doubt about this project because it is impossible to say if a child has one or more SLD starting from any written text. When they make a diagnosis, they administer a set of standardized tests based on the child's age. When there is a number of errors above a certain threshold, they can make the diagnosis. But, after some meetings, they defined our work, such as an activity of "pre-screening". So, our project will be useful to trigger doubts in tutors and to start monitoring children before the diagnosis.

In addition, after reading the focus group's answers, we realized that parents want to be understood, first of all. They do not want to be alone during the diagnosis path. They claim more attention on SLD by schools. Although they had doubted and did not understand very well how our project works, they were happy and grateful for our attention to SLD. For these reasons, the future app will be called Dis-: not a problem and will be designed as follow. First, there will be a section dedicated to users. Here there will be a guide for using the app properly.

It must be clear that this app does not make a diagnosis, but it does a sort of "pre-screening" activity. In this section, we will also answer the user's questions: what are SLD, how the app works and why they should use it. The core of the app will be a section where users upload images of child's handwritten text. In order to make the handwriting recognition more accurate, in this section, we will ask users if the text was written in upper or lower case. The output will be a "score" calculated considering the number of red flags, compared to the total number of words. Therefore, in the back-end, once the image is loaded, it will be pre-processed and segmented (such as described in Section 3.) in characters. Every character will be the input of the handwriting recognition neural network. Accordingly, the output of this neural network will be the input of the red flags detection neural network. The final output will be the desired score. Concerning the front-end part, it should be as usable as possible: we will design it in according with indications of the focus group and main web design guidelines for people with SLD (resume in Table 2).

| <b>"Problem"</b>                                      | <b>Solution</b>                          |
|---|--|
| Confusion between similar characters (p/b, m/n, etc.) | Use OpenDyslexic <sup>3</sup> font       |
| Difficulty to read lists                              | Use numbered lists                       |
| Difficulty to read with white background              | Use pastel colours                       |
| Difficulty to read large texts                        | Use two columns                          |
| Difficulty to keep attention                          | Use short sentences and emphasized texts |
| Difficulty in understanding                           | Insertion of explanatory images          |

---

3 <https://www.opendyslexic.org/>

Table 2: SLD people's issues arising from the focus group and adopted solution, according web design guidelines for people with SLD.



Figure 2: the main sections of Dis-: not a problem mock-up. Left: section where there is some information on how to make and to upload photo in the right way. Right: section where the app returns plain text corresponding to digital image and output “score”.

## 5 CONCLUSIONS

Collecting the focus group’s experiences, we have read many stories. In some cases, especially when the diagnosis was late, despite parents had had an intuition at an early age, the disorder was underestimated by teachers. In other cases, children’s self-confidence was very low, or they had problems to remain seated on a chair. In extreme cases, children were so frustrated to be led to the act of self-harm. Many parents complained about the poor preparation of teachers on the subject or their superficial attitude. Some other families complained about the very long time taken by the diagnosis process. Instead, in some cases, there was good cooperation between parents and teachers. In Italy, this is very important because the path of a diagnosis cannot start until teachers finish a six-month monitoring period. In this period, teachers try to understand if there is a persistent problem or if the child has fallen behind, compared to the rest of the class. However, most part of the focus group had not a good experience.

So, this research and app want to be a way to reach “lost children”. If this application is used in a natural environment such as school or home, more children with dysorthography, but without a tutor expert in this disorder, will be helped in time. In addition, our aims are to make this problem known as much as possible and to encourage tutors to take a path with professional teams. Our choice to use off-line approach for handwriting recognition was driven by our desire to not force users to necessarily buy an expensive tablet.

In order to improve our work, we will continue to collect data, which are very important to train neural networks correctly. Moreover, it would be useful to implement a personal area where each user could monitor every text analysed, with the corresponding score. In this way, it is possible to monitor children's performance and to observe their improvements.

In the end, a valid improvement could be to extend functions of Dis-: not a problem to the other SLD. We have decided to work with the dysorthography first for some reasons. The most important is because this disorder - together with the dyscalculia - is less known than dyslexia and dysgraphia and because there are already works related to dysgraphia.

## **6 ACKNOWLEDGEMENTS**

We thank Dr.s Paola Brovedani, Filippo Gasperini and Paola Cristofani of the IRCCS Stella Maris Hospital in Calambrone (Pisa) and the DAB-LAB Center for SLD in Pisa for the help provided in setting up the dataset and the Facebook groups DSA Roma – Dislessia, disortografia, disgrafia, discalculia; W LA DISLESSIA!; Ritardo o disturbo del linguaggio in età evolutiva; Genitori di BIMBI con ritardi e disturbi di linguaggio e/o motori; Dislessia e DSA: aiuto consigli e testimonianze; DISLESSIA – PASSODOPOPASSO; Dislessia, Disortografia e Discalculia. Diagnosi e Trattamento (Sicilia); DIS-tranoi DSA INSIEME PER CAPIRE dislessia, disortografia, discalculia; Dislessici adulti; Disturbo del Linguaggio, Spettro Autistico, Mutismo Selettivo o affini; Disgrafia e disortografia; DISLESSIA – PISA; Aiuto Dislessia and Dislessici Universitari for their participation such as focus group.

## REFERENCES

1. Asselborn, Thibault, Thomas Gargot, \Lukasz Kidziński, Wafa Johal, David Cohen, Caroline Jolly, and Pierre Dillenbourg. 'Automated Human-Level Diagnosis of Dysgraphia Using a Consumer Tablet'. *NPJ Digital Medicine* 1, no. 1 (2018): 1–9. <https://doi.org/https://doi.org/10.1038/s41746-018-0049-x>.
2. Barbiero, Chiara, Marcella Montico, Isabella Lonciari, Lorenzo Monasta, Roberta Penge, Claudio Vio, Patrizio Emanuele Tressoldi, Marco Carrozzi, Anna De Petris, and Anna Giulia De Cagno. 'The Lost Children: The Underdiagnosis of Dyslexia in Italy. A Cross-Sectional National Study'. *PloS One* 14, no. 1 (2019): e0210448. <https://doi.org/https://doi.org/10.1371/journal.pone.0210448>.
3. Choudhary, Amit, Rahul Rishi, and Savita Ahlawat. 'A New Character Segmentation Approach for Off-Line Cursive Handwritten Words'. *Procedia Computer Science* 17 (2013): 88–95.
4. Cohen, Gregory, Saeed Afshar, Jonathan Tapson, and Andre Van Schaik. 'EMNIST: Extending MNIST to Handwritten Letters'. In *2017 International Joint Conference on Neural Networks (IJCNN)*, 2921–2926. IEEE, 2017. <https://doi.org/https://doi.org/10.1109/IJCNN.2017.7966217>.
5. Consensus Conferences. *Disturbi Specifici Dell'Apprendimento*. Roma: Ministero della Salute e Istituto Superiore di Sanità, 2011.
6. Dimauro, Giovanni, Vitoantonio Bevilacqua, Lucio Colizzi, and Davide Di Pierro. 'Test-Graphia, a Software System for the Early Diagnosis of Dysgraphia'. *IEEE Access* 8 (2020): 19564–19575. <https://doi.org/10.1109/ACCESS.2020.2968367>.
7. Fenzi, Virginia, and Cesare Cornoldi. 'Le Difficoltà Ortografiche Di Adolescenti Con Dislessia'. *Dislessia* 12, no. 1 (2015).
8. Hevner, Alan R. 'A Three Cycle View of Design Science Research'. *Scandinavian Journal of Information Systems* 19, no. 2 (2007): 4.
9. Kacalak, Wojciech, and Maciej Majewsky. 'Handwriting Recognition Methods Using Artificial Neural Network'. *Proceedings of the Artificial Neural Networks in Engineering (ANNIE)* 16 (2006).
10. Mekyska, Jiri, Marcos Faundez-Zanuy, Zdenek Mzourek, Zoltan Galaz, Zdenek Smekal, and Sara Rosenblum. 'Identification and Rating of Developmental Dysgraphia by Handwriting Analysis'. *IEEE Transactions on Human-Machine Systems* 47, no. 2 (2016): 235–248. <https://doi.org/10.1109/THMS.2016.2586605>.
11. MIUR. *I Principali Dati Relativi Agli Alunni Con DSA*. Roma: Presidenza del Consiglio dei Ministri, 2019.
12. Plamondon, Réjean, and Sargur N. Srihari. 'Online and Off-Line Handwriting Recognition: A Comprehensive Survey'. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22, no. 1 (2000): 63–84. <https://doi.org/https://doi.org/10.1109/34.824821>.
13. Rehman, Amjad, Dzulkifli Mohamad, and Ghazali Sulong. 'Implicit vs Explicit Based Script Segmentation and Recognition: A Performance Comparison on Benchmark Database'. *Int. J. Open Problems Compt. Math* 2, no. 3 (2009): 352–364.

14. Rosenblum, Sara, and Gideon Dror. 'Identifying Developmental Dysgraphia Characteristics Utilizing Handwriting Classification Methods'. *IEEE Transactions on Human-Machine Systems* 47, no. 2 (2016): 293–298. <https://doi.org/https://doi.org/10.1109/THMS.2016.2628799>.
15. Santana, Vagner Figueredo de, Rosimeire de Oliveira, Leonelo Dell Anhol Almeida, and Maria Cecília Calani Baranauskas. 'Web Accessibility and People with Dyslexia: A Survey on Techniques and Guidelines'. In *Proceedings of the International Cross-Disciplinary Conference on Web Accessibility*, 1–9, 2012. <https://doi.org/https://doi.org/10.1145/2207016.2207047>.



# La tecnologia come strumento per ridurre l'ansia dal dentista nei bambini con autismo

Antonio Narzisi<sup>1</sup>, Mariasole Bondioli<sup>2</sup>, Francesca Pardossi<sup>3</sup>, Lucia Billeci<sup>4</sup>, Maria Claudia Buzzi<sup>5</sup>, Marina Buzzi<sup>6</sup>, Martina Pinzino<sup>7</sup>, Caterina Senette<sup>8</sup>, Valentina Semucci<sup>9</sup>, Alessandro Tonacci<sup>10</sup>, Fabio Uscidda<sup>11</sup>, Benedetta Vagelli<sup>12</sup>, Maria Rita Giuca<sup>13</sup>, Susanna Pelagatti<sup>14</sup>

1 IRCCS Stella Maris Foundation – anarzisi [at] inpe.unipi.it

2 Università di Pisa – mariasole.bondioli [at] unipi.it

3 Ospedale Universitario di Pisa – f.pardossi [at] yahoo.it

4 IFC-CNR – lucia.billeci [at] ifc.cnr.it

5 IIT-CNR – claudia.buzzi [at] iit.cnr.it

6 IIT-CNR – marina.buzzi [at] iit.cnr.it

7 IN-CNR – martinapinzino [at] outlook.it

8 IIT-CNR – caterina.senette [at] iit.cnr.it

9 UFSMIA Livorno (USL Toscana Nord Ovest) – valentina.semucci [at] uslnordovest.toscana.it

10 IFC-CNR

11 Università di Pisa

12 UFSMIA Pisa (USL Toscana Nord Ovest) – benedetta.vagelli [at] uslnordovest.toscana.it

13 Ospedale Universitario di Pisa; Università di Pisa

14 Università di Pisa; Associazione Autismo Pisa – susanna.pelagatti [at] unipi.it

## SINTESI

Gli strumenti informatici possono contribuire a migliorare significativamente la vita delle persone con disabilità. Il disturbo dello spettro dell'autismo può comportare ansia e paura in situazioni nuove o molto stimolanti dal punto di vista percettivo-sensoriale. Un problema comune è la difficoltà nell'effettuare visite e ricevere cure mediche. Questo articolo descrive l'esperienza effettuata nell'ambito del progetto MyDentist dedicato alla riduzione dell'ansia nella odontoiatria speciale infantile.

## PAROLE CHIAVE

ICT, web app, autismo, odontoiatria speciale, stress

# INTERVENTO

## 1 INTRODUZIONE

Il disturbo dello spettro autistico (ASD) è un grave disturbo multifattoriale caratterizzato da sintomatologia specifica nelle aree della comunicazione sociale, interessi limitati e comportamenti ripetitivi [1]. L'incidenza dell'ASD è in crescita negli ultimi anni, si stima in 1 individuo su 54 negli Stati Uniti e 1 su 87 in Italia [7, 9]. Sebbene non ci siano specifiche relazioni tra ASD e malattie del cavo orale, molti individui con ASD hanno una salute orale molto peggiore rispetto alla popolazione generale [5].

Questo può essere spiegato sia con l'esistenza di barriere nell'accesso alla cura, sia con l'esistenza di fattori che minano il grado di collaborazione dei pazienti con ASD rendendo difficile la cura, quali ipersensibilità sensoriale e livelli elevati di stress e ansia durante l'assistenza sanitaria [12].

L'idea del progetto MyDentist è nata per rispondere alle esigenze concrete di famiglie con bambini affetti da disturbo dello spettro autistico in merito alla gestione delle cure dentali, che spesso risultano molto difficili da poter essere somministrate a causa delle difficoltà legate a questo tipo di disturbo. Eventi apparentemente insignificanti, come un cambiamento nella routine quotidiana, un incontro inaspettato con un conoscente, la difficoltà nell'interpretare le aspettative altrui, il disagio nel comunicare i propri pensieri e stati d'animo, possano portare i bambini con ASD a sperimentare livelli molto elevati di stress che si manifestano nei contesti più disparati e interferiscono con le attività quotidiane a casa, a scuola, dal medico, ecc. Nelle strutture sanitarie, la gestione di pazienti con ASD è in generale una sfida difficile, perché l'ansia causata dall'interazione inusuale con il personale medico/infermieristico può portare ad atteggiamenti imprevedibili, esagerati e negativi che possono sfociare talvolta in comportamenti dirompenti o aggressivi quali colpire violentemente o mordere oggetti o persone. In particolare, nel caso delle cure dentali, l'ansia nel paziente ASD è ulteriormente acuita dalle intense sensazioni fisico/tattili e acustiche causate dalle procedure mediche. La conseguenza è che, spesso, per portare avanti le necessarie procedure e proteggere l'operatore, nei casi di comportamenti aggressivi, anche le più semplici cure dentali richiedono l'anestesia totale del paziente.

D'altra parte, l'uso di percorsi di adattamento graduali, anche autogestiti, può permettere, anche alle persone con ASD, di ambientarsi nelle situazioni nuove. Il progetto MyDentist adotta questo approccio, focalizzandosi sulla riduzione dell'ansia nei bambini con ASD nel contesto delle cure dentali tramite un protocollo di familiarizzazione e personalizzazione della gestione del percorso clinico che fa uso di strumenti ICT (Information & Communication Technologies). Alcuni studi recenti forniscono il necessario supporto preliminare per l'uso di interventi potenziati dalla tecnologia per migliorare l'igiene orale dei bambini con autismo [6, 11].

Concretamente, il progetto MyDentist offre un ambiente virtuale (Web App) che viene arricchito visita dopo visita con materiali multimediali (video, audio, foto, ecc.) spesso registrati dallo stesso bambino/a durante la seduta oppure creati a partire da risorse preesistenti a cura del dentista. L'ambiente permette la catalogazione delle esperienze del bambino/a (relative alla visita medica) e la loro rielaborazione in tempi successivi, per esempio nella preparazione della prossima visita in programma.

Per favorire la costruzione di percorsi di familiarizzazione odontoiatrica, MyDentist offre inoltre la possibilità di creare giochi multimediali personalizzati e l'assegnazione di compiti specifici al paziente a casa, come ad esempio esercizi per migliorare l'igiene orale da farsi seguendo video modelli.

## 2 PROGETTAZIONE PARTECIPATIVA

La progettazione partecipativa [8] è un processo di progettazione del software e delle interfacce utente che chiama in causa tutti gli attori parte del dominio. Essa consente di fruire in modo sinergico di idee, punti di vista, metodologie di progettazione e analisi differenti, in ambiti multidisciplinari e con il coinvolgimento degli utenti finali (o dei loro caregiver). Questa metodologia mira a massimizzare l'accessibilità e l'usabilità del sistema creato e deve essere applicata tenendo conto delle specifiche del contesto d'uso ambientale e del contesto lato utente.

La letteratura recente ci informa come sia possibile e doveroso oggi fare progettazione partecipativa anche in contesti delicati in cui gli utenti sono persone con disabilità specifiche [2, 4].

Il progetto MyDentist ha utilizzato la progettazione partecipativa attraverso una collaborazione interdisciplinare che, sotto la guida del Dipartimento di Informatica, ha coinvolto principalmente psicologi dell'IRCCS Fondazione Stella Maris, ricercatori dell'Istituto di Informatica e Telematica del Consiglio Nazionale delle Ricerche, clinici dell'Ambulatorio di Odontoiatria Infantile dell'Azienda Ospedaliera Universitaria Pisana, le logopediste della ASL Nord Ovest area Pisana e Livornese e volontari dell'Associazione Autismo Pisa (APS), attiva nel Pisano in Valdera e Valdicecina.

## 3 LO STRUMENTO ICT: L'APP

L'applicazione risultato del progetto è una RIA (Rich Internet Application) scritta in PHP, AJAX e HTML5 con architettura dati di tipo relazionale (MySQL). Tutte le funzioni del software sono implementate utilizzando jQuery e le librerie JSON, che consentono la creazione di astrazioni per interazioni di basso livello, effetti avanzati e widget di vario tipo. La tecnologia AJAX permette lo scambio di dati efficiente e facile, quindi le interfacce utente del client HTML5 sono costantemente aggiornati. L'applicazione funziona come un grande archivio, un "Toolkit" dove i materiali sono organizzati in due gruppi principali: 1) materiali multimediali preconfezionati, adatti allo specifico contesto e disponibili per tutti i bambini secondo necessità; 2) materiali personalizzati per ciascun paziente (creati dinamicamente) e costituiti da file multimediali raccolti durante ogni visita e classificati per data di visita. È previsto l'accesso a 2 aree separate:

- Area Dentista: dove il dentista crea il profilo personalizzato di ogni paziente al fine di programmare una visita, costruire un percorso clinico personalizzato, raccogliere materiali multimediali durante le visite, creare giochi personalizzati, condividere risorse e assegnare compiti a casa.
- Area Paziente: dove ogni paziente può accedere al proprio percorso personalizzato fatto di giochi che anticipano le diverse procedure dentali (es: sequenze, puzzle ecc.); video modeling (ad esempio che mostrano come usare uno spazzolino da denti), file audio per familiarizzare con i suoni della clinica odontoiatrica, storie sociali che introducono l'ambiente clinico, foto e video raccolti durante le visite.

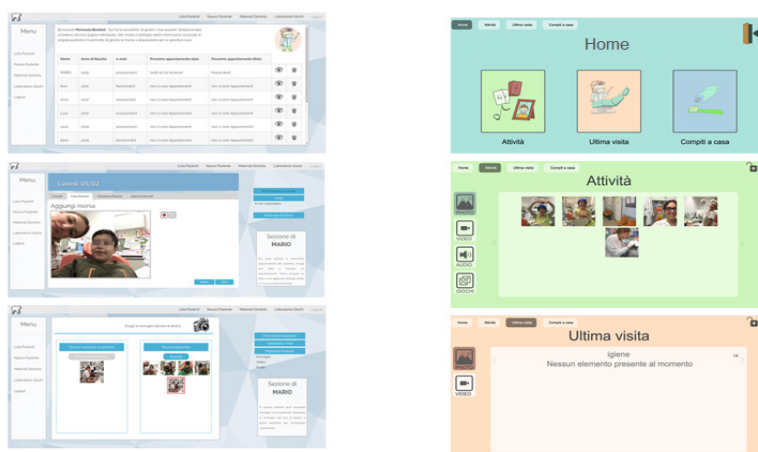


Figura 1: interfacce Utente dell'Area Dentista (a sinistra) e dell'Area Paziente (a destra)

## 4 DEFINIZIONE DEL PROTOCOLLO

Ruoli e procedure dell'approccio MyDentist sono stati definiti grazie alle diverse competenze presenti nel team multidisciplinare e hanno portato all'identificazione dei componenti chiave di un protocollo allo stesso tempo rigoroso e flessibile in grado di garantire la riproducibilità dell'intervento sia in altri ambulatori odontoiatrici sia in altri domini medici. Nello specifico, fa parte del protocollo:

- 1 Identificare le figure coinvolte, dentista, assistente, mediatore ICT. Ogni visita con il bambino/a deve garantire la presenza di queste figure, evitando presenze extra che possono disturbare.
- 2 Procedere ad una adeguata caratterizzazione dei bambini con autismo coinvolti al fine di poter garantire un intervento mirato e personalizzato. Da questo punto di vista occorre conoscere dati e abitudini del bambino/a sia in termini di pratiche odontoiatriche sia in termini di familiarità con l'uso dei dispositivi elettronici (tablet, smartphone) oltre che conoscere le sue preferenze da usare come elementi di rinforzo e ricompensa.
- 3 Identificare il dettaglio delle procedure mediche che possono fare parte dell'approccio e quelle invece da escludere e schedare opportunamente le attività distinguendo la fase preliminare (prima visita), di conoscenza reciproca (dentista-paziente), da quelle successive, anche queste modulate secondo i tempi del bambino.
- 4 Identificare i materiali da utilizzare per popolare l'applicazione che rispettino i vincoli del dominio e che siano personalizzabili sulle esigenze del paziente.

In seguito ad una lunga fase di verifica e affinamento del sistema informatico [3], volto a ottimizzare l'uso dello stesso a supporto delle procedure parte del protocollo (acquisizione dati utente, raccolta materiali delle visite, programmazione e schedulazione attività ecc.) questo strumento è stato inquadrato ufficialmente in un protocollo medico sperimentale adottato presso l'azienda Ospedaliera Universitaria Pisana, Ambulatorio di Odontoiatria Infantile (responsabile del protocollo la Prof. M. Rita Giuca) che è attualmente frequentato da circa 200 bambini con ASD provenienti principalmente dall'area pisana e livornese ma anche da altre zone della Toscana e da fuori regione. L'accesso al laboratorio è completamente gratuito e avviene attraverso una email dedicata che permette l'inserimento in una lista di attesa specifica. Gli appuntamenti vengono forniti in orari nei quali l'ambulatorio è scarsamente frequentato e le famiglie sono supportate da volontari della locale associazione di genitori Autismo Pisa.

## 5 LA SPERIMENTAZIONE

L'introduzione dell'approccio MyDentist in ambulatorio ha seguito due fasi. Nella prima fase sono stati coinvolti un gruppo di 10 bambini per un periodo di circa 3 mesi durante i quali il team ha lavorato alla progettazione dell'applicazione applicando il modello partecipativo in ambiente reale. Un ricercatore (informatico-umanista nel ruolo di mediatore ICT) ha preso parte con costanza alle visite mediche osservando le dinamiche dell'ambulatorio, introducendo gradualmente alcuni strumenti digitali definiti in precedenza con gli altri membri del team (e personalizzati sul paziente) e annotando le risposte (interazione utente-sistema-ambiente) ottenute osservando i giovani pazienti.

In ognuna di queste sedute il bambino poteva agire da utente passivo o attivo degli strumenti proposti, avendo a disposizione un tablet per accedere a diverse risorse multimediali o per registrare l'ambiente e/o se stesso durante la visita attraverso video o foto in modalità selfie.

I vari contenuti sono stati volta per volta salvati su un server e messi a disposizione dei ragazzi per essere (ri) visionati a casa. In questa fase è stata valutata la fattibilità dell'approccio (introdurre elementi extra in ambiente medico è spesso una sfida) e l'accettabilità da parte dei ragazzi destinatari dell'intervento. Questa fase è servita anche per valutare la bontà del protocollo apportando le eventuali modifiche quando necessarie. Per queste ragioni, anche queste visite sono state strutturate in modo da distinguere un momento di conoscenza (prima visita), con interviste alla famiglia e introduzione dell'ambiente al ragazzo, dagli appuntamenti successivi comprendenti ognuno:

- Attività ICT da fare a casa: giochi di apprendimento, revisione di foto e sessioni di video modeling con materiali raccolti nella visita precedente, per prepararsi alla visita successiva.
- Attività medica in clinica: diversa per ogni bambino a seconda delle sue necessità mediche.
- Attività ICT presso la clinica: a seconda del successo delle procedure mediche.

I risultati osservazionali ottenuti in questa fase esplorativa hanno confermato il potenziale della tecnologia nel ridurre l'ansia da visita odontoiatrica e aumentare il benessere e la sicurezza dei bambini, rendendo le procedure più piacevoli e incoraggiando l'igiene orale come parte della loro routine quotidiana. Inoltre è emersa ancora più chiaramente la necessità di ottimizzare la produzione e fruizione dei materiali digitali progettando un ambiente integrato in grado di ridurre i tempi di raccolta, accesso e fruizione e soprattutto rendere minimo l'impegno da parte del dentista e del mediatore ICT.

L'app ha preso così forma per affinamenti successivi dando risposta alle diverse necessità del dominio emerse sul campo.

Confermata la fattibilità dell'approccio, è stata avviata una seconda fase di sperimentazione della durata di 6 mesi circa coinvolgendo tutti i pazienti con autismo che hanno chiesto accesso all'ambulatorio dedicato. A 130 ragazzi con autismo è stato offerto un percorso di cura personalizzato e supportato dall'applicazione. Per una adeguata valutazione di efficacia dell'intervento (in termini di riduzione dello stress e ansia da visita), anche se il percorso è stato offerto a tutti, è stato necessario selezionare un sottogruppo di pazienti che, per specifiche esigenze mediche e logistiche, garantivano una continuità di intervento tale da permettere una raccolta dati puntuale e rigorosa. Alla sperimentazione hanno quindi partecipato 59 bambini (età media 9,9 anni; SD = 5.43). L'intervento ha integrato le normali procedure di cura dentale con attività per la desensibilizzazione e il controllo dello stato ansioso, fornite attraverso l'applicazione e volte ad agevolare i processi di familiarizzazione del bambino con autismo con l'ambiente medico e con le sue procedure.

Le valutazioni sono state fatte sulla base di due questionari somministrati alle famiglie al tempo T0 (primo accesso alla clinica) e al tempo T1 (dopo 6 mesi di sperimentazione). I questionari sono stati definiti dal team di ricerca ed entrambi sono stati testati prima su un sottogruppo di 10 genitori. Due esperti, entrambi con decennale esperienza in ambito autismo, hanno esaminato i questionari per la convalida della forma e del contenuto. Nello specifico, un primo questionario (questionario A) è stato somministrato al tempo T0 per indagare l'esperienza nelle pratiche dentali del singolo paziente in condizioni base. Ogni elemento del questionario identificava uno specifico comportamento e l'intervistato (il genitore del ragazzo con autismo) doveva rispondere con una scala likert a 4 punti (0=mai; 1= qualche volta; 2= spesso; 3= regolarmente).

Dopo 6 mesi è stato proposto un secondo questionario (questionario B) comprendente sia la parte rappresentata dal questionario A sia una parte addizionale volta a indagare l'esperienza del paziente con l'approccio mediato dalla tecnologia.

Quest'ultimo comprendeva 16 elementi descrittivi un orientamento positivo/negativo del ragazzo verso il supporto tecnologico. In questo caso il genitore aveva a disposizione una scala likert a 4 punti (0 = no, 1 = poco, 2 = abbastanza, 3 = molto).

## **6 RISULTATI**

I risultati della sperimentazione hanno evidenziato differenze significative nell'approccio alla cura dentale tra T0 (prima di iniziare la sperimentazione) a T1 (dopo 6 mesi dall'inizio della sperimentazione) in almeno 14 dei 18 aspetti indagati dal questionario somministrato ai genitori [10]. Le famiglie hanno valutato positivamente l'uso del supporto multimediale riconoscendo che lo stesso abbia facilitato la collaborazione del bambino nelle procedure mediche riducendo lo stress durante la visita. Secondo le risposte dei genitori, l'uso di supporti multimediali ha reso i bambini più familiari con l'igiene dentale e più collaborativi con l'ambiente in generale. Inoltre i problemi sensoriali e i comportamenti problema sono apparsi più gestibili per quasi tutti i bambini. L'uso dei supporti multimediali ha aiutato i bambini anche a trasferire le competenze acquisite al di fuori dello studio odontoiatrico e ha aumentato la loro autonomia nelle procedure per l'igiene dentale a casa. Solo il 30% dei genitori ha considerato il supporto multimediale iper-stimolante per il proprio bambino.

## 7 DISCUSSIONE E CONCLUSIONI

I risultati positivi ottenuti dai questionari e le informazioni registrate durante la sperimentazione, derivanti dall'osservazione dei giovani pazienti, mostrano come sia possibile l'accesso alla cura, anche per i bambini con autismo, riducendo gli eventi stressanti e traumatici e rendendo il bambino soggetto attivo del percorso di cura. Lo strumento digitale, nelle varie forme descritte, ha mostrato di poter essere un valido supporto per decifrare la complessità ambientale e rendere il contesto più comprensibile al bambino e dunque più accettato. Lo studio descritto ha chiaramente alcuni limiti fra i quali, la mancata rigidità nella scelta del campione, l'assenza di un gruppo di controllo con assegnazione randomizzata, la mancata raccolta di dati pre e post intervento sul grado di stress parentale; la mancanza di un protocollo o albero decisionale per l'utilizzo dell'applicazione, il protocollo prima descritto ha riguardato tutte le fasi dell'approccio ma non il supporto tecnologico nello specifico. Fra i punti di forza dell'approccio descritto occorre ribadire la gratuità del percorso di cura, considerato che l'ambulatorio coinvolto fa parte del servizio sanitario nazionale (SSN), e il fatto che lo strumento software di supporto è completamente gratuito. Inoltre è possibile prevedere che un approccio simile possa essere usato con efficacia anche in altri contesti. A questo fine anche la piattaforma necessita di essere maggiormente flessibile; per questo la sua architettura è in corso di revisione a favore di un sistema modulare basato sul Model-View-Controller (MVC) implementato con Ruby On Rails.

Complessivamente l'esperienza nell'ambito del progetto MyDentist dedicato alla riduzione dell'ansia nella odontoiatria speciale infantile ha dato risultati positivi e incoraggianti, tanto che al momento è in fase di valutazione l'estensione della metodologia ad altri ambiti medici. Gli strumenti ICT sviluppati sono resi disponibili gratuitamente per altri ambulatori pubblici e privati interessati a replicare il protocollo.

## BIBLIOGRAFIA

1. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. Washington: APA, 2013.
2. Benton, Laura, and Hilary Johnson. ‘Structured Approaches to Participatory Design for Children: Can Targeting the Needs of Children with Autism Provide Benefits for a Broader Child Population?’ *Instructional Science* 42, no. 1 (2014): 47–65.
3. Bondioli, Mariasole, Susanna Pelagatti, Maria Claudia Buzzi, Marina Buzzi, and Caterina Senette. ‘ICT to Aid Dental Care of Children with Autism’. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*, 321–322, 2017.
4. Brosnan, Mark, Sarah Parsons, Judith Good, and Nicola Yuill. ‘How Can Participatory Design Inform the Design and Development of Innovative Technologies for Autistic Communities?’ Edited by Nigel Newbutt. *Journal of Assistive Technologies* 10, no. 2 (20 June 2016): 115–20. <https://doi.org/10.1108/JAT-12-2015-0033>.
5. Du, Rennan Yanlin, Cynthia KY Yiu, and Nigel M. King. ‘Oral Health Behaviours of Preschool Children with Autism Spectrum Disorders and Their Barriers to Dental Care’. *Journal of Autism and Developmental Disorders* 49, no. 2 (2019): 453–459.
6. Grewal, Navneet, Tanvi Sethi, and Sukrit Grewal. ‘Widening Horizons through Alternative and Augmentative Communication Systems for Managing Children with Special Health Care Needs in a Pediatric Dental Setup’. *Special Care in Dentistry* 35, no. 3 (2015): 114–119.
7. Maenner, Matthew J., Kelly A. Shaw, and Jon Baio. ‘Prevalence of Autism Spectrum Disorder among Children Aged 8 Years—Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2016’. *MMWR Surveillance Summaries* 69, no. 4 (2020): 1.
8. Muller, Michael J., and Sarah Kuhn. ‘Participatory Design’. *Communications of the ACM* 36, no. 6 (1993): 24–28.
9. Narzisi, A., M. Posada, F. Barbieri, N. Chericoni, D. Ciuffolini, M. Pinzino, R. Romano, M. L. Scattoni, R. Tancredi, and S. Calderoni. ‘Prevalence of Autism Spectrum Disorder in a Large Italian Catchment Area: A School-Based Population Study within the ASDEU Project’. *Epidemiology and Psychiatric Sciences* 29, no. e5 (2018).
10. Narzisi, Antonio, Mariasole Bondioli, Francesca Pardossi, Lucia Billeci, Maria Claudia Buzzi, Marina Buzzi, Martina Pinzino, Caterina Senette, Valentina Semucci, and Alessandro Tonacci. ‘“Mom Let’s Go to the Dentist!” Preliminary Feasibility of a Tailored Dental Intervention for Children with Autism Spectrum Disorder in the Italian Public Health Service’. *Brain Sciences* 10, no. 7 (2020): 444.
11. Popple, Ben, Carla Wall, Lilli Flink, Kelly Powell, Keri Discepolo, Douglas Keck, Marilena Mademtzi, Fred Volkmar, and Frederick Shic. ‘Brief Report: Remotely Delivered Video Modeling for Improving Oral Hygiene in Children with ASD: A Pilot Study’. *Journal of Autism and Developmental Disorders* 46, no. 8 (2016): 2791–2279.
12. Stein, Leah I., Jose C. Polido, and Sharon A. Cermak. ‘Oral Care and Sensory Over-Responsivity in Children with Autism Spectrum Disorders’. *Pediatric Dentistry* 35, no. 3 (2013): 230–235.





# Designing Educational Supports for People with Intellectual Disabilities

Juan Carlos Torrado<sup>1</sup>, Javier Gomez<sup>2</sup>, Letizia Jaccheri<sup>3</sup>, Germán Montoro<sup>4</sup>  
<sup>1</sup> Norwegian University of Science and Technology – juan.c.t.vidal [at] ntnu.no  
<sup>2</sup> Universidad Autónoma de Madrid – jg.escribano [at] uam.es  
<sup>3</sup> University of Science and Technology – letizia.jaccheri [at] ntnu.no  
<sup>4</sup> Universidad Autónoma de Madrid – german.montoro [at] uam.es

## ABSTRACT

In this paper, we present two applications that have been designed to investigate the potential of mobile technologies to improve the educational tasks of people with intellectual disabilities. Specifically, we describe Leo con Lula, a mobile app that assists people with autism spectrum disorders to train global reading skills, and BART, an app based on augmented reality that assists people with intellectual disabilities to train self-evaluation skills through math exercise sheets. From the experience of designing the application and the studies to evaluate their efficacy, we come up with some design guidelines aimed to inform the creation of similar educational tools

## KEYWORDS

intellectual disabilities, educational technologies, mobile application

## TALK

### 1 INTRODUCTION

Educational technologies have been a hot topic in recent years and are becoming even more important with the lockdown situation and online teaching these days. Appropriate methods and new tools are necessary to provide a ubiquitous and high-quality learning experience. Besides online and remote learning, unsupervised activities are also relevant, in the case of special education. In these schools, classes are often limited to a certain number of students (typically 4 to 6, depending on the country's regulations) due to the constant need of support and assistance of these students. Moreover, in addition to the teacher, assistants are very common to help with some duties, such as reading, caring or during the breaks. Despite these efforts, additional supervision and help is needed, but this involves more human and time resources which are not available in most situations. Under these circumstances, educational technologies provide a new frame and opportunities to provide an unsupervised learning experience without losing teachers' track, support, and analysis.

By means of interaction recording, adapted media and tools, students can receive activities and training to be done autonomously, and teachers are provided with the appropriate data to analyse students' performance. Furthermore, educational technologies add a new level of personalisation and adaptation in different aspects. Not only can contents be adapted to the student, but also the presentation and the interaction can be personalised to fulfil user's needs. The aim of this paper is to present our vision and guidelines of how educational technologies for people with intellectual disabilities should be designed. Particularly, we present two mobile educational applications: Leo con Lula (a tablet app to introduce reading to children with Autism Spectrum Disorders) and BART (a mobile app to practice basic arithmetic operations and self-evaluation through an augmented reality experience).

## **2 RELATED WORK**

There are a plethora of studies about introducing technology to assist people with cognitive disabilities throughout their learning stages [2]. However, when it comes to self-evaluation of exercises and deliverables, the technology employed is expected to work with the material that is going to be evaluated, which is normally paper. For instance, designing a system to help children to get their writing exercises corrected and revised must be able to receive the writing exercises as input. Exercises about numerical operations also need the self-evaluating assistant to work with the paper in which the exercise has been solved. Issues that require combining physical objects with technology are often solved by using Augmented Reality (AR) [5] This technology has become recently popular for educative purposes [10] given its potential to enhance the information about the real world by inserting virtual objects and information through a screen or visual representation device.

In contrast, communication is one of the widest areas explored among technologies for the education of people with cognitive disabilities. Particularly, as it is one of the limitations many people with ASD have and, due to its imperative necessity for autonomous and proper life development, many researchers focused on developing and evaluating tools to provide augmented and alternative ways of communication (AAC) [7].

Related to communication, we find literacy. Introducing reading to children with ASD is challenging, but researchers such as Grindle [4] studied the suitability of specific software for early reading access (MimioSprout® *Early Reading*, MER) with children with autism. In their study, 4 children (aged 4 to 7 years old) participated in 80 online sessions that took place during the academic year. From the data collected and consequent analysis, the authors found that participants presented an improvement in their word recognition skills after completing the sessions.

## **3 MOBILE EDUCATIONAL APPLICATIONS**

In this section, we are going to propose two mobile educational applications aimed to help children with intellectual disabilities in two issues: global reading skills and self-evaluation skills. The former is called Leo con Lula [3], and the latter is BART [8].

## LEO CON LULA

Leo con Lula (“reading with Lula”, in Spanish) provides an introduction to reading for children with ASD. Learning methods, user’s interaction and game mechanics have been designed to adapt to these users. Particularly, the learning method is based on global reading [6] and error-free learning. As opposed to syllabic reading method, global reading is based on visual skills and begins with the picture-word identification, then the word is divided into the syllabus and so on. Regarding error-free learning, users have to choose the correct answer in order to advance. This mechanism has been implemented by means of customizable feedback in order to help the user hit.

The game is divided into three levels, increasing the difficulty. Each one is also divided into three sub-levels and, finally, each sub-level consisted of a (configurable) set of exercises. Additionally, the vocabulary is divided into groups of three words, so only a reduced number of words was trained at a time. Depending on the level and sublevel, the user is presented with a set of words and pictures and has to link them. In the lower levels, the amount of options is limited while in the higher ones (syllabic decomposition) there are more elements to connect.

Regarding the interaction and interface design, it is as simple as possible, and consist of drag and drop actions and clear and customizable elements. As an example, teachers can choose among different font types (specifically oriented to ASD teaching), amount of decorative elements and customized feedback. In order to encourage autonomous work and require less supervision from the teacher, the game records all the interaction so teachers can analyse the user’s performance afterwards. At the moment, only the basic information is provided in the app, such as the number of sessions, the cumulative number of hits and errors per user and level is showed in the game; however, further analysis can be included as more information is recorded.

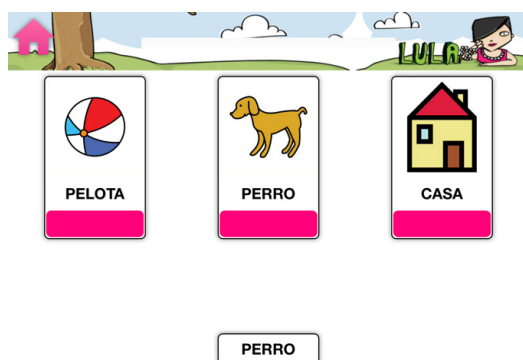


Figure 1: Leo con Lula screenshot. The user has to drag the word "perro" (dog) to the suitable space

## BART

Regarding BART, our goal is to check to what extent AR technology is suitable to help children with cognitive impairment in their self-evaluation learning process, in terms of efficacy (if they can self-evaluate successfully), efficiency (within a reasonable time) and satisfaction (taking advantage of the motivation that children usually show when technology is involved). We also wanted to explore the development of self-evaluation skill in a context of maths exercises, since this is one of the fields that requires a higher intensity of evaluation in the learning process.

For this purpose, we have developed a system that helps to solve basic arithmetic operations (BART) and designed to be used by individuals of age 10-13 with cognitive disabilities.

The system represents virtual surfaces on the real surfaces to provide information by recognising QR codes, instead of detecting shapes or representing virtual 3D models. By means of a simple and straightforward menu, it allows the following functionality:

- Create exercises: so far, we have implemented an interface to store a list of arithmetic operations.
- Generate exercise sheets: it creates a .png or .pdf with the exercise belonging to a certain mathematical operation. It contains the operation to be solved and the QR code to virtually print the solution afterwards (see Figure 2. BART exercise sample sheet)
- Visualise solutions: It opens the camera of the device and waits for the user to point towards an exercise sheet to virtually print the solution on the right side of the work of the student. Users may view this information as much time as they want, although it is intended for making agile checks and correcting the operation if it is incorrect, or perform positive reinforcement otherwise (see Figure 3. AR solution visualiser).



Figure 2: BART exercise sample sheet. On the left part, the operation is placed, with a blank space to fill the solution. On the right side, a QR code is placed. It will be replaced with the solution through the AR capabilities of the app. At the bottom of the page, the user is asked to self-evaluate the result.

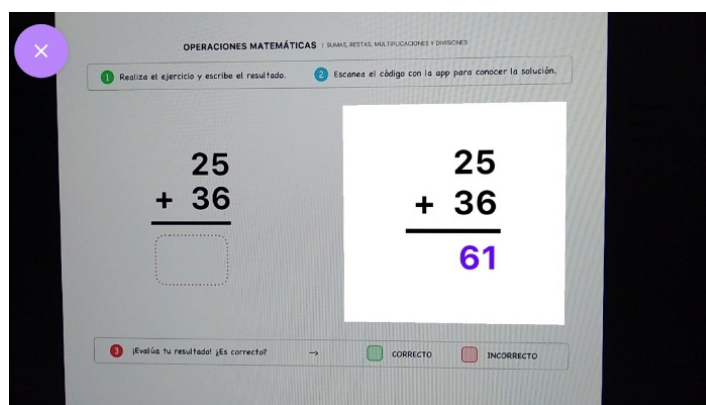


Figure 3: AR solution visualiser. Once the QR code is scanned with the app, an AR version of the result is displayed.

In addition, to train self-evaluation and math skills, this application is aimed to challenge the fine-motor skills of children with intellectual disabilities, since it makes them use a tablet, paper material and writing gear. Fine-motor skill is a term that refers to the coordination eye-hand that has todo do with manipulating small objects, doing precise movements or manipulating more than one object simultaneously.

## 4 DESIGN GUIDELINES

The two applications share design features due to the needs of the target population they have been tailored for. In this section we provide a short insight of these common features:

- **Gamification:** around 65% of digital tools for people with intellectual disabilities are abandoned or not successfully adopted by them or their families [1]. Many researchers have studied the reasons behind that, being motivation to keep using technology on a daily or regular basis is motivation [9]. Gamification is a design approach that focuses on the entertainment side of the application. Achievement metaphors such as badges, trophies, medals and progress tracking mechanisms such as bars or level system and smooth learning curves contribute to the engagement of the users and make them feel more motivated to keep using it. In BART, we used visual cues that provided with positive reinforcement when the evaluation is correct, and Leo con Lula tracks the progress by challenging the user with new exercises that build on top of each other.
- **Interactive and cognitive customization:** the UX requirements of people with intellectual disabilities are not straightforward. Language is a barrier for many users, whereas strong colors or inappropriate visual cues might jeopardize the experience of others. Therefore, the visual interface must allow customization to some realistic extent.
- **Error-free learning paradigm:** the correct-wrong binary narrative for evaluating results has proven not to be adequate for many users with intellectual disabilities, due to their difficulties to tolerate frustration.
- **Multimodality:** due to the communication requirements of different individuals with intellectual disabilities, the visual interface must be able to provide information.
- **Comprehensive activity record.**

## 5 CONCLUSIONS AND FUTURE WORK

In this paper, we present a set of guidelines for educational applications for people with ASD. These ideas are based on the empirical work carried out during the design and developed of Leo con Lula, a tablet game to introduce reading to children with ASD, and BART, an AR app to work son self-evaluation skills. Despite of the different objectives of both applications, they share a set of common points which support the creation of these guidelines.

As future lines to work on, further evaluation of both applications is proposed, in order to detect possible inconsistencies or interaction difficulties, as well as to analyse users learning processes.

## 6 ACKNOWLEDGEMENTS

This work has been carried out during the Tenure of an ERCIM Alain Bensoussan Postdoctoral Fellowship, and supported in part by projects “IndiGo!” (Grant Number PID2019-105951RB-I00, Spanish Ministry of Science and Innovation), “eMadrid” (Grant Number P2018/TCS-4307) and “Uso de TIC para favorecer la inclusion del alumnado con NEE” (Spanish Ministry of Education)

## REFERENCES

1. Dawe, Melissa. 'Desperately Seeking Simplicity: How Young Adults with Cognitive Disabilities and Their Families Adopt Assistive Technologies'. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1143–1152, 2006. <https://doi.org/10.1145/1124772.1124943>.
2. Gillespie, Alex, Catherine Best, and Brian O'Neill. 'Cognitive Function and Assistive Technology for Cognition: A Review'. *Journal of the International Neuropsychological Society* 18, no. 1 (2012): 1–19. <https://doi.org/10.1017/S1355617711001548>.
3. Gomez, Javier, Letizia Jaccheri, Juan Carlos Torrado, and Germán Montoro. 'Leo Con Lula, Introducing Global Reading Methods to Children with ASD'. In *Proceedings of the 17th ACM Conference on Interaction Design and Children*, 420–426, 2018. <https://doi.org/10.1145/3202185.3202765>.
4. Grindle, Corinna F., J. Carl Hughes, Maria Saville, Kathleen Huxley, and Richard P. Hastings. 'Teaching Early Reading Skills to Children with Autism Using MimioSprout Early Reading'. *Behavioral Interventions* 28, no. 3 (2013): 203–224.
5. Milgram, Paul, Haruo Takemura, Akira Utsumi, and Fumio Kishino. 'Augmented Reality: A Class of Displays on the Reality-Virtuality Continuum'. In *Telemanipulator and Telepresence Technologies*, 2351:282–292. International Society for Optics and Photonics, 1995. <https://doi.org/10.1117/12.197321>.
6. Serges, Jean. 'La Enseñanza de La Lectura Por El Método Global', 1958.
7. Shane, Howard C., Emily H. Laubscher, Ralf W. Schlosser, Suzanne Flynn, James F. Sorce, and Jennifer Abramson. 'Applying Technology to Visually Support Language and Communication in Individuals with Autism Spectrum Disorders'. *Journal of Autism and Developmental Disorders* 42, no. 6 (2012): 1228–1235.
8. Torrado, Juan C., Javier Gomez, and Letizia Jaccheri. 'Supporting Self-Evaluation for Children with Mental Disabilities through Augmented Reality'. In *Proceedings of the 18th ACM International Conference on Interaction Design and Children*, 635–641, 2019. <https://doi.org/10.1145/3311927.3325307>.
9. Torrado, Juan C., Ida Wold, Letizia Jaccheri, Susanna Pelagatti, Stefano Chessa, Javier Gomez, Gunnar Hartvigsen, and Henriette Michalsen. 'Developing Software for Motivating Individuals with Intellectual Disabilities to Do Outdoor Physical Activity'. In *2020 IEEE/ACM 42nd International Conference on Software Engineering: Software Engineering in Society (ICSE-SEIS)*, 81–84. IEEE, 2020. <https://doi.org/10.1145/3377815.3381376>.
10. Wu, Hsin-Kai, Silvia Wen-Yu Lee, Yi Chang Hsing, and Liang Jyn Chong. 'Current Status, Opportunities and Challenges of Augmented Reality in Education'. *Computers & Education* 62 (2013): 41–49. <https://doi.org/10.1016/j.compedu.2012.10.024>.

# MAV

## Un nuovo Manuale Audio Video di scuola guida per Bisogni Educativi Speciali

Marcella Banfi  
OPPI<sup>1</sup> Milano, Italia – marcella.banfi [at] gmail.com

### SINTESI

Il contributo analizza i passi realizzati dall'autrice dell'intervento e da Stefano e Luciana Bottoli, casa editrice EsseBì SRL, per la creazione di un innovativo Manuale di scuola guida Audio-video per Bisogni Educativi Speciali. L'intervento è stato sia sul fronte della revisione linguistica che su quello della semplificazione delle immagini e della creazione di supporti multimediali. Abbiamo realizzato un nuovo manuale multimediale semplificato, adatto a chi abbia delle fragilità nella letto scrittura, sia che esse siano di origine culturale, linguistica o neuro-cognitiva. Un manuale che accosta alcuni accorgimenti specifici per studenti con dislessia ad altri - come ad esempio una articolazione interna dei contenuti più schematica e una semplificazione lessicale - di cui possono beneficiare tutti. Un testo che, sfruttando le moderne tecnologie della comunicazione, amplia le sue potenzialità con una consistente quantità di materiale multimediale.

### PAROLE CHIAVE

inclusione, tecnologie per l'inclusione, imprenditorialità sociale

### INTERVENTO

Ci sono parole che modificano i contesti che le circondano, trasformando interpretazioni e approcci, sfumando categorie e strutturando riflessioni più puntuali. “Dislessia” è certamente una di queste. Il più noto e, spesso, anche il più frainteso tra i termini che afferiscono ai disturbi specifici dell'apprendimento. La sua introduzione nella legislazione italiana - nel 2010 - ha costretto a riconsiderare molte delle categorie in uso nelle aule scolastiche. Ma cosa è successo nel mondo di fuori? possiamo dire che qualcosa sia cambiato?

In classe, ormai, si seguono numerose buone prassi che permettono agli studenti con DSA di compensare le loro difficoltà e conseguire, così, la loro forma di eccellenza personale; purtroppo lo stesso, ancora, non si può dire del mondo al di fuori.

---

1 Organizzazione Professionale Per la formazione degli Insegnanti.



Per quanto la natura evolutiva dei disturbi specifici permetta una loro compensazione e per quanto si possa affermare che le abilità di apprendimento vengano esercitate sostanzialmente all'interno della scuola, è altresì vero che le società del terziario avanzato mettono sempre più in gioco le competenze di autoformazione e apprendimento continuo. In questo senso gli studenti con DSA dovrebbero avere il diritto di utilizzare strumenti adatti alla loro specificità neurobiologica anche nelle situazioni di tutti i giorni: consultando una pagina web, leggendo un quotidiano, preparandosi all'esame di guida.

Ci sembrava, quindi, che fosse doveroso predisporre un manuale di scuola guida più rispettoso dei bisogni specifici dell'apprendimento di tutti quei ragazzi, diagnosticati o meno, che hanno esigenze speciali legate all'apprendimento. In questo senso il nostro nuovo manuale semplificato è adatto a chiunque abbia delle fragilità nella letto scrittura, sia che esse siano di origine culturale, linguistica o neuro-cognitiva.

Abbiamo introdotto, quindi, nella riscrittura, alcuni accorgimenti specifici per studenti con dislessia e altri - come ad esempio una articolazione interna dei contenuti più schematica e una semplificazione lessicale - di cui possono beneficiare tutti. Inoltre, sfruttando le moderne tecnologie della comunicazione, il manuale cartaceo è stato ampliato con una consistente quantità di materiale multimediale: audio, video e audio-animazioni. Spesso, infatti, un'immagine ben costruita è più efficace di molte parole; se a questa si aggiungono animazione e un commento vocale, la percentuale di persone raggiunte si amplia notevolmente. Vedremo, ora, nel dettaglio le modifiche introdotte alla versione standard del manuale.

Per quanto tutti beneficino di una spaziatura ampia tra le righe e di un font senza grazie, queste componenti dello scritto sono fondamentali per un ragazzo con dislessia. Le particolarità del suo movimento oculare, la saccade, fanno sì che, se le linee del testo sono troppo addossate, parti della parola sottostante interferiscano con quella sopra, creando gravi difficoltà nella decifrazione. In considerazione di ciò, le prime modifiche al manuale sono state l'utilizzo di un'interlinea doppia e di un carattere di dimensioni adeguate, privo di grazie.

Per quanto riguarda l'impaginazione del testo si è scelto di non giustificare le righe ma di mantenere l'allineamento a sinistra: per quanto impercettibilmente, infatti, con la giustificazione, sia l'intervallo tra i caratteri delle parole che gli spazi bianchi si modificano in base alla lunghezza da occupare. Questo, che può sembrare utile ai fini dell'ordine della pagina, crea problemi agli studenti con dislessia perché li obbliga a ricalibrare continuamente la saccade, appesantendo ancor di più il lavoro di decodifica. Viceversa, l'allineamento a sinistra mantiene invariata la distanza tra le lettere e gli spazi bianchi, agevolando la regolarità dei movimenti oculari.

Si conosce da tempo l'importanza di limitare al massimo gli sforzi compiuti per la lettura del testo. Ognuno di noi ha, infatti, un limite nella quantità di energia che può attivare per un compito, è importante, quindi, che per gli studenti con DSA sia creato un ambiente di lettura che, limitando gli sforzi della decodifica fonologica, lasci più spazio alla comprensione del significato.

Un correlato logico della fatica e della lentezza nella lettura, infatti, è la difficoltà di comprensione. Per questa, che si riscontra non solo negli studenti con DSA, ma anche in una fascia sempre più ampia di persone che sembra incapace di strutturare gerarchicamente e articolare concetti tratti da un testo lineare, si è svolto un lavoro di riordino e ri-assemblamento dei contenuti del manuale.

I paragrafi sono stati riorganizzati in macro-categorie, che a loro volta ne contengono altre. I contenuti sono stati organizzati in maniera logicamente più ferrea, ponendo attenzione nell'esplicitare le connessioni tra le parti ed evitando di far comparire stessi contenuti in sedi diverse per non creare confusione. Un esempio potrà spiegare meglio il lavoro di riorganizzazione; prendiamo i segnali previsti per il passaggio a livello. Questo argomento è piuttosto difficile perché ci sono diversi tipi di passaggio che hanno alcuni segnali comuni e altri differenti. Nella versione del manuale non semplificata i segnali vengono ripetuti ad ogni tipologia di attraversamento ferroviario senza nessun tipo di introduzione. Il compito di sintetizzare e desumere dal testo che esistono diversi tipi di passaggi a livello e che alcuni cartelli sono comuni e altri no è lasciato allo studente a fine lettura.

Nel manuale semplificato, invece, gli impliciti di articolazione vengono palesati: per prima cosa si elencano i diversi tipi di passaggi a livello, poi si avvisa che alcuni segnali saranno comuni ed altri no; si prosegue mostrando tutti i segnali con i loro significati specifici e solo poi si elencano le diverse tipologie di attraversamenti abbinandole solo ai segnali che mutano. Si chiude con una tabella sinottica di raffronto.

In sostanza il manuale semplificato mostra i criteri di articolazione con maggiore dettaglio, riorganizza e raggruppa le nozioni in sottoinsiemi di contenuto più specifici e numerosi, gerarchizza.

In questo senso nella prima edizione<sup>2</sup> è stata lasciata la paragrafatura come segno grafico del passaggio da un argomento all'altro.

Non solo il testo è stato differentemente organizzato ma si è svolta anche un'attenta analisi dei contenuti. Ci si è chiesti: quali di questi garantiscono sia il superamento dell'esame che le conoscenze adeguate per formare conducenti preparati e competenti? Quanto non sembrava funzionale è stato espunto, riducendo il testo all'osso, ma si è vagliato ogni singolo quiz dell'esame per garantire che tutto ciò che può presentarsi sia trattato nel manuale.

Non siamo intervenuti solo sul contenuto, ma anche sulla forma sintattica e lessicale. Ci sono, infatti, alcuni accorgimenti linguistici che rendono un testo più leggibile. Per muoverci in questa direzione, in primo luogo, abbiamo riordinato i componenti della frase, riportando quasi tutti i casi all'ordine tipico SVO soggetto, verbo, complemento oggetto; poi abbiamo risolto le doppie negazioni; in seguito abbiamo accorciato la lunghezza delle frasi, preferendo una concatenazione per paratassi piuttosto che per ipotassi. Le frasi implicite sono state, quando possibile e utile, esplicitate. Infine nelle frasi nominali è stato inserito un verbo.




Si veda un esempio:

- formulazione del manuale non semplificato: “Nei centri abitati, causa l'elevata concentrazione di utenti della strada e abitanti presenti contemporaneamente, numerose e particolari sono le situazioni di pericolo che possono presentarsi.”
- formulazione del manuale semplificato: “Nei centri abitati la circolazione è più difficile perché ci sono molti fattori da controllare contemporaneamente. Quindi bisogna fare molta attenzione e usare prudenza.”

---

2 Stiamo lavorando alla seconda edizione migliorata e abbiamo sottoposto il testo a una revisione incrociata tra specialisti del settore e studenti con BES

Il periodo originale, composto da una subordinata causale implicita che divideva a metà la principale, è stato suddiviso in due frasi separate; si è integrata sciolto l'implicito ed è stato ripristinato l'ordine naturale della frase. La semplificazione ha investito anche il lessico: si sono scelte parole ad alta disponibilità, quando possibile, e quando non lo era si sono introdotti i termini specifici sempre muniti da glosse interpretative che li agganciassero a pre-conoscenze comuni, per facilitarne la memorizzazione. Quando alcune immagini o spiegazioni creavano ambiguità o si presentava la necessità di raffronti di contenuti trattati in punti lontani, si sono creati box separati, aggiunti alla fine del paragrafo per non interrompere lo svolgimento del testo.

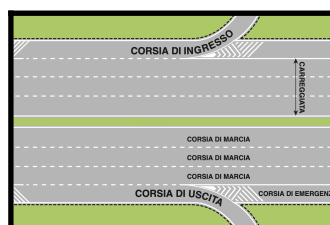
|   |   |  |
|---|---|--|
|  |  |  |
| Obbligo di andare dritto  | Obbligo di svoltare a sx  | Obbligo di svoltare a dx   |
| Vietato: svoltare a dx e a sx   | Vietato: proseguire dritto o svoltare a dx  | Vietato: proseguire dritto o svoltare a sx   |

Il lavoro sulle immagini ha comportato una serie di scelte ben definite: mentre nella versione non semplificata si trovano alcune fotografie commiste a immagini, per il nuovo manuale si è scelto di utilizzare solo illustrazioni prodotte in computer grafica.

Si è considerato, infatti, che uno studente con DSA fatica se posto di fronte a una molteplicità di stimoli grafici; se l'immagine è ricca di particolari non lo aiuterà ad assimilare un concetto, al contrario, genererà solo confusione e malessere. Perciò da un lato le fotografie, con la loro presa dal vivo fitta di particolari, rendevano difficile una focalizzazione adeguata sul concetto da illustrare, dall'altro non garantivano la plasticità necessaria alla semplificazione che desideravamo.

Tutte le immagini sono state pulite, lavorando nell'ottica del "less is more" lasciando solo i dettagli significativi e, quando possibile, utilizzando una veduta in pianta.

Si confrontino le due immagini qui di seguito:



L'evoluzione va letta da sinistra verso destra. Come si può vedere i numeri di riferimento, faticosi perché comportano un passaggio continuo dalla figura al testo e viceversa, sono stati sostituiti con scritte inserite nella figura. Inoltre la prospettiva centrale è stata modificata per aumentare l'intuitività dell'immagine; è stata tolta tutta la segnaletica perché non era oggetto di analisi specifica del capitolo e si è aumentato il grado di astrazione dell'immagine in funzione di una maggior iconicità. Ogni immagine del manuale è stata considerata singolarmente e sistemata in base alle sue peculiarità.

Grazie all'enorme mole di dati statistici posseduta da Essebi Italia Srl, è stato possibile selezionare i quesiti più difficili per i corsisti. Quando un quiz superava il 30% di risposte errate era catalogato come quiz difficile. Successivamente ogni singola domanda veniva osservata e, nel caso, semplificata linguisticamente secondo le stesse regole usate per il manuale. Grazie alla piattaforma on line, è stato poi possibile inserire un video multimediale per supportare la comprensione di ognuno di questi quesiti.

Grande attenzione è stata data alla parte on-line del manuale; in primis la lettura integrale del manuale non è stata ottenuta attraverso un sintetizzatore vocale, bensì attraverso la lettura di una vera attrice. Questo perché la lettura sintetica è ancora piena di imprecisioni prosodiche, ovvero errori legati sia al posizionamento degli accenti interni alle parole, sia al ritmo di elocuzione, il che complica più che semplificare la comprensione del testo per gli studenti con DSA. Per questi, infatti, l'ascolto non integra la lettura del testo, bensì la sostituisce poiché mantenere alta l'attenzione simultanea nei due canali, visivo e uditivo, è molto complicato.

I video, infine, non illustrano solo i quesiti difficili, ma anche le parti del testo che, sempre attraverso l'analisi statistica delle risposte errate ai quiz, abbiamo individuato come più ostiche<sup>3</sup>.

In conclusione riteniamo questo nostro manuale un primo passo, sicuramente migliorabile<sup>4</sup>, verso un sistema per l'ottenimento della patente più equo, che garantisca a tutti una più efficace assimilazione dei concetti di base per un uso più consapevole e responsabile della strada.

Un supporto che garantisca maggior leggibilità e quindi un possibile miglioramento nella preparazione dei corsisti ci sembra importante in un Paese, come il nostro, in cui i morti sulle strade sono mediamente 6 punti percentuale<sup>5</sup> più alti rispetto alla media europea.

---

3 Sono stati vagliati singolarmente i 7171 quesiti ministeriali, tra questi sono state selezionate come difficili 730 domande. Ogni domanda ha un supporto audio-video. Inoltre abbiamo preparato 120 video aggiuntivi per i contenuti ritenuti più ostici.

4 Stiamo lavorando a un trial di verifica per migliorare la prima edizione del manuale e per avere conferma del suo funzionamento. Purtroppo lo stop imposto alle scuole guida causa Covid 19 ha rallentato il percorso e, ad oggi, non abbiamo che pochi dati a sostegno del nostro lavoro.

5 Fonte ISTAT:  
[www.istat.it/it/archivio/232366](http://www.istat.it/it/archivio/232366)

## ULTERIORI RIFERIMENTI

1. AID (Associazione Italiana Dislessia). *La dislessia raccontata agli insegnanti 1. Come riconoscerla*. Firenze: Libri Liberi, 2007.
2. Cornoldi, Cesare, a c. di. *Difficoltà e disturbi di apprendimento*. Bologna: Il Mulino, 2007.
3. Ianes, Dario, Daniela Lucangeli, e Irene Cristina Mammarella. *La discalculia e altre difficoltà in matematica*. Erikson, 2010.
4. Notabartolo, Daniela. *Competenze testuali per la scuola*. Roma: Carocci, 2014.
5. Stella, Giacomo. *Dislessia oggi. Prospettive di diagnosi e intervento in Italia dopo la legge 170*. Trento: Erickson, 2011.

# **Sessione IV**

## **Amartya Sen**



# La lotta informatica per la Democrazia cibernetica

Giacomo Tesio  
giacomo [at] tesio.it

## SINTESI

L'invenzione di processori general purpose, la rapida miniaturizzazione dell'elettronica e la costruzione di una rete internazionale di scambio dati, hanno avviato una serie di rivoluzioni sociali, culturali e politiche analoghe a quelle prodotte dalle precedenti innovazioni informatiche, ma su scala planetaria ed in rapidissima successione.

Con la propria pervasività, l'informatica sta infatti creando una nuova società cibernetica che lega, integrando sempre più profondamente, complessi sistemi automatici e persone inconsapevoli del loro funzionamento.

La realizzazione, la manutenzione ed il controllo di questi sistemi, automatici ma non autonomi, è affidata ad una nuova categoria di lavoratori, gli informatici, che iniziano lentamente ad acquisire una propria coscienza di classe, esprimendo esigenze, valori e metodi di lotta solo in parte riconducibili a quelli del passato.

Analizzando le caratteristiche di questa nuova lotta di classe "informatica", è possibile evidenziare le dinamiche di potere che operano nella società cibernetica in cui viviamo, dinamiche invisibili per chiunque non sia in grado di comprendere e modificare personalmente il funzionamento degli automatismi che lo circondano.

## PAROLE CHIAVE

informatica, informazione, dato, cultura, potere, lavoro, automatismo, algoritmo, programma, bug, sicurezza, complessità



# INTERVENTO

## 1 INFORMAZIONE E DATO

*Computer Science is no more about computers than astronomy is about telescopes.*

Queste parole del grande Dijkstra esprimono molto bene il dramma di una disciplina che per ragioni storiche si è sviluppata intorno agli strumenti della propria ricerca, senza aver chiaro il proprio oggetto di studio.

Il termine Informatica deriva dalla contrazione dell'espressione "Informazione Automatica".

Definiamo automatico un qualsiasi meccanismo che, una volta avviato, produce una serie di effetti pre-determinati senza richiedere ulteriore intervento dell'uomo per svolgere le operazioni che lo caratterizzano. L'intelligenza umana opera pesantemente nella costruzione dell'automatismo, nella sua messa in funzione e nella gestione degli effetti prodotti, ma non influenza l'esecuzione dell'automatismo stesso (a meno di non voler alterare gli effetti per cui è progettato).

L'automatismo non ha però alcuna autonomia: anche quando cambia il proprio comportamento in funzione dei dati che riceve, tale variazione è predeterminata, in modo consapevole o meno, da chi lo ha realizzato.

I computer sono sempre sistemi automatici, siano essi venduti come smartphone, dispositivi medici o server "nel cloud".

L'informatica tuttavia non si limita alla creazione di automatismi riproducibili elettro-meccanicamente, ma attraverso di essi studia l'informazione, come possa essere trasferita, preservata, rappresentata, interpretata e trasformata, nonché le tecniche che applicano queste conoscenze.

Il termine Informazione, dal latino in formare, denota un'idea, un costrutto della mente umana che può essere comunicato in modo preciso ad altri esseri umani. L'informazione esiste esclusivamente nella mente di un essere umano, come esperienza soggettiva e cosciente di pensiero comunicabile, sia essa elaborata autonomamente, derivata da esperienze dirette o appresa da altri esseri umani. L'insieme delle informazioni disponibili ad un essere umano definisce le sue Conoscenze, mentre l'insieme di informazioni condivise dai membri di una comunità definisce la loro Cultura.

L'informazione nasce circa 164.000 anni fa come effetto collaterale dell'evoluzione umana, prodotta dal vantaggio evolutivo fornito dal linguaggio. Ogni informazione nasce per essere comunicata e si distingue da altre componenti dell'esperienza soggettiva umana, come intuizioni o esperienze mistiche, proprio per questa sua comunicabilità. Sono informazioni i numeri, le teorie scientifiche, le riflessioni filosofiche, le analisi storiche o politiche così come le favole, le notizie, le credenze religiose o le osservazioni su un fenomeno fisico.

Fino all'invenzione della pittura, circa 40 mila anni fa, la condivisione delle informazioni avveniva esclusivamente in modo sincrono: due o più persone dovevano incontrarsi fisicamente e condividere l'esperienza che dava origine all'informazione o metterla in comune attraverso un linguaggio sufficientemente espressivo.

Con l'invenzione della pittura e successivamente con la scrittura, l'uomo ha imparato a comunicare in modo asincrono, attraverso rappresentazioni simboliche trasferibili ed interpretabili da altri esseri umani.

Nacque così il dato, dal participio passato di dare, ovvero la rappresentazione di un'informazione che può essere appunto data, trasmessa ad altri, condivisa con altri esseri umani lontani nel tempo o nello spazio.

Il Dato è una delle possibili rappresentazioni di un'informazione, impressa su un supporto fisico ed interpretabile dall'uomo. Sebbene quasi sempre spontaneo ed inconsapevole per la mente umana, il processo di interpretazione del dato non è riducibile ad una relazione matematica: il dato veicola sempre informazioni aggiuntive rispetto al messaggio intenzionalmente espresso dal mittente e la sua decodifica in informazione da parte del destinatario dipende sempre dalle informazioni di cui questo già dispone, dalle sue conoscenze e dalle culture delle comunità di cui è membro.

Grazie a Gödel, sappiamo che è possibile dimostrare matematicamente la coerenza fra dati, ma non la loro veridicità.

Questa relazione organica bidirezionale fra informazione e dato costituisce la chiave per comprendere l'informatica.

Per migliaia di anni la scrittura è stata la forma prevalente di dato, prodotto, elaborato, riprodotto e distribuito manualmente da personale specializzato la cui enorme influenza politica derivava direttamente dai limiti dei loro strumenti, fossero essi geroglifici nell'antico Egitto, logogrammi nella Cina imperiale o le raffinate calligrafie dei monaci amanuensi. Dunque per millenni la creazione, l'elaborazione e la riproduzione del dato è stata prerogativa di uomini al servizio del potere.

Con l'invenzione della scrittura a caratteri mobili di Gutenberg, venne introdotta una prima automazione in questo processo, ma esclusivamente nella fase di riproduzione dei dati: l'elaborazione e la distribuzione dei testi scritti rimaneva prerogativa umana, ma la loro riproduzione semi-automatica abilitò rivoluzioni culturali e politiche impensabili in precedenza, dal Rinascimento all'Illuminismo, dalla Riforma Protestante all'alfabetizzazione di massa.

Le conseguenze di questa piccola innovazione informatica divennero presto chiare alle istituzioni politiche dell'epoca che cercarono di arginarne il potenziale rivoluzionario attraverso liste di libri proibiti, licenze governative che garantivano piccoli monopoli editoriali in cambio del controllo statale sui contenuti pubblicati, fino all'introduzione nel 1710 del concetto di Copyright nello Statuto di Anna.

## **2 ALGORITMO E PROGRAMMA**

Nel 1645, con l'invenzione della prima calcolatrice meccanica da parte del giovane Blaise Pascal, nacque il primo meccanismo automatico in grado di elaborare dati. L'interfaccia utente era estremamente semplice: l'operatore inseriva i dati in input esprimendo i numeri da sommare o sottrarre attraverso alcune rotelle attivate da uno stilo ed interpretava i dati in output restituiti da alcuni dischi numerati. Alla Pascalina seguirono diversi calcolatori meccanici, tutti vincolati ad elaborazioni predeterminate e specifiche, come la calcolatrice di Leibniz o il calendario perpetuo di Giovanni Plana.

La macchina analitica ideata da Charles Babbage e proposta pubblicamente nel 1837 fu la prima a prevedere l'esecuzione di un automatismo variabile, espresso e fornito alla macchina sotto forma di schede perforate: Babbage ideò il Programma ovvero la rappresentazione di un algoritmo eseguibile automaticamente da una macchina.

Per diverse ragioni incidentali, la macchina analitica non fu mai realizzata e fu necessario aspettare quasi cento anni prima che Turing ideasse la Universal Turing Machine, un modello matematico in grado di descriverne ed astrarne il funzionamento.

L'ideazione della Universal Turing Machine, una macchina programmabile in grado di riprodurre il funzionamento di qualsiasi altra macchina programmabile (con gli stessi dispositivi di input/output), ha posto le basi per realizzare linguaggi interpretati, macchine virtuali, emulatori, VPS e business model come l'IaaS (infrastructure as a service).

Malgrado ciò, la ricerca di una definizione formale del concetto di Algoritmo ha impegnato per circa 200 anni filosofi, matematici ed informatici come Church, Gödel, Minsky e Knuth, ma a tutt'oggi non esiste una definizione universalmente accettata. La ragione di questo ritardo risiede nella confusione fra informazione e dato, la cui relazione organica e bidirezionale si ripresenta fra algoritmo e programma.

Un Algoritmo è una particolare tipologia di informazione che descrive una sequenza di azioni applicabili ad un sistema.

In quanto informazione, esiste sempre ed esclusivamente nella mente di un essere umano. La sua espressione, la sua trasformazione in dato trasferibile ed eseguibile automaticamente, è invece un Programma.

E' importante sottolineare come, contrariamente ai calcolatori meccanici precedenti, limitati alla specifica elaborazione per cui erano costruiti, le macchine Turing-complete permettono l'esecuzione di elaborazioni variabili, limitate esclusivamente dalle possibilità espressive del linguaggio in cui vengono rappresentate e dall'hardware disponibile.

Per poter intuire la portata di questa innovazione (e del potere che ne deriva) bisogna considerare che quando un computer esegue un programma, ciò che sostanzialmente è un atto di pura immaginazione della mente di un programmatore, dopo essere stato espresso in un certo linguaggio, inizia ad operare sulla realtà fisica e sociale, riproducendone la volontà ed imponendola a coloro che con tale programma interagiscono per percepire, interpretare o agire sulla realtà che li circonda.

Comprendere la differenza fra Algoritmo e Programma è dunque fondamentale per sventare i tentativi di nascondere, dietro un velo di matematica neutralità, gli interessi e la responsabilità delle scelte che il software esprime.

In quanto dato, in quanto rappresentazione, il programma è infatti soggetto a diverse categorie di errore:

1. errori di comprensione dell'algoritmo, nella mente del programmatore;
2. errori di rappresentazione dell'algoritmo, introdotti durante la scrittura del programma;
3. errori di esecuzione del programma, dovuti alla macchina che lo esegue automaticamente;
4. errori di utilizzo del programma, dovuti per lo più a problemi di documentazione o a scelte commerciali.

La seconda e la terza categoria attengono alla rappresentazione dell'algoritmo, al programma, e sono tipicamente definiti Bug. Le altre due categorie affliggono invece gli interlocutori, che l'esecuzione automatica del programma mette in comunicazione asincrona attraverso la collaborazione che veicola: i programmatori e gli utenti.

Le probabilità di queste categorie di errore variano in modo indipendente: la comprensione dell'algoritmo dipende dalla cultura del programmatore e dalla sua esperienza del problema che intende risolvere, la probabilità di errori di rappresentazione dipende anzitutto dalla dimensione complessiva del programma (che include le sue dipendenze dirette o indirette) e la probabilità degli errori di esecuzione dipende dalla complessità dell'hardware, reale o virtuale, che lo esegue. Infine la probabilità di errori nell'utilizzo di un software da parte dell'utente dipende dalla sua comprensione del suo funzionamento, ovvero degli algoritmi che questo implementa.

La probabilità che un programma funzioni correttamente corrisponde al prodotto dei complementi di queste probabilità.

La probabilità che un sistema di programmi sviluppati indipendentemente ma integrati per realizzare una determinata applicazione funzioni correttamente, decade in modo esponenziale all'aumentare dei programmi integrati.

### **3    PROTOCOLLI E RETI**

Parallelamente allo sviluppo dei processori general purpose, il progresso dell'elettronica ha permesso di miniaturizzare una straordinaria varietà di componenti: qualsiasi smartphone è composto di microcamere, microfoni, giroscopi, ingressi USB, ricevitori GPS, Bluetooth, WiFi, LTE etc, componenti che appaiono ormai anche su automobili, forni e frigoriferi.

Tutti questi componenti registrano dati dall'ambiente e li scambiano fra loro e con sistemi automatici esterni attraverso programmi che rappresentano una particolare categoria di algoritmi: i "protocolli".

Un Protocollo definisce la sintassi e la sequenza dei messaggi che diversi sistemi automatici possono scambiare per coordinare l'esecuzione di un'elaborazione più complessa, talvolta distribuita nel tempo o nello spazio. Ciascuno di questi protocolli determina trade-off tecnici ed economici che favoriscono determinate topologie (centralizzate, decentrate, distribuite) e determinate relazioni fra i software (client/server, P2P, federazione, incompatibilità reciproche) determinando complesse relazioni di potere fra le persone che programmano, dirigono o partecipano a vario titolo nell'elaborazione.

Veicolate da un'unica rete globale sempre attiva, questi programmi automatizzano contemporaneamente la distribuzione di due categorie di dati: quelli che le persone esprimono intenzionalmente e quelli che le descrivono contro la loro volontà.

Questa distribuzione di dati avviene attraverso la creazione virtualmente istantanea di innumerevoli copie non tracciabili, distribuite attraverso la rete ad un numero non determinabile di intermediari automatici, data controller e data processor.

Azioni che appaiono semplici, come la visita di una pagina Web o l'ingresso in una stanza dotata di IoT, avviano complesse elaborazioni automatiche che travalicano legislazioni e continenti, continuando ad operare per anni attraverso i profili comportamentali che alimentano. Poiché la copia dei dati è istantanea, gratuita e non lascia tracce, privacy e sicurezza dipendono dal controllo costante su chi vi ha accesso: sono Pubblici i dati reperibili attraverso informazioni pubblicamente disponibili, mentre sono Riservati quelli accessibili solo a coloro che detengono specifiche credenziali riservate.

## 4 COMPLESSITÀ E POTERE

*Everyone knows that debugging is twice as hard as writing a program in the first place. So if you're as clever as you can be when you write it, how will you ever debug it?*

Quando Brian Kernighan scrisse queste parole, nel 1978, Unix consisteva di 14.795 linee per il kernel e 141.833 per tutto il resto del sistema, includendo un compilatore C, diversi linguaggi di scripting ed alcuni giochi. All'inizio del 2020, il solo kernel Linux contava quasi 28 milioni di righe di codice scritte da decine di migliaia di programmatori, con una crescita di circa 4 volte rispetto al 2005. Tuttavia un kernel, pur nella sua complessità, è un programma indipendente.

La stragrande maggioranza dei programmi esistenti dipende da altri programmi, sia durante la scrittura che per l'esecuzione.

Su Debian GNU/Linux, Firefox dipende direttamente o indirettamente da 157 programmi diversi<sup>1</sup>. Chromium (Google) conta addirittura 241 dipendenze<sup>2</sup>. Su Android o su Windows, questi software hanno dipendenze completamente diverse.

Questa enorme complessità, in parte accidentale ed in parte scientemente perseguita, ha determinato la nascita di una nuova categoria di lavoratori, gli informatici, in grado di articolarla e dirigerla verso obiettivi economici o politici.

Aziende ed agenzie di intelligence hanno interesse a massimizzare la diffusione dei dati espressi (contenuti o programmi) o emessi (personal data) dai cittadini, minimizzando al contempo la diffusione dei dati che descrivono la propria operatività (trasparenza) e controllando l'accesso ai dati che producono (intellectual properties).

Queste organizzazioni, ben consapevoli del fatto che la conoscenza è potere solo dove l'ignoranza è diffusa, sfruttano la complessità dell'informatica contemporanea in molti modi: per scongiurare regolamentazioni efficaci, per rendere incomprensibili le proprie clausole contrattuali, per instaurare monopoli commerciali, per costringere la stampa sulle proprie piattaforme di distribuzione (controllandone capillarmente la circolazione) o per marginalizzare il software libero attraverso software open source talmente complessi da vanificare le 4 libertà o trasformarle in privilegi elitari.

Per decenni, attraverso salari molto elevati, gratifiche, benefit aziendali ed una retorica del culture fit e dell'auto-imprenditorialità, le aziende del comparto IT hanno persuaso migliaia informatici a condividere valori e obiettivi.

Tuttavia queste aziende combattono da sempre una strenua lotta contro l'autonomia intellettuale dei propri dipendenti.

Nel 2015 Apple, Google, Intel e Adobe hanno sborsato 415 milioni di dollari porre fine alla High-Tech Employee Antitrust Litigation, evitando una condanna che vietasse gli accordi con cui riducevano la concorrenza per i "top talent".

---

1 Contando solo le dipendenze runtime: `apt-rdepends firefox-esr|grep Depends|awk '{print $2}'|sort|uniq|wc -l`

2 Contando solo le dipendenze runtime: `apt-rdepends chromium|grep Depends|awk '{print $2}'|sort|uniq|wc -l`

A fine 2019, Google ha licenziato i quattro dipendenti che avevano organizzato i Google Walkout contro la gestione delle molestie sessuali dell'azienda. A inizio 2020 il top management di Facebook discuteva su come realizzare e vendere un sistema di content control, per identificare e censurare le discussioni dei dipendenti riguardo al concetto di "unionize".

Sempre nel 2020, Amazon ha pubblicato due offerte di lavoro per un "Intelligence Analyst" ed un "Sr Intelligence Analyst" in grado di identificare "labor organizing threats against the company", per poi assumere il Generale Keith Alexander (ex-direttore del NSA). Successivamente, la diffusione di alcuni documenti riservati ha rivelato diverse attività di sorveglianza adottate da Amazon per contrastare la diffusione di contenuti sindacali ed ambientalisti fra i dipendenti, anche in Europa.

Il timore che questa ostilità evidenzia, deriva dalla consapevolezza che gli informatici costituiscono la prima categoria di lavoratori della storia a detenere il pieno controllo dei mezzi di produzione, saldamente ancorati sulle proprie spalle.

Nonostante l'egemonia culturale della Silicon Valley, è possibile osservare i segni di questa emergente coscienza di classe lungo tutta la storia dell'informatica moderna. Sin dalla caccia alle streghe degli anni '80, durante la quale molti giovani hacker e phracker statunitensi furono arrestati per aver ignorato i tabù posti a protezione dei dati di molte aziende, apparve evidente come l'accumulo di denaro non costituisca l'interesse prevalente di molti informatici.

Emblematico, già nel 1989, fu l'arresto dei membri del P.H.I.R.M., rei di aver pubblicato una guida su come accedere ai dati gestiti dal sistema di home banking di Bank Of America... dopo averlo appreso dalle brochure dell'azienda.

O le più recenti disclosure di Phineas Fisher che, nell'agosto 2014 e nel luglio 2015, pubblicò i documenti ottenuti dai sistemi informatici di due aziende (GammaGroup e HackingTeam) specializzate nella creazione di software di spionaggio e sorveglianza per regimi totalitari, rivendicando l'importanza politica del proprio gesto in una guida all'hacking informatico intitolata "HackBack! A DIY Guide for those without the patience to wait for whistleblowers".

Nel febbraio 2018, uno studente di 22 anni, Luigi Gubello, fu denunciato dal Movimento 5 Stelle e processato per aver rivelato come tutti i dati della piattaforma Rousseau fossero pubblici, accessibili a chiunque sapesse come farseli restituire automaticamente dal software del partito, senza bisogno di credenziali private.

Simili "data-breach" vengono effettuati routinariamente da giovani sistemisti, agenzie di spionaggio e gruppi di attivisti informatici come Anonymous, che usano comuni conoscenze informatiche per ricevere automaticamente dati che gli utenti credono riservati, ma che le aziende non proteggono per non sostenere i costi di un'infrastruttura IT appropriata.

Analogamente, molti informatici consapevoli dell'inadeguatezza delle leggi sul copyright, ideato per proteggere un sistema tardo-feudale ed incompatibile con democrazie in cui la sovranità appartenga al popolo, ne combattono la natura oscurantista, sia attraverso licenze copyleft dal reach molto esteso (dalla AGPLv3 alla HESSLA), sia attraverso software libero (DeCSS, BitTorrent etc...) che ignora i ridicoli tabù posti a sua protezione (CSS, DRM etc...).

Sul fronte della privacy e del controllo sociale, diversi add-ons liberi come uBlock Origin o nuTensor, vengono sviluppati da volontari per ridurre il danno prodotto da Web browser open source come Firefox o Chrome, veri e propri strumenti di sorveglianza di massa.

Altri gruppi provano coraggiosamente a forkare Android, per rimuovere le funzionalità di tracciatura inserite nel sistema da Google. Altri ancora sperimentano protocolli alternativi, meno ostili agli utenti, come 9P2000, Gemini o GNUnet.

Piccole comunità di informatici provano ad attaccare il potere alla radice, sviluppando sistemi operativi distribuiti come 9front, che “fanno pensare” antepo- nendo la semplicità e la leggibilità dei programmi alla facilità d’uso per i profani.

Ignoti sono il numero ed il costo dei bug scientemente introdotti nel software per protesta, ed in diverse occasioni informatici indignati per il comportamento predatorio o per le scelte politiche di un’azienda software, ne hanno danneggiato fortemente la credibilità (ed i fatturati), semplicemente rimuovendo dal Web il software che avevano sviluppato o mantenuto gratuitamente per anni. Nel 2016 Azer Koçulu riuscì ad interrompere il funzionamento di milioni di siti Web rimuovendo 11 righe di JavaScript da NPM. Tre anni dopo, Seth Vargo rimosse da GitHub una propria libreria Ruby interrompendo per ore le attività di sviluppo dei clienti di Chef, fra cui Facebook, SAP, IBM e Walmart.

## **5 LA LOTTA INFORMATICA... PER LA DEMOCRAZIA**

Appare dunque evidente che i valori, le istanze e le metodologie di lotta degli informatici, esattamente come le strutture di potere con cui si confrontano, non sono completamente riconducibili alle categorie novecentesche.

In una società cibernetica, chi non è in grado di comprendere e modificare il funzionamento degli automatismi da cui è circondato, diventa un ingranaggio inconsapevole nelle mani di chi li controlla. Invece di inseguire una frettolosa “alfabetizzazione digitale” in un contesto che è ancora ai geroglifici, la lotta informatica per la democrazia richiede

1. nuovo software, che anteponga la semplicità intrinseca ad una manipolativa facilità d’uso;
2. nuove leggi, che sostituiscano l’IP protection tardo-feudale con norme al servizio di un popolo davvero sovrano;
3. nuove reti FTTH, che permettano ai cittadini il completo controllo fisico dell’hardware che elabora i propri dati;
4. una nuova didattica, che permetta ai cittadini di comprendere la società cibernetica in cui vivono, partecipando attivamente alla sua costruzione: in una Democrazia cibernetica, la Scuola rimane più importante del Parlamento.

L’alternativa è un mondo in cui un giovane giornalista come Daniel Verlaan, può accedere ad una videoconferenza segreta dei Ministri della Difesa dell’Unione Europea, trovando su Twitter credenziali fotografate dal Ministro olandese.

Un mondo in cui tale intromissione produce grandi risate, ma nessun interrogativo su chi altri abbia avuto accesso a simili riunioni, “segrete” ma pubbliche, perché visibili per chiunque sappia come interrogare gli automatismi che le veicolano.

La posta in gioco è alta: dobbiamo scegliere fra una democrazia cibernetica ed una oligarchia del controllo globale.

È la scelta fra diventare hacker o robot. Cittadini o schiavi.

## ULTERIORI RIFERIMENTI

1. Bietti, Elettra. «From ethics washing to ethics bashing: a view on tech ethics from within moral philosophy». In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 210–219, 2020.
2. Brooks, Frederick P., e No Silver Bullet. «Essence and accidents of software engineering». *IEEE Computer* 20, n. 4 (1997).
3. Dark Creaper. «Hacking Bank Of America's Home Banking System», 1989.
4. Dijkstra, Edsger W. «On the nature of computing science». In *Control Flow and Data Flow: concepts of distributed programming*, 1–4. Springer, 1986.
5. Doctorow, Cory E. «IP». Locus Magazine 716, 2020. <https://locusmag.com/2020/09/cory-doctorow-ip/>.
6. Fisher, Phineas. «HackBack - A DIY Guide For Those Without The Patience To Wait For Whistleblowers». Exploit Database, 8 agosto 2014. <https://www.exploit-db.com/exploits/41913>.
7. Kernighan, Brian W., e P. J. Plauger. *The Elements of Programming Style*. 2nd Edition. New York: McGraw-Hill, 1978.
8. Knuth, Donald E. «Computer science and its relation to mathematics». *The American Mathematical Monthly* 81, n. 4 (1974): 323–343.
9. Krug, Steve. *Don't make me think!: Web & Mobile Usability: Das intuitive Web*. MITP-Verlags GmbH & Co. KG, 2018.
10. Lessing, Lawrence. *Code: Version 2.0*. New York: Basic Books, 2006. <http://codev2.cc/download+remix/Lessig-Codev2.pdf>.
11. Mahieu, René, e Jef Ausloos. «Recognising and Enabling the Collective Dimension of the GDPR and the Right of Access», 2020.
12. Mentor. «The Conscience of a Hacker». *Phrack Inc* 1, n. 7 (1986).
13. Morozov, Evgeny. «The Meme Hustler: Tim O'Reilly's Crazy Talk». *The Baffler*, n. 22 (2013): 66–147.
14. Snowden, Edward J. *Permanent Record*. London: Metropolitan Books, 2019.
15. Stallman, Richard. «Viewpoint Why "open source" misses the point of free software». *Communications of the ACM* 52, n. 6 (2009): 31–33.





# IoRestoACasa.work - isolati ma non soli

Luca Ferroni<sup>1</sup>, Massimo Carboni<sup>2</sup>  
beFair – luca [at] befair.it  
GARR – massimo.carboni [at] garr.it

## PAROLE CHIAVE

videoconferenze, smartworking, smartlearning, community, webRTC, banda ultralarga

## INTERVENTO

Il progetto IoRestoACasa.work nasce nel periodo di lockdown per mettere a disposizione una rete di server per video-chiamate di gruppo, libere, private e gratuite, per continuare a studiare, lavorare, socializzare nel periodo del lockdown.

Il progetto nasce su iniziativa del team beFair, una rete di professionisti dell'IT di Fabriano già attivi nelle community del software libero con l'associazione PDP Free Software User Group e il MakerSpace della Biblioteca di Fabriano. È proprio da queste esperienze e da questo background comune che si sono ritrovati a credere che grazie al software libero si possono innescare meccanismi virtuosi per la società.

Così IoRestoACasa.work predispone un sistema di facile installazione per Jitsi Meet e lancia un appello a tecnici, hacker, provider e datacenter italiani per mettere a disposizione risorse riscuotendo l'adesione di importanti soggetti. Tra i primi il provider SeeWeb, l'istituto di ricerca CNR e il GARR. È a quest'ultimo soggetto che si deve una fondamentale spinta al progetto: l'introduzione del software libero EduMeet sviluppato nell'ambito del progetto europeo GEANT.

Nei primi 15 giorni IoRestoACasa.work ha raccolto l'adesione di 46 server, che hanno effettuato 20.300 video-chiamate per 6.000 ore di videoconferenza. Da subito ha aggregato una community con 60 tecnici specializzati, fino ad arrivare ai 241 membri della chat Telegram attuale interessati a far crescere il progetto.

Si desidera ringraziare Richard Stallman per aver avviato il movimento del Software Libero, i professori Angelo Raffaele Meo e Renzo Davoli per aver trasmesso questi valori in Italia a tanti studenti in modo che potessero cogliere l'opportunità di essere hacker sociali, oltre che aver creato opportunità di lavoro ed innovazione tecnologica.

Si aggiungono i ringraziamenti ai membri più attivi della community Telegram che credono nel progetto e in particolare ai più attivi che offrono gratuitamente il loro servizio di supporto.



# Privacy Risk Assessment on Network Data

Roberto Pellungrini<sup>1</sup>, Anna Monreale<sup>2</sup>

<sup>1</sup> Università di Pisa – roberto.pellungrini [at] di.unipi.it

<sup>2</sup> Università di Pisa – anna.monreale [at] unipi.it

## ABSTRACT

In the modern Internet era, the usage of social networks such as Twitter, Instagram and Facebook is constantly increasing.

The analysis of this type of data can help us understand interesting social phenomena because these networks intrinsically capture the new nature of user interactions. Unfortunately, social network data may reveal personal and sensitive information about users, leading to privacy violations. As a consequence, before applying any analytical framework on social network data, it is important to empirically assess the privacy risk of users to identify risky data to be treated with appropriate privacy-preserving technique. In this paper, we propose a study of privacy risk for social network data. In particular, we empirically analyse a set of privacy attacks on social network data by using the privacy risk assessment framework PRUDence. After simulating the attacks on real data, we analyze the distribution of privacy risk.

## KEYWORDS

privacy, privacy risk, social networks

## TALK

### 1 INTRODUCTION

Social networks are used by people every day for different purposes: for interacting with friends (Facebook), for professional activities (LinkedIn), for spreading information, news and multimedia material (Twitter and Instagram). Nowadays, the analysis of social network data is fundamental to study and understand social phenomena. The social network analysis can help in understanding customer interactions and reactions [5], marketing strategies based on communities or singles users, and migration flows, fake news diffusion or virus spread [10]. However, social network data may contain sensitive and private information about the real people that actively operate in the network. Backstrom et al. [1] showed that basic anonymization techniques are not enough for privacy protection as malicious adversaries still may succeed in re-identifying individuals using a background knowledge attack. In order to enable a practical application of the privacy-preserving techniques proposed in the literature, Pratesi et al. [4] proposed PRUDence, a framework for systematic privacy risk assessment.

This framework follows the idea of the EU General Data Protection Regulation, which explicitly imposes on data controllers the responsibility of assessing privacy risk for data mining processes.<sup>1</sup>

In [4] Pratesi et al. shows the applicability of their framework on mobility data. In this paper, we propose to apply PRUDence framework for the privacy risk assessment in social network data. We show how PRUDence can be applied to social network data and what kind of insights we can gather from the analysis of privacy risk in social networks. We show that even when a social network is represented with aggregative data structures, it is still possible to conduct background knowledge attacks and re-identify individuals in the network. This requires to first formally define a set of privacy attacks on social network data, then simulate them on real data to empirically evaluate the individual privacy risks.

## 2 DATA DEFINITION

Social networks have traditionally been modelled as graphs  $G=(V, E, L, \Gamma)$  where  $V$  is the set of vertices representing individuals,  $E \subseteq V \times V$  is the set of edges representing the relationships between individuals,  $L$  is a set of labels, and  $\Gamma: V \rightarrow l$  is a labelling function that maps each vertex to a subset of labels  $l \subseteq L$ .

To keep our definition simple, we assume that edges do not have any labels and that all relationships are mutual (undirected graph). From the social network represented as a graph, it is possible to derive other data structures, representing aggregated information. The purpose of these data structures is to expose less information than the original form while enabling the computation of standard network metrics. Here, we define some of these structures:

- Friendship Vector  $F_v$  of an individual  $v \in V$  is a set of vertices  $F_v = \langle v_1, v_2, \dots, v_n \rangle$  representing individuals connected to  $v$  in the social network graph. It represents the neighbourhood of  $v$  at distance 1.
- Label vector of an individual  $v$  is a set of labels  $LA_v = \langle la_1, la_2, \dots, la_m \rangle$ . Each  $la_j = (f, l)$  with  $j \in \{1, 2, \dots, \vee L \vee\}$  is a pair of feature name  $f$  and label value  $l$ . The label vector of an individual can be empty. Each label describes a profile feature of an individual, such as gender or education etc.
- Degree vector of an individual  $v$ , denoted by  $D_v = \langle d_{v_1}, d_{v_2}, \dots, d_{v_n} \rangle$ , represents the number of neighbours of each neighbor of  $v$ . Thus,  $d_{v_i} = \text{len}(F_{v_i})$ .
- Mutual Friendship vector of an individual  $v$ ,  $MF_v = \langle mf_1, \dots, mf_n \rangle$  represents the number of common neighbours of  $v$  with each one of its neighbours  $v_i$ .

Given the definitions above, we can define a Social Network Dataset is a set of data structures  $S = \{S_1, S_2, \dots, S_k\}$  where  $S_v (1 \leq v \leq k)$  is the social network data structure of an individual  $v$ .

---

<sup>1</sup> [https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L\\_.2016.119.01.0001.01.ENG&toc=OJ:L:2016:119:TOC](https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2016.119.01.0001.01.ENG&toc=OJ:L:2016:119:TOC)

### 3 PRIVACY RISK ASSESSMENT FRAMEWORK

Given the rapid growth in the number of services and applications based on social networks, there is increasing concern about privacy issues in published social network data. The prevention of node/individual re-identification is one of the critical issues. With some background knowledge about an individual in a social network, an adversary may perform a re-identification attack and disclose the identity of the individual. To preserve privacy, it is not sufficient to remove all identifiers, as shown by Chih-Hua Tai et al. [9] or Bin Zhou and Jian Pei [10]. In this paper we want to empirically study the privacy risk in social network data using the PRUDence risk assessment framework [4] PRUDence enables a privacy-aware ecosystem for sharing personal data. In this framework, Data Providers must perform a privacy risk assessment before releasing the data to ensure the privacy of the individuals. The privacy risk assessment component of the framework produces a quantitative measure of privacy risk. Such a measure depends on the kind of privacy attack simulated, the kind of data, and on the aggregation on the data itself. The simulation of a privacy attack requires two phases: first, we assume that a malicious adversary gathers, in some way, background knowledge about an individual and then the adversary uses the acquired background knowledge to re-identify the individual in the social network dataset. Background knowledge is essentially a portion of an individual's data, used by the adversary to conduct an attack. It represents what kind of information the adversary has. The length of background knowledge is the number of elements known by the adversary. Since, in principle, an adversary may have any kind of *background knowledge*, a worst-case scenario analysis has to be conducted. To this aim, PRUDence allows for the systematic generation of *background knowledge instances* of a certain length are generated to simulate an attack, and the one *background knowledge instance* that presents the highest probability of re-identification for an individual is the final privacy risk for that individual. The probability of re-identification is computed according to the principles of *k-anonymity* [8]. Thus, the more common the background of an individual is, the harder it is to re-identify him. The probability of re-identification depends then on the actual number of individuals that share the same *background knowledge instance* that the adversary possesses. In order to simulate how the adversary matches the background knowledge against the data, the attack must be formalized as a *matching function*, i.e. a function that indicates whether a *background knowledge instance* is present in the data of a certain individual.

We can then compute the probability of re-identification as:  $PR_s(s=v|b) = \frac{1}{|M(S,b)|}$  where  $M(S,b) = \{s \in S | match(s,b) = True\}$ ,  $s \in S$  is a record in the dataset, and  $b$  is a *background knowledge instance*, and *match* is the matching function. With this definition, privacy risk can vary between 0 (no risk) and 1 (maximum risk). The attack simulation process for each individual in the data, is summarized in Figure 1.

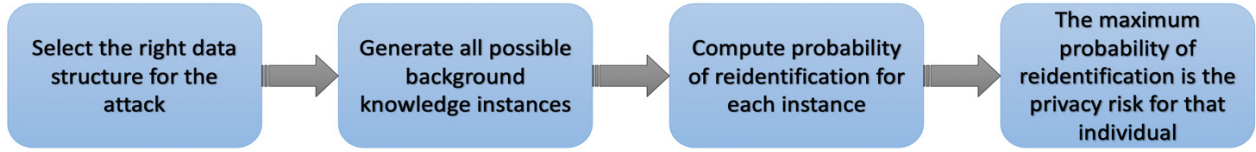


Figure 1: attack simulation process

Given the privacy framework we presented, the definition of an attack depends entirely on the matching function used to understand if a particular background knowledge instance can be found in the data structure of an individual. We define a set of attacks based on the data structures of a *Social Network Dataset* as previously defined.

- *Neighbourhood Attack*: the adversary knows a certain number of friends/neighbours of an individual. It was introduced by Chih-Hua et al. [9] The matching function is

$$match(b, F_v) = \begin{cases} true & \text{if } b \subseteq F_v \\ false & \text{otherwise} \end{cases}$$

- *Label Pair Attack*: the adversary knows a certain number of pairs of features with their values of an individual. It was introduced by Chenyang Liu et al. [3]. The matching function is

$$match(b, LA_v) = \begin{cases} true & \text{if } b \subseteq LA_v \\ false & \text{otherwise} \end{cases}$$

- *Friendship Degree Attack*: the adversary knows the degree of some neighbours of the victim and the degree of the victim. Introduced by Chih-Hua et al. [9]. The matching function is

$$match(b, D_v) = \begin{cases} true & \text{if } b \subseteq D_v \\ false & \text{otherwise} \end{cases}$$

- *Mutual Friend Attack*: the adversary knows the number of mutual friends of the victim and some of its neighbours. Introduced by Chong-Jing et al [7]. The matching function is

$$match(b, MF_v) = \begin{cases} true & \text{if } b \subseteq MF_v \\ false & \text{otherwise} \end{cases}$$

## 4 EMPIRICAL PRIVACY RISK ASSESSMENT

For experimental evaluation of the defined attacks, we use the Facebook Dataset provided by Stanford University's "Stanford Large Network Dataset Collection" [2]. This dataset includes node features (profiles), circles and ego networks. Nodes have been anonymized by replacing the Facebook-internal ids for each user with a new value.

Feature vectors from this dataset have also been provided while the interpretation of those features has been anonymized. After aggregating all data, we obtain a social network graph of 4039 nodes and 88,234 edges. Almost half of all individuals have 30 friends/neighbours or less. In Figure 2, we can see some summary of the data.

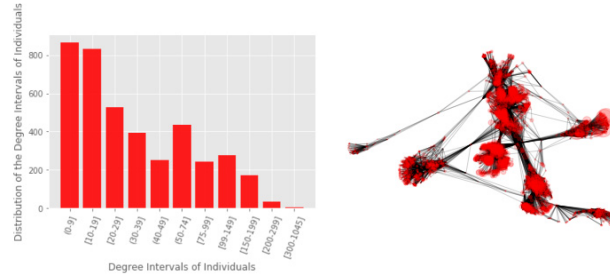


Figure 2: degree distribution of nodes in the network and visualization of the Facebook Dataset

We simulated the privacy attacks defined in Section 3 with various background knowledge length (1,2,3 and 4) and computer the cumulative distribution of privacy risk, varying the length of background knowledge, for each attack. The cumulative distribution indicates on the y-axis the portion of individuals that present a privacy risk lower than the respective risk value on the x-axis. Therefore, the lower the curve, the higher the number of individuals with risk near 1. A higher curve instead indicates lower levels of risk overall. We present some of the results in Figure 3.

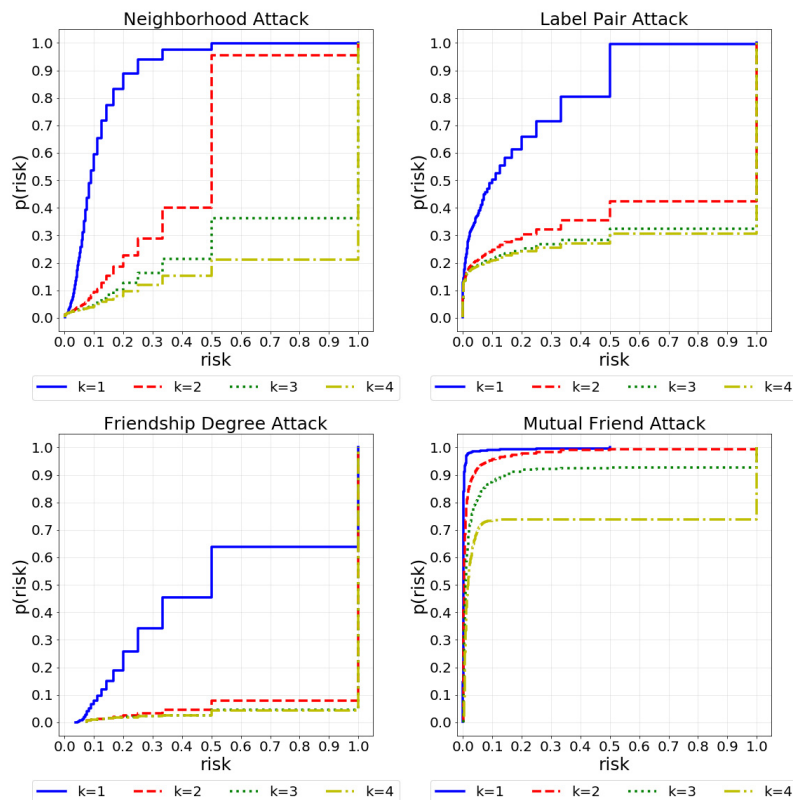


Figure 3: cumulative distribution of risk for privacy attacks on network data



From Figure 3, we can see that, unsurprisingly, increasing the length of the background knowledge, i.e. increasing the amount of information that the adversary has regarding the victim, the privacy risk also increases. However, we see that, after the first increase from  $k=1$  to  $k=2$ , privacy risk almost reaches a plateau. This behaviour has also been observed in other contexts, for example, in mobility data [11]. This suggests that increasing the length of the background knowledge does not raise privacy risk proportionally. Also, interestingly, we see that the Mutual Friend Attack is the least powerful of all attacks, while the Friendship Degree Attack is the most dangerous. These results show that it is possible to analyze privacy risk empirically on network data, obtaining information that is easy to understand and to use in order to protect the data and the users in it.

## REFERENCES

1. Backstrom, Lars, Cynthia Dwork, and Jon Kleinberg. ‘Wherefore Art Thou R3579X? Anonymized Social Networks, Hidden Patterns, and Structural Steganography’. In *Proceedings of the 16th International Conference on World Wide Web*, 181–190, 2007.
2. Leskovec, Jure, and Rok Sosič. ‘Snap: A General-Purpose Network Analysis and Graph-Mining Library’. *ACM Transactions on Intelligent Systems and Technology (TIST)* 8, no. 1 (2016): 1–20.
3. Liu, Chenyang, Dan Yin, Hao Li, Wei Wang, and Wu Yang. ‘Preserving Privacy in Social Networks against Label Pair Attacks’. In *International Conference on Wireless Algorithms, Systems, and Applications*, 381–392. Springer, 2017.
4. Pratesi, Francesca, Anna Monreale, Roberto Trasarti, Fosca Giannotti, Dino Pedreschi, and Tadashi Yanagihara. ‘PRUDence: A System for Assessing Privacy Risk vs Utility in Data Sharing Ecosystems’, 2018.
5. Rossetti, Giulio, Letizia Milli, Fosca Giannotti, and Dino Pedreschi. ‘Forecasting Success via Early Adoptions Analysis: A Data-Driven Study’. *PloS One* 12, no. 12 (2017): e0189096.
6. Rossetti, Giulio, Letizia Milli, Salvatore Rinzivillo, Alina Sîrbu, Dino Pedreschi, and Fosca Giannotti. ‘NDlib: A Python Library to Model and Analyze Diffusion Processes over Complex Networks’. *International Journal of Data Science and Analytics* 5, no. 1 (2018): 61–79.
7. Sun, Chongjing, S. Yu Philip, Xiangnan Kong, and Yan Fu. ‘Privacy Preserving Social Network Publication against Mutual Friend Attacks’. In *2013 IEEE 13th International Conference on Data Mining Workshops*, 883–890. IEEE, 2013.
8. Sweeney, Latanya. ‘K-Anonymity: A Model for Protecting Privacy’. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 10, no. 05 (2002): 557–570.
9. Tai, Chih-Hua, Philip S. Yu, De-Nian Yang, and Ming-Syan Chen. ‘Privacy-Preserving Social Network Publication against Friendship Attacks’. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1262–1270, 2011.
10. Zhou, Bin, and Jian Pei. ‘Preserving Privacy in Social Networks against Neighborhood Attacks’. In *2008 IEEE 24th International Conference on Data Engineering*, 506–515. IEEE, 2008.

# Digital Humanities, Education and Cultural Heritage across Academic and Museum Settings

Giuditta Cirnigliaro<sup>1</sup>, Angelica Federici<sup>2</sup>, Luca Marcozzi<sup>3</sup>, Manfredi Merluzzi<sup>4</sup>

<sup>1</sup> Università Roma Tre - giuditta.cirnigliaro [at] uniroma3.it

<sup>2</sup> Università Roma Tre - angelica.federici [at] uniroma3.it

<sup>3</sup> Università Roma Tre - luca.marcozzi [at] uniroma3.it

<sup>4</sup> Università Roma Tre - manfredi.merluzzi [at] uniroma3.it

## ABSTRACT

The wide variety of research practices described as Digital Humanities are unified by their emphasis on creating, connecting, interpreting and collaborating. In the last decade, the relationship between technology and the humanities has grown significantly transforming itself into something of a symbiosis. Being aware of the humanities' capability of evolving as a discipline by using technological knowledge to reach a wide and varied audience, the Humanities Department (DSU) community at Roma Tre University founded the Digital Humanities Working Group and Laboratory in 2018 in order to promote transdisciplinary collaborations among the academic and museum environments, with the aim to reach a wider public and create contents available for cultural heritage fruition, enhancement, and preservation. The DH Lab is currently developing or participating in wider research projects, such as VR Sant'Agnese, Petrarch's ITINERA and LiLeo Digital Atlas, which focus on the creation and presentation of cultural heritage resources to a wider public, in collaboration with prominent cultural organizations and museum institutions. These projects showcase the ability of Digital Humanities to challenge all disciplines of Art History, Curatorial, and Preservation Studies ultimately to engage with new interdisciplinary methodologies, innovative technologies, and original approaches to pedagogy.

## KEYWORDS

infosphere, digital education, digital public humanities, digital humanities and traditional tools, cultural heritage preservation, academic and museum settings, transdisciplinarity

# TALK

## 1 INTRODUCTION

The increasing diversification and specialisation in the analysis of literary, artistic and scientific sources have favoured the birth over the last century of multiple disciplines focusing on projects in the fields of either Italian Studies, Art History, or Computer Science. Wholly transdisciplinary research is still needed to initiate a dialogue about these disciplines, as the recent rise of Digital Humanities initiatives and laboratories testifies. Often considered just as technical tools applied to humanistic thought, Digital Humanities offer the unique opportunity to introduce a common methodological vision in the context of the more established literary-artistic and technical-scientific studies. Moreover, Digital Humanities projects are a social enterprise that favour inclusivity and dissemination of research [2; 3].

Indeed, the phrase Digital Humanities describes the humanities in the plural. But, as heterogeneous as it is, Digital Humanities is unified by its emphasis on creating, connecting, interpreting and collaborating [1]. DH is concerned not only with the use of digital technology for humanities projects but also with how the use of digital technology for humanities projects changes the user experience. We are convinced that this sector is not only interesting, not only culturally and artistically relevant, not only of great potential for increasing the audience of museums, touching people who, thanks to certain technological shortcuts, will be able to stop considering cultural experience as something detached from itself. We are convinced that this sector is something more: that it is strategic for the country as a system [6].

In the last decade, the relationship between technology and the humanities in contemporary society has grown significantly transforming itself into something of a symbiosis. In time, the humanities have been capable of evolving as a discipline using technological knowledge to reach a wide and varied audience [5]. The synergy that exists between technology and humanities has also been brought to the university by promoting exchange and dialogue between different Departments. We may go as far as saying that the synergy between the departments together with the accessibility brought by the interaction of the humanities and technology has made the university a more inclusive, democratic and globalised pole. For example, in the field of Art History, the DH were born to unite and share knowledge on medieval artistic artefacts which are often of difficult interpretation and complex to explain. The Nasher Museum Project attempted to integrate detached fragments of architecture into a virtual space that would restore its original function. We may go as far as saying that the Discipline has left the academic circle – to join societal needs for change and inclusivity [4].

Italy is among the leaders in the field of Digital Cultural Heritage or the use of cultural heritage through the use of new technologies. Exhibitions and multimedia events are now a fundamental driver in the field of edutainment. Cultural heritage breathes a new life with digital technologies. Institutions can reach a wider audience, and develop creative and accessible content for the enjoyment, enhancement and conservation of cultural heritage<sup>1</sup>.

---

1 <https://ec.europa.eu/digital-single-market/en/digital-cultural-heritage>

Following the new direction in Digital Cultural Heritage promoted by the European Commission, the Humanities Department (DSU) community at Roma Tre University, under the direction of Professor Manfredi Merluzzi, founded the Digital Humanities Working Group and Laboratory in 2018 in order to promote transdisciplinary collaborations among the academic and museum environments, with the aim to reach a wider public and create contents available for cultural heritage fruition, enhancement, and preservation. DH projects inside the university were selected by the lab in accordance with best practice research standards as a methodological benchmark for future investigations in the field. These include but are not limited to the meticulous study of primary sources, attentive scrutiny of extant artefacts, factual display of 3D reconstructions. Indeed the objective of these projects is to use new digital tools in a broader setting, with the scope of reaching broader and diverse users to breach the gap between specialised investigators and laymen. This holistic methodology will hopefully democratise the humanities fostering e-equality, participation, rights and values in the digital age.

## 2 THE ROMA TRE DIGITAL HUMANITIES LAB

The Digital Humanities Laboratory (DH Lab) at Roma Tre University merges new approaches in the digital humanities with traditional object-based research, by promoting a broad range of interdisciplinary projects and initiatives at the core of the Humanities Department mission. From the onset of the lab's foundation, a dialogue was initiated between different research branches inside the Department. The Roma Tre DH Lab supports collaborative work across different departments and divisions, such as the Cultural Heritage, Art and Literature, Engineering, History, History of Science and Geography Laboratories, and is equipped with a multimedia archive for both research and teaching in the Digital and Public Humanities.

After hosting prestigious initiatives, such as the web portal ALI<sup>2</sup>, the DH Lab is currently developing or participating in wider research projects which focus on the creation and presentation of cultural heritage resources to a wider public, in collaboration with prominent cultural organizations and museum institutions.

The current research strands of the Roma Tre DH Lab are represented by three projects, namely, *VR Sant'Agnes: The Virtual Reconstruction of a Medieval Monastery* in collaboration with RilTec laboratory, Department of Architecture; *Petrarch's ITINERA: Italian Trecento Intellectual Network and European Advent*, which is a part, based in Roma Tre, of a PRIN National research led by Siena University, and *LiLeo Digital Atlas: Leonardo Word-and-Image Digital Editions*, developed in collaboration with the Museo Galileo in Florence. Each of the DH Lab representative projects uses model open-source, web-based, and interoperable technologies.

In particular, *VR Sant'Agnes* expands traditional building drawing design by incorporating digital visualization software tools to transpose the 2D drawing into a 3D Virtual Reality space rendering;

---

2 ALI is a web portal for the study of the autographs of Italian writers from the Origins until the end of the sixteenth century. The portal is the result of the work of literary historians, philologists, historians of language, palaeographers, librarians and archivists.  
[www.autografi.net](http://www.autografi.net)

*Petrarch's ITINERA* employs Cadmus and Semantic Web to catalogue and track the movements of the poets' correspondences thus transforming the written word into semantic web diagrams; *LiLeo Digital Atlas* experiments with Omeka, Project Mirador, and IIIF to collect, annotate, and share original sources, and employs abstract models, which are the product of scholarly interpretation, to build a database of patterns, by emphasizing and making them explicit<sup>3</sup>.

*VR Sant'Agnese*, *Petrarch's ITINERA* and *LiLeo Digital Atlas* are highly interdisciplinary projects, which are developed by a team of international scholars specialized in either textual, visual, and scientific fields of investigation, web designers and developers, library science technical staff, and research assistants with experience in Digital Humanities, information studies, and computer science. A distinctive feature of these initiatives is their user-centricity and educational output and implications.

Being aware of the fundamental role that DH has in education, we encourage students to participate in the life of the DH Lab, by collaborating on these projects as academic assistants under the direction of the technical staff. For instance, students have been involved in data collection and organization of the image files to be included in the *LiLeo Digital Atlas*, and in the creation of derivative images containing selected texts and pictures from the original files, and spreadsheets for tracking project metadata and for bulk upload of items to the digital database. In addition, students of the Renaissance Literature course (MA) will use Cadmus to catalogue part of Petrarch's correspondence to be included in the *ITINERA* database. Under the direction of the community of scholars and specialists from Roma Tre and partner institutions, students could learn technical skills to apply to their future work and could also serve as an ideal audience for the projects themselves. In this way, Digital Humanities creates bonds between universities and cultural institutions, such as the Galileo Museo in Florence, and the National Museum of Science and Technology "Leonardo da Vinci", where students could conduct professional work during the 2019 Roma Tre Digital Humanities Summer School edition.

By stimulating differentiated and active learning both inside and outside of the class (e.g., the museum environment), Roma Tre DH Lab projects enhance the variety of current students' educational and cultural backgrounds and skills. Student collaborations in the DH Lab projects are always linked to the courses and workshops which are part of their study plan, and therefore perfectly integrated in their academic curriculum. In this perspective, *Petrarch's ITINERA* would be the preferred environment for students to interpret some of the primary sources and cultural context of the Italian Literature course and the seminars on the poet.

Thanks to the comparative analysis and interactive maps tools, students would be able to interpret and connect different texts, and to track the paths of writers and their correspondences across time and

---

3 Omeka is a free open source web-publishing platform for sharing digital collections and creating media-rich online exhibits in library, museum, archive, and academic settings. Project Mirador is an open-source, web based, multi window image viewing platform with the ability to zoom, display, compare, and annotate images belonging to different repositories. IIIF (International Image Interoperability Framework) is a growing community of the world's leading research libraries and image repositories, which collaborate to produce an interoperable technology and community framework for image delivery. IIIF provides uniform and rich access to image-based resources hosted around the world; it also develops interfaces that support interoperability between image repositories and technologies for viewing, comparing, manipulating and annotating images.

<https://omeka.org/>

<https://projectmirador.org/>

<https://iiif.io/>

space. Furthermore, participation in international academic and public institutional projects constitutes an invaluable experience for knowledge growth and civil awareness of students, which are faced with the public uses of history and culture. In this way, student learning is modelled on the problems and aims of Public and Digital History and not on a simplistic competence-based approach.

We hope that research projects like *VR Sant'Agnese* may foster the future involvement of students, to physically interact with the architectural spaces that they study in their art history classes, and visualize their function and implications thanks to object-based spatial analysis and 3D virtual model reconstruction.

### **3 CONCLUSION**

The presentation of the current research strands of the Roma Tre DH Lab has shown that Digital Humanities offers multiple tools for data visualization and examination, which allow to question historical research on multiple levels and encourage different possibilities of making use of history. In addition, digital and public history research and learning initiatives facilitate cultural heritage preservation and fruition and open culture circulation. We believe that the ongoing projects are undertaken at the Lab showcase the ability of Digital Humanities to challenge all disciplines of Literature, Art History, Curatorial, and Preservation Studies ultimately to engage with new interdisciplinary methodologies, innovative technologies, and original approaches to pedagogy. Digital imaging, manuscript encoding and visual annotation tools – including 3D Modelling, Virtual/Augmented Reality and Image Interoperability Framework Viewer – revolutionize our critical engagement with the arts and the humanities in both the academic and museum environments. This agenda has deeply influenced the museum world with new innovative projects that catalyzed the need for inclusivity and dissemination.

### **4 ACKNOWLEDGEMENT**

The authors wish to thank Professors Giulia Bordi and Maurizio Fiorilla, and the Roma Tre Digital Humanities Laboratory collaborators for their insights throughout the development of the projects and for institutional support. A special thanks to Professor Marco Canciani, Director of RilTec Laboratory at Roma Tre University, for his invaluable support in collecting data through laser scanning and point cloud meshes in the *VR Sant'Agnese* project; Carlo Vecce, Professor Paolo Galluzzi, Dr Andrea Bernardoni, Dr Dinni Rolfo, Dr Francesca Giannetti, and the staff of the Museo Galileo in Florence and the Digital Humanities Lab at Rutgers (NJ), who took the time to discuss these projects in their early stages and offered helpful commentary.

## REFERENCES

1. Burdick, Anne, Johanna Drucker, Peter Lunenfeld, Todd Presner, and Jeffrey Schnapp. *Digital\_Humanities*. Mit Press, 2012.
2. Foster, Ian. 'How Computation Changes Research'. *Switching Codes: Thinking through Digital Technology in the Humanities and the Arts*, 2011, 15–37.
3. Kirschenbaum, Matthew G. 'What Is Digital Humanities and What's It Doing in English Departments?' In *Defining Digital Humanities*, 211–220. Routledge, 2016.
4. Pinotti, Andrea, and Antonio Somaini. *Cultura Visuale: Immagini, Sguardi, Media, Dispositivi*. Vol. 653. Einaudi Torino, 2016.
5. Soncini, Elisa. *Memorie Sociali, Memorie Mediali. La Costruzione Del Passato Nell'era Delle Comunicazioni Di Massa*. Milano: Franco Angeli, 2008.
6. Tonelli, Massimiliano. 'Mostre Multimediali: Una Risorsa Strategica?' *Artribune Magazine* 38 (2017).

## APPENDIX: ROMA TRE DH RESEARCH PROJECTS

### VR SANT'AGNESE.

The Virtual Reconstruction of a Medieval Monastery combines established tools of scholarly enquiry including but not limited to archival research, empirical display of extant objects, and meticulous study of building structures with photogrammetry/point cloud, building information modelling and virtual reality. This project aims to reconstruct the medieval facies of the monastery of Sant'Agnese fuori le mura in Rome in an integrated study of architecture, liturgical furnishing and painting. The transformation of physical spaces over time was originally approached through traditional building design, largely reliant upon two-dimensional technical drawings (plans, elevation, sections).

Building information modelling extends our understanding of art and architecture by augmenting the three primary spatial dimensions (width, height, and depth) with time.

The instalment of female communities in existing structures, the loss of a significant number of convents, and the perishable nature of the materials used has frequently impaired the study of female monastic architecture. When it comes to women religious fixed architectural schemes cannot be traced on the basis of surviving elements.<sup>4</sup>

Starting from the pioneering experience of the *Atlante percorsi visivi* (La pittura medievale a Roma. 314-1431) by Maria Andaloro, aimed at recontextualizing murals and mosaics in the churches of Rome with the help of floor plans and axonometric views, recreating specific architectural contexts, this study aspires to a precise spatial analysis that goes beyond the descriptive limit of photographic reproductions.<sup>5</sup> By creating a 3D virtual model of the monastery in its late medieval facies, this application aims to be both a research tool and aid to visiting the monument (Figure 1).<sup>6</sup> Ultimately, the outcome of this project will be to engage with architectural history with both new and traditional methodologies, by challenging the way we study and visualize the change of space over time. In addition to publishing our findings, we intend to create an open-access platform which will allow downloading these immersive digital environments. The greater/future scope of this project is to both track the benefits in the pedagogy of multiple users in a 3D virtual reality medium, which in many fields have already replaced two-dimensional flat screen teaching and to study the impact of digital reconstructions in relation to the dissemination of cultural heritage to impaired visitors.<sup>7</sup>

---

4 Zappasodi, Emanuele. *Sorores reclusae. Spazi di clausura e immagini dipinte in Umbria fra XIII e XIV secolo*, Mandragora, 2018; Bruzelius, Caroline. "Hearing is Believing Clarissan Architecture, ca. 1213-1340", *Gesta* 31/2 (1992): 83-91; Bruzelius, Caroline. "Nuns in Space: Strict Enclosure and the Architecture of the Clarisses in the Thirteenth Century" in Ingrid Peterson, ed., *Clare of Assisi: A Medieval and Modern Woman, Clarefest Selected Papers*, vol. III, Columbia University Press, 1996, 53-74.

5 Andaloro, Maria. *La pittura medievale a Roma, Atlante percorsi visivi 1*, Jaca Book, 2006.

6 Burdick, Anne, Johanna Drucker, Peter Lunenfeld, Todd Presner, and Jeffrey Schnapp. *Digital Humanities*. The MIT Press, 2012; Schreibman, Susan, Ray Siemens, and John Unsworth. *A Companion to Digital Humanities*, Blackwell Publishing, 2004.

<https://www.oculus.com>

<https://vr.google.com/cardboard/>

7 <https://www.cam.ac.uk/research/news/virtual-reality-journey-through-a-tumour-cambridge-scientists-receive-ps40-million-funding-boost>



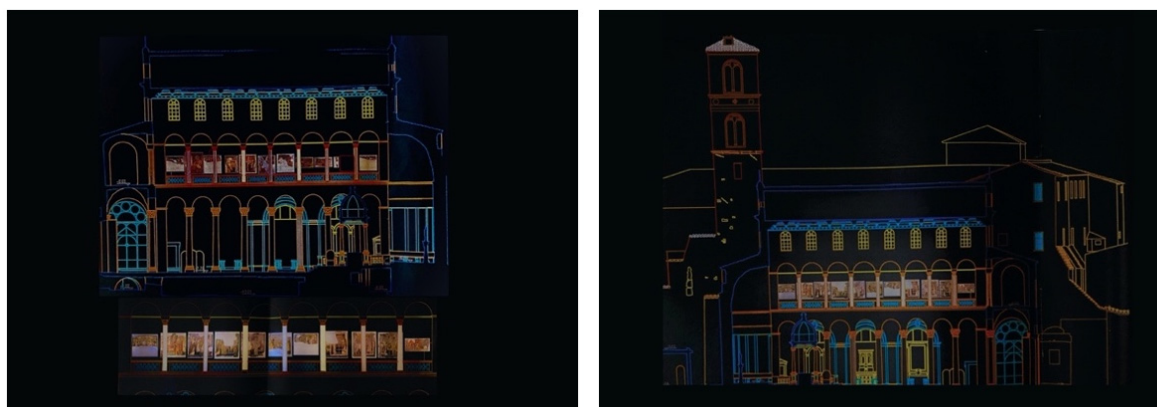


Figure 1-2: axonometric View of Sant'Agnese fuori le mura with virtual restoration of detached fresco cycle.

### PETRARCH'S ITINERA.

Petrarch's influence durably shaped European literary and cultural consciousness. His figure is one of crucial importance in a number of fields – literary, intellectual, ethical and philosophical, political and diplomatic – where the transformation of medieval thought into the intellectual mindset of the Renaissance is noticeable. Petrarch's correspondence in particular preserves overwhelming evidence of his engagement with the contemporary world, by encompassing a wide range of recipients including prominent figures from the literary and the political establishment, as well as individuals of intellectual, ecclesiastical, juridical and diplomatic distinction.<sup>8</sup> Petrarch's Latin letters are identified as the project's primary object of investigation as they provide a matchless overview of fourteenth-century European culture and society – from England to Germany, from Sicily to Provence and Bohemia. The texts will be systematically card-indexed by means of a standard form, which will enable users to cross-check the collected data in the most flexible and extensive way. In particular, Roma Tre unit, coordinated by Luca Marozzi, will achieve a card-indexing of Petrarch's correspondents and participate to the setting up of the database and to the output of results and their interrogation. This activity has been guided so far by prof. Paul Weston (Pavia University), who is working in cooperation with the Siena Unit, directed by Natascia Tonelli. Students and junior researchers from Roma Tre will record and analyze data in the framework of the DH Lab.<sup>9</sup>

The materials will be made available in an open-access database, designed according to criteria of scalability, portability and interoperability and suitable for rapid and intuitive consultation. This last point has special importance, insofar as the materials stored on digital platforms are making a decisive contribution to the study of Italian and Latin medieval literature and more generally to the advancement of historical-philological disciplines.

---

8 Marozzi, Luca, and Francisco Rico. Petrarca, Francesco, in *Dizionario biografico degli Italiani*, vol. 82, Istituto della Enciclopedia Italiana, 2015, 671-84.

9 Recent significant digital projects on Petrarch are Petrarchive: An Edition of Petrarch's Songbook "Rerum Vulgarium Fragmenta", and RDP. Le "rime disperse" di Petrarca: l'altra faccia del Canzoniere.

<http://dcl.slis.indiana.edu/petrarchive/>

<https://www.unige.ch/petrarca/it/homepage/>

The database and its biographical entries constitute the project's core business, with its multiple search options, its cross-referencing pathways and interactive maps developed along with the principles of literary cartography. The database will also include data on the poems of correspondence and its protagonists, the texts themselves and the information on their manuscript and printed tradition, accompanied with HD reproductions of the most important manuscripts (a task of the Napoli and Perugia units, coordinated by Marco Cursi and Daniele Piccini), all interconnected by a system of hypertextual links and cross-references. The ultimate objective is the production of an atlas detailing the circulation of humanistic knowledge, ideas, genres and literary forms in the transition from the Middle Ages to the Renaissance.

#### LILEO DIGITAL ATLAS.

Leonardo Word-and-Image Editions focuses on Leonardo da Vinci's writings and drawings, and investigates their relationship with scientific studies, in order to make sense of the artist's aims and combinatory devices.<sup>10</sup> By designing a diagrammatic model that enables the juxtaposition and layering of drafts belonging to similar projects in Leonardo's works and sources, LiLeo Digital Atlas tracks the evolution of forms through different media and underlines recurrent patterns in his output.

The identified diagrams are digitized as part of the Omeka digital archive Lileo<sup>11,12</sup>, which is the basis for the creation of the hyper-medial edition LiLeo Digital Atlas, developed at Roma Tre University, in collaboration with the Galileo Museum of Florence<sup>13</sup>.

As part of the LiLeo Digital Atlas project, subjects that feature both in Leonardo's codices and in his textual and visual sources are catalogued and grouped according to their typology, format, and field of investigation. The goal of the project is to digitally enable the juxtaposition and layering of drafts belonging to similar projects in Leonardo's works and sources, in order to identify aesthetic and scientific patterns in his output, uncover the origins of his interdisciplinary research, and reveal his creation of a hieroglyphic language.<sup>14</sup> LiLeo Digital Atlas will show digital editions on the work of Leonardo, his sources (Aesop, Pliny, etc.) and the work of his contemporaries (Giuliano da Sangallo, Francesco di Giorgio), which feature related visual and textual structures (Figure 3-4).

---

10 Barkan, Leonard. *Michelangelo: A Life on Paper*, Princeton University Press, 2011, 2-14; Zwijnenberg, Robert. *The Writings and Drawings of Leonardo da Vinci: Order and Chaos in Early Modern Thought*, Cambridge University Press, 1999; Cernigliaro, Giuditta, and Carlo Vecce. *Leonardo da Vinci: Favole e profezie. Scritti letterari*, Garzanti, 2019.

11 <https://blogs.libraries.rutgers.edu/lileo/>

12 Cernigliaro, Giuditta. "The Digital Reconstruction of Leonardo's Library: Revealing Formal Patterns in Early Modern Thought." In Special Issue on "Digital Humanities for Academic and Curatorial Practice", *Studies in Digital Heritage* 3, 2 (2019): 128-143.

13 Exemplary work with the aim to broaden access to Leonardo's codices is represented by e-Leo Digital Archive of History of Technology and Science, and "La Biblioteca di Leonardo".

<https://www.leonardodigitale.com>

<https://bibliotecadileonardo.museogalileo.it/>

14 Cf. Vecce, Carlo. "Parola e immagine nei manoscritti di Leonardo da Vinci." In *Percorsi tra parole e immagini (1400-1600)*, edited by Angela Guidotti, and Massimiliano Rossi, Maria Pacini Fazzi, 2000, 19-35; Cursi, Marco, "Leonardo da Vinci: Books." In *Collecting, Organizing, and Transmitting Knowledge. Miscellanies in Late Medieval Europe*, edited by Sabrina Corbellini, Giovanna Murano, and Giacomo Signore, Brepols, 2018, 233-248.

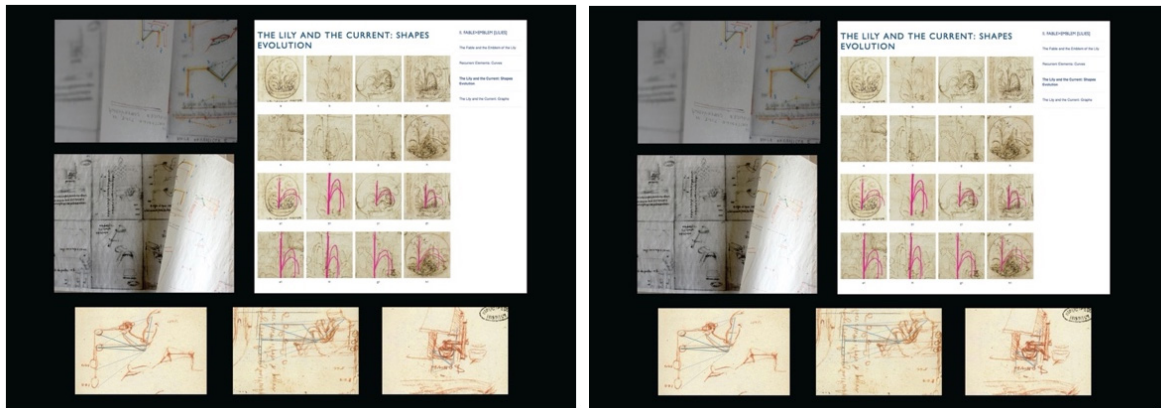


Figure 3, 4: LiLeo Digital Atlas diagrammatic tables and digital tools ‘Visualize>word/image’ and ‘Connect>Leonardo/sources’

By combining new approaches in the digital humanities with traditional artistic tools, LiLeo Digital Atlas expands the work of the already existing digital archives on Leonardo’s codices; it provides original insights for early modern research on intertextual languages and makes original sources of early modern culture available to a wider public. LiLeo Digital Atlas is an interoperable, durable and scalable digital product – which not only explores Leonardo’s transdisciplinary modes of investigation, but it also opens to multiple uses across different forms, media, and fields of study.

**Sessione V**  
**Roberto Busa**



# OCR Correction for Corpus-assisted Discourse Studies: A Case Study of Old Newspapers

Dario Del Fante<sup>1</sup>, Giorgio Maria Di Nunzio<sup>2</sup>

<sup>1</sup> Università di Padova – dario.delfante [at] phd.unipd.it

<sup>2</sup> Università di Padova – giorgiomaria.dinunzio [at] unipd.it

## ABSTRACT

The use of OCR software to convert printed characters to digital text is a fundamental tool within diachronic approaches to Corpus-assisted discourse Studies. However, OCR software is not totally accurate, and the resulting error rate may compromise the qualitative analysis of the studies. This paper proposes a mixed qualitative-quantitative approach to OCR error detection and correction in order to develop a methodology for compiling historical corpora. We present a case study on newspapers of the beginning of the 20<sup>th</sup> century for the linguistic analysis of the metaphors representing immigrants.

## KEYWORDS

corpus-assisted discourse studies, OCR detection, OCR correction

## TALK

### 1 INTRODUCTION

In Corpus-assisted Discourse Studies [9], the processes of corpus design and corpus compilation have a marked impact on the entire research and, depending on it, the results may shift dramatically. Especially for diachronic studies, there is a scarcity of digitized version of paper documents, and it is often necessary to manually transcribe the texts under analysis or to use Optical Character Recognition (OCR) software which have a fundamental role in the study of digitizes manuscripts [6]. However, OCR errors may significantly affect the compilation of a corpus in CADs [2]. There are procedures which are adopted to correct OCR errors [1] may not work properly in those cases where the quality of the scan is poor. In this paper, we propose a replicable semi-automatic method for detection and correction of OCR errors. The outcome of this project consists of a set of rules which are, eventually, valid for a different context and applicable to different corpora and which can be reproduced and reused.

The proposed procedure, in terms of computational readability, is aimed at making more readable and searchable the vast array of historical text corpora which are, at the moment, only partially usable given the high error rate introduced by an OCR software.

## 2 SEARCHING FOR METAPHORS TO REPRESENT IMMIGRANTS IN 1900

Our case study is the analysis of the metaphors used in the newspapers to represent migration to/from the United States of America and Italy from a diachronic perspective since the beginning of the XX century. Given the limited space available in this paper, we will refer to only one particular moment in history which have a significant value in relation to migratory movement: from 1900 to 1914, just before World War I, because this time period, as in [4], is a particularly significant moment for migratory movements, specifically from Europe to the U.S. The availability of data, the newspaper political leaning, and the registration fees were additional constraints that narrowed the selection of the newspapers to the New York Herald<sup>1</sup> for the U.S., and La Stampa<sup>2</sup> for Italy.

Afterwards, we needed to select the keywords to filter the articles useful for our study. The starting point for English was the set of words identified by [5] named under the acronym RASIM: *refugee\**,<sup>3</sup> *asylum seeker\**, *immigrant\**, and *migrant\**. We added a fifth word to this list: *emigrant\**. As for Italian, we needed to select a set of comparable search terms between English and Italian. We consulted the diachronic Diacoris Corpus,<sup>4</sup> a 15 million words collection of Italian texts produced between 1861 and 1945. The best candidate translation for migrant, immigrant and emigrant were *migrant\**, *immigrat\**, *immi-grant\**, *emigrant\**, *emigrat\**; for refugee and asylum seeker the candidate Italian terms were *rifu-giat\**, *profug\**, *clandestin\** and *richiedent\* asil\**.

In the first two rows of Table 1, we show a summary of the statistics for each compiled corpus. The tokens and types of values represent the total number of occurrences versus the number of unique words, respectively. We report both the type/token ratio (TTR) and the standardized type/token ratio (STTR).

## 3 OCR ERROR DETECTION AND CORRECTION

In this section, we propose a semi-automatic mixed approach to OCR detection, which brings together the dictionary-based and the context-based approaches. A careful analysis of a sample of texts showed that there were a lot of misspellings or non-meaningful words in both corpora caused by the OCR software. The first problem in our case study concerns the fact that we did not have the corresponding ground truth version of the corpora. Therefore, we decided to compile two contemporary newspaper corpora whose texts were digitalized since the beginning: The New York Times for the U.S., La Stampa for Italy (see last two rows of Table 1).

---

1 <http://chroniclingamerica.loc.gov>

2 <http://www.archiviolaStampa.it>

3 We use the symbol ‘\*’ to indicate the possibility of plural, or feminine/masculine for the Italian words.

4 <http://corpora.dslo.unibo.it/DiaCORIS/>

The error detection correction task consisted of a three-step procedure:

1. Detection of errors by comparing the list of words of the old corpus with the new corpus. The words that do not appear in the latter, or that have a statistically significant difference in frequency compose a list of plausible error candidates.
2. Analysis and categorization of the error in the list of candidates: i) an error containing the same number of characters than the respective correct form.; ii) an error containing a higher/smaller number of characters than the correct form; iii) a word interpreted by the OCR as two distinct words (i.e., ‘department’ vs ‘depart’ and ‘ment’).

Define the error correction rule as a regular expression.

The implementation of these procedure follows the principles described by [10] where the idea is to mine textual information from large text collections in an efficient and effective by means of pipelines allowing for a sequential process of text analysis. For our experiments, we used the R programming language, which has a set of packages, named ‘tidyverse’<sup>5</sup>, that implements this idea<sup>6</sup>. A total of 2,313 errors for English and 269 errors for Italian have been individuated and, respectively, as many correcting rules have been written for each language.

| Corpus          | Years     | Documents | Tokens     | Types     | TTR   | STTR   |
|-----------------|-----------|-----------|------------|-----------|-------|--------|
| New York Herald | 1900-1914 | 8,540     | 55,796,968 | 2,326,897 | 4.17% | 50.24% |
| La Stampa       | 1900-1914 | 3,092     | 18,773,664 | 817,865   | 4.36% | 56.63% |
| New York Times  | 2000-2014 | 125       | 58,915,060 | 308,251   | 0.52% | 48.39% |
| La Stampa       | 2000-2014 | 62        | 15,332,063 | 275,103   | 1.79% | 62.78% |

Table 1: statistics

In general, it is not easy to predict in what way OCR correction will work (see Table 2). On one side, the Italian corpus dimensions have been increased in relation to the number of tokens. The increase of tokens might be because many errors were not previously recognized as valid tokens. On the other side, the English corpus dimension has been decreased in relation to the number of tokens. The decrease might be due to the correction of split errors, such as *depart* and *ment*, corrected in *department*.

| Corpus                  | Tokens (B) | Types (B) | Tokens (A) | Types (A) | Δ Tokens | Δ Types |
|-------------------------|------------|-----------|------------|-----------|----------|---------|
| New York Herald 1900-14 | 55,796,968 | 2,326,897 | 55,555,708 | 2,323,790 | -0.43 %  | -0.13 % |
| La Stampa 1900-14       | 18,773,664 | 817,865   | 18,778,210 | 817,858   | 0.02%    | -0.001% |

Table 2: Statistics about errors before (B) and after (A) OCR corrections

5 <https://www.tidyverse.org>

6 <https://github.com/gmdn>



## **4 FINAL REMARKS AND FUTURE WORK**

In this paper, we presented a semi-automatic method for detection and correction of OCR errors for the discourse analysis of old newspaper documents. The outcome of this project consists in a set of rules which are, eventually, valid for a different context and applicable to different corpora and which can be reproduced and reused. There are still open questions that we will investigate in this line of work: how many documents have we missed during the compilation of the corpus given that a search keyword may be subject to OCR correction as well? How these types of keyword search error can affect a CADS analysis? For this reason, we intend to use error models to predict the relative risk that queried terms mismatch targeted resources due to OCR errors, as suggested by [3]. We also want to compare our analysis with other approaches that make use of BERT pre-trained neural networks to post-hoc error correction [8], especially in those cases where the context is not clear given multiple OCR errors in the same paragraph, or that take advantage of multiple OCR engines by aligning and comparing their different outputs in order to reduce the error rate [7].

## REFERENCES

1. Bassil, Youssef, and Mohammad Alwani. ‘Ocr Post-Processing Error Correction Algorithm Using Google Online Spelling Suggestion’. *ArXiv Preprint ArXiv:1204.0191*, 2012.
2. Bazzo, Guilherme Torresan, Gustavo Acauan Lorentz, Danny Suarez Vargas, and Viviane P. Moreira. ‘Assessing the Impact of OCR Errors in Information Retrieval’. In *European Conference on Information Retrieval*, 102–109. Springer, 2020.
3. Chiron, Guillaume, Antoine Doucet, Mickaël Coustaty, Muriel Visani, and Jean-Philippe Moreux. ‘Impact of OCR Errors on the Use of Digital Libraries: Towards a Better Access to Information’. In *2017 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, 1–4. IEEE, 2017.
4. Cohen, Robin. *The Cambridge Survey of World Migration*. Cambridge University Press, 1995.
5. Gabrielatos, Costas. ‘Selecting Query Terms to Build a Specialised Corpus from a Restricted-Access Database.’ *ICAME Journal* 31 (2007): 5–44.
6. Kettunen, Kimmo, Eetu Mäkelä, Teemu Ruokolainen, Juha Kuokkala, and Laura Löfberg. ‘Old Content and Modern Tools-Searching Named Entities in a Finnish OCRed Historical Newspaper Collection 1771-1910’. *ArXiv Preprint ArXiv:1611.02839*, 2016.
7. Lund, William B., and Eric K. Ringger. ‘Improving Optical Character Recognition through Efficient Multiple System Alignment’. In *Proceedings of the 9th ACM/IEEE-CS Joint Conference on Digital Libraries*, 231–240, 2009.
8. Nguyen, Thi Tuyet Hai, Adam Jatowt, Nhu-Van Nguyen, Mickael Coustaty, and Antoine Doucet. ‘Neural Machine Translation with BERT for Post-OCR Error Detection and Correction’. In *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020*, 333–336, 2020.
9. Partington, Alan, Alison Duguid, and Charlotte Taylor. *Patterns and Meanings in Discourse: Theory and Practice in Corpus-Assisted Discourse Studies (CADS)*. Vol. 55. John Benjamins Publishing, 2013.
10. Wachsmuth, Henning. *Text Analysis Pipelines: Towards Ad-Hoc Large-Scale Text Mining*. Vol. 9383. Springer, 2015.



# La digitalizzazione del GDLI: un approccio linguistico per la corretta acquisizione del testo?

Eva Sassolini<sup>1</sup>, Marco Biffi<sup>2</sup>, Francesca De Blasi<sup>3</sup>, Elisa Guadagnini<sup>4</sup>, Simonetta Montemagni<sup>5</sup>

1 ILC-CNR – eva.sassolini [at] ilc.cnr.it

2 Accademia della Crusca | Università di Firenze – marco.biffi [at] unifi.it

3 ILC-CNR – francesca.deblasi [at] ilc.cnr.it

4 ILC-CNR – elisa.guadagnini [at] ilc.cnr.it

5 ILC-CNR – simonetta.montemagni [at] ilc.cnr.it

## SINTESI

In questo articolo sono discussi metodi e strategie in via di elaborazione per la correzione (propedeutica alla successiva strutturazione) dei contenuti del Grande dizionario della lingua italiana (GDLI) fondato da Salvatore Battaglia, estratti da un formato digitale non standard. La presenza, in questo formato, di errori distribuiti di vario tipo ha condizionato la scelta dell'approccio all'estrazione e messo in luce tutte le difficoltà dell'operazione. Le sperimentazioni fatte sino a oggi portano a privilegiare una strategia di correzione multilivello, che procede scomponendo in sezioni distinte l'individuazione e la correzione degli errori, in modo da rendere gestibili interventi complessi di correzione semi-automatica, altrimenti improponibili, e consentire un loro affinamento progressivo. Parallelamente alla definizione di regole di riconoscimento di struttura e formato, stiamo analizzando metodi e procedure in grado di migliorare la qualità dell'input e specializzare i moduli di estrazione per i singoli campi della voce a partire dal "lemma". Le finalità del lavoro sono duplici: l'estrazione e strutturazione dei contenuti e la produzione di un formato standard di rappresentazione dei dati. Si tratta di un percorso difficile perché il formato dei dati rende l'uso di strumenti reperibili in letteratura non applicabile. Solamente al termine del lavoro potremo capire se esistono le condizioni per trasformare l'approccio adottato in un protocollo di intervento replicabile.

## PAROLE CHIAVE

dizionari digitali, risorse linguistiche, estrazione dell'informazione, riconoscimento di errori, correzione del testo post OCR

# INTERVENTO

## 1 INTRODUZIONE

Il *Grande dizionario della lingua italiana (GDLI)* è uno strumento lessicografico monumentale, che descrive dettagliatamente la semantica e gli usi delle parole ricorrendo in particolare a una ricchissima esemplificazione tratta dai testi italiani (soprattutto letterari) dalle Origini al XX secolo<sup>1</sup>. I potenziali vantaggi della digitalizzazione e strutturazione del *GDLI* sono chiari agli studiosi: il vocabolario costituisce un ricchissimo corpus diacronico dell'italiano e potrebbe supportare analisi raffinate, che si gioverebbero moltissimo delle possibilità offerte da una consultazione anche non "lineare" (che è l'unica possibile su carta). L'urgenza dell'intervento è quindi nota ma, non avendo potuto migliorare la qualità dell'Optical Character Recognition (OCR) avvalendoci di tecniche di pre- e/o post-elaborazione dell'output, come viene attualmente proposto in letteratura [8], l'input del lavoro di estrazione e strutturazione è oggi costituito da oltre 23.000 pagine di testo in formato digitale non standard, prodotte da procedure di OCR in cui l'originale cartaceo non è stato sottoposto a nessun tipo di collazione, parziale o totale. L'acquisizione risultante, a causa delle caratteristiche particolarmente complesse della stampa<sup>2</sup>, risulta quindi imperfetta tanto a livello puramente testuale (corretto riconoscimento dei caratteri) quanto riguardo alla segmentazione del testo (corretta individuazione delle voci e delle parti della voce). La situazione è ulteriormente complicata dal fatto che il *GDLI* unisce alla forte problematicità formale l'estrema variabilità sostanziale, dovuta al fatto che la lingua del testo accetta – data la massa degli esempi citati e la loro appartenenza a una diacronia lunghissima – una profonda variabilità sia grafico-fonetica sia morfosintattica. In questo contesto, il progetto di digitalizzazione strutturata è iniziato con molte incognite e una lunga strada davanti. L'individuazione e la correzione degli errori OCR di tipo ortografico è un problema noto e studiato nella comunità NLP (Natural Language Processing) sin dagli anni Settanta. Una rassegna dei primi lavori su questo problema è descritta da Kukich [5]. In passato la maggior parte dei tradizionali sistemi d'indagine sugli errori OCR si concentrava sulla costruzione di complesse "matrici di confusione" di caratteri (coppie) per rilevare la presenza di parole inesistenti; più recentemente l'utilizzo di informazioni sul contesto linguistico in cui l'errore appare migliora sperimentalmente la precisione [3]. Un esempio di questo tipo è quello proposto da Bassil *et al.* [1] per l'inglese, utilizzando modelli di n-grammi o da Kissos, in cui si valuta l'impatto relativo di diverse fonti di informazione, combinando caratteristiche di modelli linguistici, informazioni sul processo OCR e informazioni sul contesto del documento [4].

- 
- 1 La pubblicazione del *GDLI* è iniziata nel 1961 e finita nel 2002, presso la casa editrice torinese UTET, per cura di Salvatore Battaglia (per i primi sei volumi) e poi di Giorgio Barberi Squarotti. Ai 21 volumi che compongono l'opera si sono aggiunti due supplementi, pubblicati rispettivamente nel 2004 e nel 2009 (sotto la direzione di Edoardo Sanguineti).
  - 2 Come descritto in [8; 9], pongono grosse difficoltà a una lettura OCR il formato di stampa (i volumi di grande dimensione con testo distribuito su tre colonne; il corpo dei caratteri estremamente piccolo per le parti di definizione ed etimologia, e ancora minore per gli esempi citati, dati l'uno di seguito all'altro separati dal solo "punto fermo"; la presenza di grandi porzioni di testo che si distribuiscono di seguito con pochissimi a capo), la qualità non uniforme della carta, la presenza di caratteri speciali, i tratti particolarmente sottili dello stile corsivo, ecc.

In un ambito più affine si collocano esperienze che hanno affrontato output OCR di lessici storici trovando metodi per rilevare e rappresentare automaticamente il contesto semantico di un testo [10], sul modello già adottato da Zhang e Chang [11] che utilizza una combinazione lineare di modelli linguistici per correggere gli errori.

Anche riguardo l'estrazione della struttura abbiamo valutato altre esperienze che cercano come mappare le informazioni tipografiche ottenute sull'output di testo riconosciuto dall'OCR [7].

Partendo da queste esperienze possiamo vedere il nostro lavoro come un altro esperimento in un complesso tentativo di correzione/strutturazione automatica dei risultati di OCR con errori su dati di rilevanza storico-linguistica. È importante rilevare che, se è vero che molti grandi vocabolari delle lingue occidentali redatti prima del XXI secolo sono oggi disponibili in una versione elettronica, va però sottolineato che essi sono stati acquisiti – o per lo meno integralmente corretti – manualmente, grazie alla ricopiatura/revisione del testo su supporto digitale da parte di operatori umani<sup>3</sup>.

La correzione manuale pone a sua volta una serie di problemi, poiché i correttori fatalmente sbagliano, non correggendo errori già presenti nel testo come anche introducendone di nuovi (che, a quel punto, saranno aleatori e dunque assai difficilmente reperibili). Inoltre, prevedere anche soltanto una rilettura integrale di opere vaste come un grande vocabolario comporta tempi lunghi e un non trascurabile impegno economico.

## 2 L'APPROCCIO LINGUISTICO ALLA CORREZIONE

Con la finalità di arrivare alla digitalizzazione del GDLI con un input di questo tipo, la domanda che ci si pone è: è possibile impiegare intelligenza linguistica per mettere a punto una strategia di correzione efficace, che cerchi di ridurre al minimo la necessità di interventi manuali? Questa sfida riguarda l'estrazione e strutturazione dei contenuti e la loro conversione in un formato standard di rappresentazione, che rispetto alla digitalizzazione rappresenta una base di partenza necessaria. La definizione di un approccio e di norme (buone pratiche) di intervento che possano essere riutilizzati in casi analoghi è un obiettivo di lungo termine di cui ora si possono solo intravedere le basi. L'idea è nata due anni fa e oggi l'approccio adottato può già avvalersi di un insieme di regole e procedure software messe a punto nelle prime sperimentazioni, in grado di produrre una segmentazione progressiva del testo e in cui l'identificazione delle caratteristiche distintive è tradotta in vincoli di corretta attribuzione e impostata in modo incrementale all'interno del processo di parsing del testo digitale [8; 9].

Queste prime sperimentazioni sono state fondamentali per individuare le strategie più adatte ad affrontare un lavoro così complesso, iniziando con una prospettiva "a grana grossa", di macro-interventi a tutto testo, gli errori sono stati suddivisi in "bloccanti" e "non bloccanti"<sup>4</sup>.

---

3 Per citare soltanto alcuni casi notevoli, è stato così per le prime quattro impressioni del *Vocabolario* della Crusca, per il "Littré électronique" e il *Trésor de la langue française informatisé*, per l'*Oxford English Dictionary*.

<http://www.lessicografia.it/>

<http://atilf.atilf.fr/>

4 I primi impediscono la definizione dei confini della voce, inficiando in modo grave il processo di estrazione successivo; i secondi producono un'entrata con errori di vario tipo.

Successivamente l'analisi dei risultati ha mostrato come questa gerarchia non fosse funzionale al lavoro di recupero puntuale di alcune porzioni strutturate di testo e si è manifestata l'esigenza di implementare anche micro-interventi.

A fronte dell'estrema complessità del compito, si è deciso di tentare la strada di una strategia di correzione multilivello, che sfrutti il più possibile criteri di tipo linguistico.

In generale, un testo fortemente strutturato quale è un dizionario presenta "campi testuali" ben distinti, che hanno al loro interno delle omogeneità sfruttabili per la correzione: così, per esempio, il campo "lemma" è caratterizzato da una bassa variabilità morfologica (e per questo motivo si può mettere in relazione con una serie di risorse intra- ed extra-testuali, per cui vedi infra §4).

D'altra parte, l'acquisizione OCR produce errori che sono trasversali a tutti i campi, che ricorrono "a tutto testo" con le medesime caratteristiche e possono essere ricondotti a due tipologie fondamentali (vedi tab. 1). In questa fase la priorità è l'estrazione del lemmario corretto, per questo ci siamo concentrati su due livelli di correzione: la correzione degli errori "a tutto testo", con la finalità di ottenere una ripulitura generale del testo acquisito da OCR; la messa a punto di strategie specifiche di correzione del campo "lemma".

Gli interventi correttivi si collocano in fasi e tempi diversi del processo di estrazione: il formato word del testo per gli errori a tutto testo; i testi dei campi della voce in fase di parsing, valutando la combinazione di diverse caratteristiche (per esempio nel caso del lemma si valuta la presenza di particolari suffissi in rapporto alla categoria grammaticale); i dati strutturati una volta estratti.

In quest'ultimo caso l'approccio alla correzione può essere anche molto complesso e utilizzare più risorse linguistiche come lessici e liste di forme per le ipotesi di correzione.

|                          |  |
|--------------------------|--|
| 1) cattiva segmentazione | 1.1) introduzione di spazi bianchi dove non ci sono  |
|                          | 1.2) univernazione di parole separate  |
| 2) grafie scorrette      | 2.1) omissione di singoli caratteri o sequenze   |
|                          | 2.2) cattivo riconoscimento di singoli caratteri o di sequenze di caratteri                |
|                          | 2.3) cattivo riconoscimento di caratteri che produce l'inserimento di caratteri aggiuntivi |

Tabella 1: tipologie di errori

### 3 STRATEGIE DI CORREZIONE DEGLI ERRORI “A TUTTO TESTO”

Delle due tipologie di errore “a tutto testo”, tralasciamo per il momento il tipo 1), la “cattiva segmentazione”, che richiede strategie di correzione diversificate a seconda del campo della voce (vd. infra §4 per la correzione di questo errore nel campo “lemma”). Concentriamoci quindi sugli errori di tipo 2), vale a dire le grafie scorrette: che si tratti della fattispecie 2.1), 2.2) o 2.3) questi errori possono essere individuati (ed eventualmente corretti automaticamente) a diversi “livelli di correzione”, a seconda dell’area di testo in cui si situa l’intervento correttivo richiesto:

| LIVELLO DI CORREZIONE | AREA DI TESTO INTERESSATA DALL’ERRORE                                  | DESCRIZIONE DELL’ERRORE  | STRATEGIA DI INDIVIDUAZIONE/ CORREZIONE DELL’ERRORE  |
|-----------------------|--|--|--|
| livello 0             | singolo carattere  | presenza di un carattere non alfabetico (o non appartenente all’alfabeto italiano) | <ul style="list-style-type: none"> <li>analisi degli errori della “sezione <i>gold</i>” (sostituzioni 1:1)</li> </ul>                          |
| livello 1             | sequenza di 2 o più caratteri (ogni carattere di per sé è ammissibile) | sequenza di caratteri non ammessa in lingua italiana                               | <ul style="list-style-type: none"> <li>analisi degli errori della “sezione <i>gold</i>”</li> <li>“criterio fonotattico”</li> </ul>             |
| livello 2             | parola (caratteri e sequenze di caratteri di per sé sono ammissibili)  | parola errata  | <ul style="list-style-type: none"> <li>analisi degli errori della “sezione <i>gold</i>”</li> <li>confronto con altri <i>corpora</i></li> </ul> |

Un primo strumento che abbiamo utilizzato per la correzione di errori grafici è, come si vede dalla tabella, la “sezione gold”.

Essa consiste in uno specimen di GDLI, pari a 30 colonne complessive (scelte aleatoriamente), corretto a mano<sup>5</sup> e impiegato per reperire le corrispondenze univoche tra un riconoscimento OCR errato e una correzione, rispetto a uno o più caratteri: la “sezione gold”, quindi, non ha rilevanza statistica e non è servita per l’analisi degli errori, ma soltanto per una ripulitura del dettato rispetto agli errori univoci e ricorrenti. Dato che alcuni di questi errori interessano marche d’uso, grammaticali e semantiche, la loro correzione è funzionale tanto per la successiva strutturazione del testo, quanto per l’implementazione delle possibilità di ricerca trasversale del vocabolario. Sono esempi di questo tipo di errore la forma “aw.” per l’abbreviazione “avv.” nel GDLI a stampa; lo stesso vale per “medie.” per “medic.”, “pass.” per “pass.” (nell’abbreviazione “part. pass.”), ecc. La sezione gold ha permesso anche la correzione di altri errori ricorsivi, quali il nesso “-z?-” per “-zz-”, ecc.

Per l’individuazione degli errori di livello 1) abbiamo elaborato anche il “criterio fonotattico”, che permette al sistema di isolare come erronee le parole che contengono i nessi consonantici (di 2 e 3 caratteri) non ammessi in italiano (cfr. [2; 6]). In questo caso la correzione non è automatica, ma sono segnalate alla correzione manuale singole parole certamente errate: l’applicazione del “criterio fonotattico” individua come sbagliata, per esempio, la forma “impervietà” (poiché il nesso triconsonantico -pcr- non è ammissibile in lingua italiana), che dovrà essere corretta in “impervietà”.

5 La correzione è stata fatta da Cecilia Palatresi e Silvia Dardi, collaboratrici dell’Accademia della Crusca.



Per la gestione degli errori di livello 2), che sono evidentemente i più complicati da individuare, è di grande ausilio il ricorso a risorse esterne, quali le banche dati testuali esistenti e i vocabolari digitalizzati (per un'applicazione specificamente al campo lemma cfr. §4).

## **4 STRATEGIE DI CORREZIONE DEGLI ERRORI NEL CAMPO “LEMMA”**

Accanto alle strategie di correzione già descritte, abbiamo messo a punto una serie di procedure mirate per il campo “lemma”. Va premesso che si è deciso di riprodurre il lemmario del GDLI cartaceo rispetto alla sequenza dei caratteri, facendo astrazione dai diacritici: le entrate presentano l'indicazione dell'apertura/chiusura delle “e” e “o” toniche e l'indicazione della natura sorda/sonora delle affricate e sibilanti. Posto che il cattivo riconoscimento del diacritico è un errore ricorsivo dell'OCR, che il diacritico utilizzato per le consonanti (un puntino sottoscritto) rappresenta un carattere non standard, e che la presenza dei diacritici ostacola il confronto con altri lemmari, si è deciso di non gestire questo ordine di informazioni<sup>6</sup>.

Tornando, quindi, ai livelli di correzione degli errori (tab. 2), posto che gli errori di livello 0 sono gestiti interamente dalle strategie di correzione “a tutto testo”, abbiamo messo a punto delle strategie specifiche per i livelli di errore 1 e 2, per la cui gestione possiamo quindi contare sui risultati del parsing; abbiamo inoltre pensato a dei criteri per individuare e correggere gli errori di cattiva segmentazione (tab. 1).

Per gli errori di livello 1, abbiamo ideato una strategia di correzione che si è rivelata efficace per gli errori di sequenza a inizio di parola. Poiché i lemmi devono presentarsi in sequenza alfabetica, il mancato rispetto dell'ordinamento alfabetico è un sicuro segnale di errore.

Uno strumento di correzione valido consiste nel confronto del lemma con l'intervallo alfabetico segnalato dalle “testatine” (vale a dire le entrate della prima voce della prima colonna e dell'ultima voce della terza colonna, segnalati nell'intestazione di ogni pagina del GDLI): grazie a questo confronto è possibile formulare delle ipotesi di correzione della parte iniziale del lemma – è così possibile per esempio correggere l'errato “bfdicotomia” in “broncotomia”.

Per gli errori, sempre di livello 1, di sequenza a fine di parola, abbiamo pensato di testare un “criterio morfologico”: incrociando le terminazioni ammissibili in italiano per le diverse categorie grammaticali, sarà possibile individuare errori di terminazione e, in alcuni casi, formulare ipotesi di correzione (ad es., per correggere “chetonemla” in “chetonemia”).

Per gli errori di livello 2 e gli errori di tipo 1) (cattiva segmentazione), un ulteriore strumento di correzione del lemmario è la collazione con i lemmari di altri vocabolari italiani: stiamo testando in particolare il confronto con il lemmario del Tommaseo-Bellini (di cui il GDLI costituisce l'esplicito “aggiornamento” e completamento).

“Normalizzando” anche le entrate del T-B (cioè eliminando gli accenti) e considerando le sole entrate monorematiche, i risultati della collazione sono promettenti: il confronto consente la correzione dei lemmi del GDLI in cui sono scorretti singoli caratteri, ma anche dei lemmi in cui sono stati omessi singoli caratteri o che sono segmentati scorrettamente.

---

6 Ma sarà naturalmente accentata la vocale finale degli ossitoni, conformemente all'ortografia italiana.

## 5 CONCLUSIONI

Il progetto complessivo di digitalizzazione e strutturazione è ancora in fase di realizzazione delle risorse e delle procedure, il livello “case study” non può dirsi concluso. L’inquadramento e la definizione della direzione del lavoro hanno avuto una lunga gestazione ma sono ora maturi per essere affrontati in modo sistematico. Questa consapevolezza ha permesso, nel giugno 2020, di aggiungere al gruppo di lavoro il progetto TrAVaSI<sup>7</sup>, che, con l’apporto di diverse competenze specifiche, offrirà al progetto di estrazione un respiro più ampio, sfruttando l’interdisciplinarietà degli ambiti di competenza per specializzare i criteri di estrazione per tutti i fenomeni che lo richiedono. Allo stato attuale l’indagine preliminare sulla qualità del testo OCR di partenza – che ricordiamo non essere descrittiva ma finalizzata unicamente alla realizzazione di procedure d’intervento correttivo quanto più efficaci possibile – può dirsi conclusa. Così anche la fase di lavori preparatori volti ad avviare la correzione assistita del lemario: sulla base delle evidenze emerse nello studio della sezione gold e dei criteri di individuazione degli errori sopra descritti, è stata estratta la lista di lemmi ed è stato stilato un preciso protocollo di correzione automatica e manuale che si avvale di un opportuno strumento software di supporto dedicato.

---

7 Progetto POR FSE 2014 - 2020, Asse A Occupazione - Priorità di investimento A.2 – Obiettivo A.2.1 – Azione A.2.1.7. “ASSEGNI DI RICERCA IN AMBITO CULTURALE” (Bando per progetti congiunti di alta formazione attraverso l’attivazione di assegni di ricerca): il progetto, a cui partecipano ILC e l’Accademia della Crusca, mira a formare giovani ricercatori con professionalità e competenze specifiche.

## BIBLIOGRAFIA

1. Bassil, Youssef, e Mohammad Alwani. «Ocr context-sensitive error correction based on google web 1t 5-gram data set». arXiv Preprint, 2012.
2. Chiari I., Castagna S. «La fonotassi statistica dell'italiano e del tedesco: i nessi consonantici.», 2005.
3. Evershed, John, e Kent Fitch. «Correcting noisy OCR: Context beats confusion». In *Proceedings of the First International Conference on Digital Access to Textual Cultural Heritage*, 45–51, 2014.
4. Kissos, Ido, e Nachum Dershowitz. «OCR error correction using character correction and feature-based word classification». In *2016 12th IAPR Workshop on Document Analysis Systems (DAS)*, 198–203. IEEE, 2016.
5. Kukich, Karen. «Techniques for automatically correcting words in text». *Acm Computing Surveys (CSUR)* 24, n. 4 (1992): 377–439.
6. Muljadic, Zarko. *Fonologia della lingua italiana*. il Mulino, 1972.
7. Reul, Christian, Sebastian Göttel, Uwe Springmann, Christoph Wick, Kay-Michael Würzner, e Frank Puppe. «Automatic Semantic Text Tagging on Historical Lexica by Combining OCR and Typography Classification: A Case Study on Daniel Sander's Wörterbuch der Deutschen Sprache». In *Proceedings of the 3rd International Conference on Digital Access to Textual Cultural Heritage*, 33–38, 2019.
8. Sassolini, Eva, e Marco Biffi. «Strategie e metodi per il recupero di dizionari storici», 2020, 235–39. <https://doi.org/10.6092/unibo/amsacta/6316>.
9. Sassolini, Eva, Anas Fahad Khan, Marco Biffi, Monica Monachini, e Simonetta Montemagni. «Converting and Structuring a Digital Historical Dictionary of Italian: A Case Study». In *Electronic lexicography in the 21st century. Proceedings of the eLex 2019 conference. 1-3 October 2019, Sintra, Portugal*, 603–621. Lexical Computing, 2019.
10. Wick, Michael, M. Ross, e E. Learned-Miller. «Context-sensitive error correction: Using topic models to improve OCR». In *Ninth International Conference on Document Analysis and Recognition (ICDAR 2007)*, 2:1168–1172. IEEE, 2007.
11. Zhang, DongQing, e Shih-Fu Chang. «A Bayesian framework for fusing multiple word knowledge models in videotext recognition». In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2:II–528. Citeseer, 2003.

# Soluzioni human-centred per la lessicografia mobile

Valeria Caruso<sup>1</sup>, Roberta Presta<sup>2</sup>, Roberto Montanari<sup>3</sup>

<sup>1</sup> Università degli Studi di Napoli 'L'Orientale' – vcaruso [at] unior.it

<sup>2</sup> Università degli Studi Suor Orsola Benincasa - roberta.presta [at] centroschianuova.it

<sup>3</sup> Università degli Studi Suor Orsola Benincasa - roberto.montanari [at] unisob.na.it

## SINTESI

Al fine di sviluppare dizionari più completi, innovativi e adatti allo smartphone come mezzo di fruizione, è stato intrapreso un percorso di progettazione “human-centred” e finalizzato alla compilazione di *Idiomatica*, un dizionario delle espressioni idiomatiche italiane per apprendenti stranieri. Il contributo descrive le prime due tappe del percorso di progettazione. Nella prima sono stati ideati i primi prototipi attraverso un processo collaborativo tra lessicografi, studenti cinesi di italiano L2 e specialisti di fattori umani. Due diverse tipologie di strumenti sono state prototipate: una ‘scrollabile’ e, quindi, simile ai dizionari comunemente in uso per smartphone, ed una ‘cliccabile’ che presenta modalità nuove di accesso ai dati.

Nella seconda fase, il tipo cliccabile è stato sottoposto ad una sessione di valutazione formativa con parlanti stranieri per verificare la presenza di possibili criticità nell’uso. Le indicazioni ricavate da questo esperimento saranno utilizzate per perfezionare il prototipo e, successivamente, compararne l’usabilità rispetto ad uno scrollabile.

## PAROLE CHIAVE

risorse linguistiche, software, mutamenti nella lettura e nell’apprendimento, opportunità di accesso ai dati

## INTERVENTO

### 1 DIZIONARI DIGITALI PER SMARTPHONE: COMPLETEZZA DESCRITTIVA E USABILITÀ

Il passaggio dalla lingua scritta su carta alla sua digitalizzazione nei dispositivi elettronici ha contribuito a creare nuove forme di fruizione dei testi e di accesso alle conoscenze. Negli ipertesti letterari, ad esempio, il senso di un romanzo viene ricostruito mediante una esplorazione non lineare delle lessie che lo compongono (Ponzio 2008, 99). In senso generale la ricerca di nuove forme e modelli testuali, ottimizzati per diversi tipi di dispositivi ospiti, può contribuire a rinnovare soprattutto i generi con finalità informative, favorendo la capacità dei destinatari di acquisire conoscenze in tempi rapidi.

L'ubiquità degli smartphone, ad esempio, ha contribuito secondo Jones (2014) all'eversione del cyberspazio (Gibson, 2010) garantendo una massiccia intrusione di dati digitali nella vita quotidiana. A dispetto di ciò, i software per questi dispositivi rappresentano un ambito d'interesse ancora piuttosto marginale nella ricerca scientifica: i dizionari, ad esempio, migrano dalle versioni desktop agli smartphone senza adeguamenti significativi (Gao, 2013), mentre il mercato delle app sembra particolarmente vivace. Winestock e Jeong (2014) riportano che nel 2013 tra le cento app per smartphone più remunerative dell'ecosistema Apple comparivano sessantuno dizionari.

I formati ridotti dei loro schermi e le situazioni di mobilità in cui gli smartphone vengono consultati necessitano di un'accurata progettazione se si vuole garantire ricchezza informativa e modalità agili di accesso ai dati. Su questa direttrice si colloca la ricerca intrapresa per lo sviluppo di *Idiomatica*, un dizionario per smartphone delle espressioni idiomatiche italiane indirizzato a parlanti stranieri. Il target di utenti richiede una descrizione particolarmente accurata di questi fraseologismi, stando alle diverse difficoltà che gli stranieri hanno nel gestire le caratteristiche semantiche e morfo-sintattiche delle locuzioni idiomatiche (Sivanova-Chanturia e Spina, 2019 per una sintesi su tutti questi aspetti). In particolare, poi, per imparare gli idiomi, esperti del settore come Nation (2006) raccomandano esplicitamente l'uso dei dizionari.

La ricerca così intrapresa si sviluppa sul doppio canale della ricerca linguistica (Caruso 2016), per offrire dati completi, e di quella ergonomica (Caruso et al. 2019), per rendere le informazioni accessibili. In questa sede ci soffermeremo a discutere della necessaria mediazione tra queste due componenti nel processo di app design che è stato condotto in sinergia tra lessicografi e specialisti di fattori umani<sup>1</sup>, presentando le due diverse fasi di progettazione di *Idiomatica* realizzate finora. La prima ha consentito di ideare due diverse tipologie di prototipi: uno offre una fruizione di tipo 'scrollabile dei dati, l'altro necessita invece di 'click' su etichette metalinguistiche per accedere alle informazioni ed è stato pertanto chiamato 'cliccabile'. Nella seconda fase di progettazione alcune componenti del prototipo cliccabile sono state testate con apprendenti stranieri di italiano L2 al fine di verificare eventuali criticità nell'accesso ai dati.

Le indicazioni ricavate da questo esperimento saranno utilizzate per perfezionare il prototipo e, successivamente, compararne l'usabilità rispetto ad uno scrollabile.

## 2 SOLUZIONI HUMAN-CENTRED PER IDIOMATICA

L'attenzione verso l'efficientamento dei dizionari in vista dei loro utenti ("Benutzereffizienz") caratterizza la ricerca linguistica già dagli anni '80 (Wiegand, 1987:179). Nell'epoca dei dizionari cartacei, questo obiettivo è stato perseguito con analisi strutturali delle risorse esistenti e con questionari di gradimento per raccogliere i desiderata degli utenti. Ma col diffondersi dei dizionari digitali, già prima degli anni 2000, i saperi linguistici hanno cominciato a coniugarsi sempre di più con quelli statistico-informatici, soprattutto per l'estrazione dei dati dai corpora che hanno rinnovato anche la ricerca lessicologica<sup>2</sup>.

---

1 Professionisti che studiano l'interazione tra l'uomo e gli altri elementi di un sistema socio-tecnico, applicando teorie e metodi per progettare tali sistemi al fine di ottimizzare il benessere dell'uomo e le prestazioni del sistema stesso (ripreso da IEA).

2 Si pensi ai pionieristici studi sulle collocazioni e le espressioni multiparola di Sinclair (1987).

Tuttavia, come osserva Heid (2011), due sono le componenti necessarie all'efficientamento dei dizionari: la selezione dei dati e la loro corretta presentazione al fine di non sopraffare l'utente con troppe informazioni ("swamp the user" secondo Atkins 1996: 257).

È proprio di Heid (2011) il primo studio sui dizionari elettronici condotto con i protocolli di usabilità delle scienze informatiche per testarne principalmente l'efficacia, ovvero la presenza delle informazioni di cui l'utente ha bisogno, e l'efficienza, il tempo necessario per finalizzare la ricerca.

Tuttavia, nella progettazione di *Idiomatica* l'interazione tra protocolli informatici e conoscenze linguistiche è andata ben oltre le fasi di testing di un prodotto già finito e ha riguardando l'intero processo di sviluppo del dizionario.

In linea con i principi dello human-centred design (HCD, Giacomini, 2014), la progettazione è partita dal problema di come rendere accessibili sullo schermo di uno smartphone i numerosi dati dei tredici campi informativi di cui è composta una voce del dizionario<sup>3</sup>. Come evidenziato nello standard internazionale che lo definisce (International Organization for Standardization, 2010), l'obiettivo della progettazione human-centred è quello di realizzare prodotti soddisfacenti dal punto di vista dell'usabilità e della soddisfazione dell'utente, pertanto i bisogni e le esigenze di quest'ultimo devono essere indagati e compresi fin dall'inizio del processo ideativo. Nel caso di *Idiomatica*, i futuri utenti del dizionario (studenti di italiano L2) e gli "esperti in materia" (i lessicografi o "subject matter experts") sono stati parte attiva del processo di sviluppo fin dagli esordi e, successivamente, hanno preso parte alla valutazione delle soluzioni realizzate. Grazie a questa partecipazione si possono infatti realizzare prodotti facili da usare e capaci di rispondere ai bisogni reali delle persone nella risoluzione dei compiti per cui sono stati progettati (Davis 1989, Hornbaek 2017).

### 3 SESSIONE DI CO-DESIGN E PROTOTIPI

Il percorso ideativo di *Idiomatica* è partito con una sessione di "co-design" (o "design partecipato") durante la quale gli specialisti di fattori umani ed i designer hanno potuto confrontarsi con tutte le parti interessate all'utilizzo del dizionario: 4 lessicografi e 14 studenti cinesi di italiano L2 (livello B2 e C1 del QCER). In questo modo è stato possibile comprendere i bisogni e le aspettative degli studenti che consultano un dizionario degli idiomi monolingue su smartphone e, al contempo, realizzare uno strumento in linea con lo stato dell'arte dei saperi veicolati. I materiali raccolti con la sessione di co-design contengono intuizioni significative, ispirazioni e bisogni ben focalizzati provenienti dalla comunità degli stakeholder del progetto (studenti, lessicografi e designer). Sono state così individuate tre macroclassi di caratteristiche a cui il dizionario deve rispondere:

- I. **obiettivi:** gli scopi per i quali l'utente potrebbe desiderare di utilizzare un simile strumento; tra essi, nella prospettiva degli studenti, oltre alle finalità di apprendimento, sono emersi anche la curiosità di indagare gli aspetti culturali della lingua. Per i lessicografi, invece, emerge il desiderio di avere a disposizione uno strumento più completo dal punto di vista dei contenuti rispetto a quelli attualmente a disposizione da poter sfruttare per la didattica.

---

3 Elenco dei campi informativi di *Idiomatica*: Significato, Significato letterale, Significato figurato, Emozioni, Registro stilistico, Atti linguistici, Struttura linguistica, Varianti, Contesti d'uso: Situazioni, Contesti d'uso: Testi, Connettori, Flessione, Diatesi. Si veda anche Caruso et al. 2019.

II. **funzionalità generiche**: funzionalità che l'utente si aspetta di trovare nell'app perché presenti in molte altre app di uso comune; ad esempio, la possibilità di condividere contenuti, esprimerne un rating, usare il dizionario anche offline.

III. **caratteristiche specifiche**: funzionalità e contenuti specifici per il dizionario. Riguardano le ovvie esigenze di comprendere il significato degli idiomi ma non solo: poterne ascoltare la pronuncia, conoscerne le origini, avere più esempi d'uso, comprenderne il registro e i contesti d'uso per poterlo usare correttamente nella vita reale.

L'insieme di queste caratteristiche è diventata la base per progettare una prima serie di prototipi, alcuni dei quali verranno poi selezionati per essere testati con utenti reali come materiali di prova, secondo quanto prescritto dallo HCD. In particolare, la struttura funzionale e informativa dell'app è stata declinata secondo due paradigmi alternativi di interazione con i contenuti. Un primo prototipo, di tipo "scrollabile" (fig. 1a), presenta tutto il contenuto, suddiviso per campi informativi, in una singola schermata e viene letto "a scorrimento", sfruttando il movimento di "scroll" che sposta verso l'alto i contenuti già letti per poter visualizzare quelli a seguire. Questa tipologia di prototipo corrisponde alla struttura dei dizionari per app più comunemente diffusa oggi. Dal punto di vista dell'interazione, essa mette a disposizione dell'utente tutte le informazioni sulla stessa schermata. Un secondo prototipo (fig. 1b), di tipo "cliccabile" (o più precisamente "tappabile", dato che il gesto di selezione sul *touch screen* è chiamato "tap"), presenta il contenuto informativo organizzato per tipi in una forma "indicizzata". L'accesso ai contenuti avviene selezionando una etichetta metalinguistica corrispondente al tipo di dato desiderato. Questa tipologia di prototipo consente di avere una visione sintetica di tutti i tipi di dati disponibili e di indirizzare le ricerche verso quelli più utili allo svolgimento di un compito specifico. Ulteriori indicatori suggeriscono all'utente quali informazioni usare per i compiti di comprensione ("Informazioni per la comprensione") e di produzione ("Informazioni per la produzione"), proponendo una settorializzazione delle informazioni in linea con le proposte della "Function Theory of Lexicography" (Tarp 2008).

Confrontando le due tipologie di prototipi emergono i punti di forza e criticità di ciascuno. In particolare, nel dizionario 'cliccabile' la compattezza microstrutturale è sicuramente maggiore rispetto ad uno scrollabile, ma l'utente può accedere alle informazioni solo se le etichette metalinguistiche usate sono chiare. Inoltre, il dizionario contiene almeno un tipo di descrizione assolutamente nuova per gli utenti. La così detta "struttura linguistica", che illustra aspetti morfo-sintattici di coniugazione dell'idiom evidenziando anche la semantica dei partecipanti all'evento, secondo il modello dei *Frame* di Fillmore (Fillmore et al 2003) e le annotazioni di FrameNet, riproposte tuttavia in maniera semplificata e intuitiva come in Figura 1-destra.

In un percorso di progettazione human-centred la ricerca deve proseguire con dei test con gli utenti finali del prodotto per identificare quale tipologia di prototipo soddisfa maggiormente i loro bisogni. Per un dizionario, questa valutazione può avvenire attraverso prove di usabilità comparativa dei prototipi realizzati nel contesto dell'esecuzione di task lessicografici.

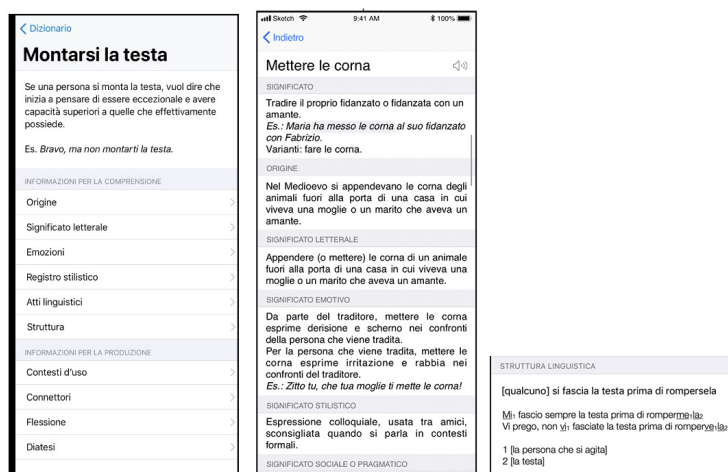


Figura 1: Schermata del prototipo cliccabile (sinistra), schermata del prototipo scrollabile (centro) e rappresentazione della “Struttura linguistica” (destra).

## 4 VALUTAZIONE FORMATIVA DEL PROTOTIPO CLICCABILE

A causa delle criticità nell’accesso ai dati che il prototipo ‘cliccabile’ potrebbe presentare è stato condotto un test di valutazione formativa con utenti per identificare fin dalle prime fasi del ciclo di sviluppo eventuali problematiche d’uso nella soluzione ideata. I test di valutazione formativa, in contrapposizione a quelli di valutazione sommativa, servono per ottenere feedback sul prototipo nelle fasi in cui esso prende forma (Lazar, 2017). Nel nostro caso, il test ha avuto come obiettivo primario quello di osservare, in via preliminare, come utenti reali interagiscono con un dizionario tanto diverso da quelli comunemente in uso e, in seconda istanza, verificare la comprensibilità delle etichette metalinguistiche e l’utilità dei campi informativi per condurre ricerche più rapide.

Date le difficoltà di accesso ai laboratori e le restrizioni imposte dal distanziamento sociale durante i mesi di lockdown, la valutazione formativa si è svolta da remoto sfruttando le potenzialità degli strumenti di videoconferencing online.

10 apprendenti cinesi di livello B1 hanno partecipato alla sperimentazione come volontari. È stato chiesto loro di usare il prototipo dell’app nella sua versione cliccabile per risolvere 5 quesiti lessicografici su 2 idiomi italiani che per essere svolti correttamente avrebbero richiesto l’accesso a diversi campi informativi.

Il prototipo dell’app è stato messo a disposizione dei partecipanti in una sua versione interattiva online realizzata tramite la piattaforma Marvel<sup>4</sup>, mentre i task lessicografici sono stati proposti utilizzando la piattaforma Moodle<sup>5</sup>.

4 <https://marvelapp.com/>

5 <https://moodle.org/>



Ai partecipanti è stato chiesto di condividere lo schermo in modo che i ricercatori potessero vedere sia la schermata su cui essi svolgevano l'esercizio, sia la schermata tramite cui interagivano con il prototipo online.

Gli utenti sono stati invitati ad esprimere ad alta voce i loro pensieri, dubbi e osservazioni durante l'interazione con l'app (Think Aloud Protocol, McDonald, 2012). In questo modo, i ricercatori hanno potuto annotare le osservazioni degli utenti e ricavare prime indicazioni sui principali punti di forza e di debolezza del prototipo proposto. Dopo l'esecuzione di ogni task, le annotazioni dei ricercatori sono state discusse con gli utenti, per approfondire e verificare le loro impressioni sul prototipo utilizzato.

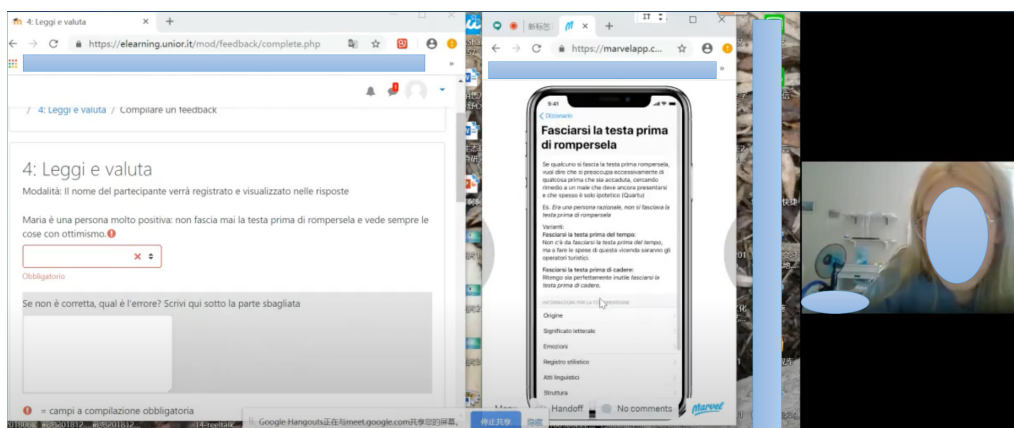


Figura 2: immagine tratta dal test di valutazione formativa. Sulla destra, la schermata con il task lessicografico in esecuzione. Al centro, il prototipo interattivo di Idiomatica utilizzato durante l'esecuzione del task. A sinistra, il partecipante.

Dopo la sperimentazione, le videoregistrazioni sono state analizzate da due diversi ricercatori seguendo il protocollo indicato da Wolfer e colleghi (2018). I video sono stati valutati secondo i seguenti parametri: (i) numero di tipi di informazioni consultate, (ii) tempo speso per ogni tipo di informazione consultata, (iii) correttezza del tipo di informazione consultata (è quella che dovrebbe essere consultata per il task in questione?), eventuali commenti ad alta voce (iv) durante la consultazione e (v) dopo la consultazione.

Ciascuno di questi parametri è stato quindi valutato in funzione dell'esito di ogni prova, ovvero se le risposte corrette venissero date consultando il tipo di dato più adatto, con quali tempistiche e dopo quante altre consultazioni errate. I commenti degli utenti hanno fornito indicazioni sui loro desiderata ma anche critiche e apprezzamenti che verranno usati per migliorare soprattutto il layout grafico dell'app. Analizzando i risultati di queste valutazioni, che saranno presentate nel dettaglio in altra sede, sono emersi i punti di forza e le debolezze del prototipo realizzato. Gli utenti apprezzano il dizionario per la sua ricchezza informativa, l'accessibilità microstrutturale e l'efficacia descrittiva di alcuni campi informativi, come l'illustrazione della corrispondenza tra la morfo-sintassi e la semantica delle locuzioni basata sui *Frame* di Fillmore nel campo "Struttura linguistica". Ugualmente apprezzate sono le indicazioni sui tipi di atti linguistici che le singole locuzioni tipicamente realizzano; mentre le informazioni sul registro non sono state comprese dai partecipanti, soprattutto per l'etichetta metalinguistica usata ("Registro stilistico"), e le tavole flessive non sembrano sufficienti a coniugare correttamente le locuzioni più complesse, come quelle provviste di doppio clitico (ad es. 'fasciarsi la testa prima di rompersela').

Poco percepibile risulta la bipartizione microstrutturale tra le “Informazioni per la comprensione” e le “Informazioni per la produzione”.

Alla luce di questi risultati, prima di procedere con la sperimentazione comparativa di usabilità tra il prototipo scrollabile e quello cliccabile verrà somministrato un questionario online per scegliere etichette metalinguistiche più trasparenti e un layout che evidenzii i suggerimenti sui tipi di dati da consultare per task di comprensione e di produzione.

## BIBLIOGRAFIA

1. Atkins, Beryl T. Sue. «Bilingual dictionaries: Past, present and future». *EURALEX'96 Proceedings. Göteborg: Department of Swedish, Göteborg University*, 1996, 515–546.
2. Caruso, Valeria. «Dizionari elettronici e apprendimento delle espressioni idiomatiche: monitoraggio dei bisogni e prospettive future», 2016.
3. Caruso, Valeria, Balbi Barbara, Johanna Monti, e Presta Roberta. «How Can App Design Improve Lexicographic Outcomes? Examples from an Italian Idiom Dictionary». In *ELEX 2019: SMART LEXICOGRAPHY*, 374–396. Lexical Computing CZ sro, 2019.
4. Davis, Fred D. «Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology». *MIS Quarterly* 13, n. 3 (1989).
5. Fillmore, Charles J., Christopher R. Johnson, e Miriam RL Petruck. «Background to frame-net». *International journal of lexicography* 16, n. 3 (2003): 235–250.
6. Gao, Yongwei. «On the application of dictionaries: from a Chinese perspective». In *Proceedings of the eLex 2013 Conference, Tallinn, Estonia*, 213–224, 2013.
7. Giacomini, Joseph. «What is human centred design?» *The Design Journal* 17, n. 4 (2014): 606–623.
8. Gibson, W. «Google's Earth». *New York Times*, August 2010. <http://nytimes.com/2010/09/01/opinion/01gibson.html>.
9. Heid, U. «Electronic Dictionaries as Tools: Toward an Assessment of Usability». In *E-lexicography: the internet, digital initiatives and lexicography*, a cura di Pedro Antonio Fuertes Oliveira e Henning Bergenholtz. London ; New York: Continuum, 2011. P327 .E19 2011.
10. Hornbæk, Kasper, e Morten Hertzum. «Technology acceptance and user experience: A review of the experiential component in HCI». *ACM Transactions on Computer-Human Interaction (TOCHI)* 24, n. 5 (2017): 1–30.
11. International Ergonomics Association. «Definition of Human Factors and Ergonomics», s.d. <https://iea.cc/what-is-ergonomics/>.
12. International Organization for Standardization. *Ergonomics of human-system interaction*. Genève: ISO, 2010.
13. Jones, Steven E. *The emergence of the digital humanities*. New York: Routledge, 2014. AZ195 .J66 2014.
14. Lazar, Jonathan, Jinjuan Heidi Feng, e Harry Hochheiser. *Research methods in human-computer interaction*. Morgan Kaufmann, 2017.

15. McDonald, Sharon, Helen M. Edwards, e Tingting Zhao. «Exploring think-alouds in usability testing: An international survey». *IEEE Transactions on Professional Communication* 55, n. 1 (2012): 2–19.
16. Nation, P. «Vocabulary: Second Language». In *Encyclopedia of Language & Linguistics*, a cura di E. K. Brown. Amsterdam: Elsevier, 2006. <http://www.sciencedirect.com/science/works/0080448542>.
17. Ponzio, Augusto. «Hypertext and translation». *Translation and The Machine: Technology, Meaning, Praxis*, ed. Steve Berneking and Scott S. Elliott, 2004, 95–102.
18. Sinclair, John M., a c. di. *Looking up: an account of the COBUILD Project in lexical computing and the development of the Collins COBUILD English language dictionary*. 3. repr. London: Collins, 1993.
19. Siyanova-Chanturia, Anna, e Stefania Spina. «Multi-word expressions in second language writing: A large-scale longitudinal learner corpus study». *Language Learning* 70, n. 2 (2020): 420–463.
20. Tarp, Sven. *Lexicography in the borderland between knowledge and non-knowledge: general lexicographical theory with particular focus on learner's lexicography*. *Lexicographica* 134. Tübingen: Max Niemeyer Verlag, 2008. P327.5.S82 T378 2008.
21. Wiegand, H. E. «Zur handlungstheoretischen Grundlegung der Wörterbuchbenutzungsfor- schung». *Lexicographica* 3, n. S (1987).
22. Winestock, Christopher, e Young-kuk Jeong. «An analysis of the smartphone dictionary app market». *Lexicography* 1, n. 1 (2014): 109–119.
23. Wolfer, Sascha, Martina Nied, Idalete Dias, Carolin Müller-Spitzer, e María José Domínguez. «Combining quantitative and qualitative methods in a study on dictionary use», 2018.

# MITAO: a tool for enabling scholars in the Humanities to use Topic Modelling in their studies

Ivan Heibi<sup>1</sup>, Silvio Peroni<sup>2</sup>, Luca Pareschi<sup>3</sup>, Paolo Ferri<sup>4</sup>

<sup>1</sup> Università di Bologna – ivan.heibi2 [at] unibo.it

<sup>2</sup> Università di Bologna – silvio.peroni [at] unibo.it

<sup>3</sup> Università di Roma Tor Vergata – luca.pareschi [at] uniroma2.it

<sup>4</sup> Università di Bologna – p.ferri [at] unibo.it

## ABSTRACT

Automatic text analysis methods, such as Topic Modelling, are gaining much attention in Humanities. However, scholars need to have extensive coding skills to use such methods appropriately. The need for having this technical expertise prevents the broad adoption of these methods in Humanities research. In this paper, to help scholars in the Humanities to use Topic Modelling having no or limited coding skills, we introduce MITAO, a web-based tool that allows the definition of a visual workflow which embeds various automatic text analysis operations and allows one to store and share both the workflow and the results of its execution to other researchers, which enables the reproducibility of the analysis. We present an example of an application of the use of Topic Modelling with MITAO using a collection of English abstracts of the articles published in “Umanistica Digitale”. The results returned by MITAO are shown with dynamic web-based visualizations, which allowed us to have preliminary insights about the evolution of the topics treated over the time in the articles published in “Umanistica Digitale”. All the results, along with the defined workflows, are published and accessible for further studies.

## KEYWORDS

topic modelling, MITAO, tool

# TALK

## 1 INTRODUCTION

The quantitative analysis of textual contents is an approach that has been adopted since the end of the 19th century, when John Gilmer Speed, a former New York World editor, used quantitative methods to compare the content of four New York dailies published between 1881 and 1893 [15].

These methods (and, more generally, the recent computational text mining techniques) are considered an interdisciplinary topic nowadays with relations with computer science, statistics, linguistics, sociology, and other social sciences. Today we have massive digital archives (e.g. Project Gutenberg and Google Books) and new computational tools and sophisticated systems for studying them.

In the past years, such computational tools have attracted scholars from the Humanities and Social Sciences domains. Indeed, an increasing number of scholars in the Humanities have started to take advantage of the potentials of such computational tools for text processing and analysis [10]. In particular, within the Digital Humanities, such tools are crucial in addressing research questions [8], as also demonstrated by the recently-born Computational Humanities Research community.<sup>1</sup>

Topic Modelling is one of the textual processing methodologies mentioned above. Topic Modelling aims at determining what events or concepts a document or a collection of documents is discussing [16], something which may have several applications in the Humanities [2] and the Digital Humanities [11].

Technically speaking, Topic Modelling is an automatic technique based on Bayesian statistics, and in particular, on an algorithm called Latent Dirichlet Allocation [9]. Topic Modeling analyses texts and creates ‘topics’, which are bags of words that often co-occur together in the original texts [12]. The underlying idea is that the algorithm is able to elicit a latent structure of topics that constitute the texts in a corpus [9]. As a result, all the words in the textual sources are coded to a topic, and all the original texts are constituted by several topics according to different percentages. Topics can be ontologically different [4], as they can be interpreted as themes, or discourse, or frames [5]. The work done by DiMaggio, Nag, and Blei [3] identifies four relevant characteristics of the Topic modelling analysis that helps researchers cover the limitations arisen from their previous approaches (mainly based on a manual approach):

- a) the generated data are available for other researchers willing to test the interpretations and conclusions inferred;
- b) the automatic analysis permits dealing with large amounts of documents;
- c) the process is an inductive one, such that it allows researchers to discover the structure of the corpus before imposing their priors on the analysis, and also enables the usage of the same corpus to pursue other different research questions;
- d) it deals with polysemy, such that it is able to recognize that the meaning of a term depends on its context (therefore the same term might have different meanings depending on its context).

---

1 <https://www.computational-humanities-research.org/>

While useful as an approach per se, one has to master at least a programming language and a tool for a visual representation of data to use Topic Modelling correctly.

However, scholars in the Humanities are not usually trained in these skills – and, thus, the adoption of Topic Modelling in their research is perceived as an insuperable barrier. Interfaces and visual tools are needed in order to mediate the use of Topic Modelling appropriately, so as to hide its technical implementation to non-experts.

In this paper, we introduce MITAO, a tool which makes Topic Modelling techniques reusable by scholars with no or limited coding skills by means of a visual interface which enables them to create a workflow for processing textual content [6]. MITAO uses the Latent Dirichlet Allocation (LDA) Topic Modelling, one of the earliest and well-known methods, originally developed by Blei, Ng, and Jordan [1]. Any workflow produced using MITAO can be stored and shared among scholars for reproducibility purposes.

In Section 2, we provide an introduction to MITAO, we give an overview of its architecture, we discuss its main features available, and we introduce how to perform a Topic Modelling analysis and to get the tabular and graphical results using it. In Section 3, we show an example of an application of MITAO using a collection of the abstracts of the articles published in “Umanistica Digitale”<sup>2</sup>. In Section 4, we conclude this article sketching out some future works.

## 2 MITAO

MITAO (Mashup Interface for Text Analysis Operations) is an open-source, user-friendly, modular, and flexible software written in Python and Javascript for performing several kinds of text analysis. MITAO can be run locally on a machine by using any modern Web browser. MITAO is licensed under the ISC License, and source code and documentation are available on GitHub.<sup>3</sup>

We developed MITAO to help scholars with no or limited skills in coding to overcome two of the main issues they have to deal with: (a) using computational text analysis techniques for their own research using a particular programming language, and (b) describing and discussing the technical aspects of their analysis. To overcome these limitations, the current version of MITAO (downloadable from its GitHub repository) can:

- convert documents (from PDF to TXT);
- clean textual content (e.g. stopword removal or removal of parts of a text through the use of regular expressions);
- perform Topic Modelling;
- provide a quantitative measure of the results (through perplexity score and topic coherence [13]);
- visualize the topic model created with dynamic web-based visualizations;
- save the data and visualizations produced

In addition, users can save the workflow defined in MITAO and afterwards share it with other colleagues or publish it, so as to foster reproducibility of the results of research.

---

2 <https://umanisticadigitale.unibo.it/>

3 <https://github.com/catarsi/mitao>

The GUI of MITAO is simple and user friendly. The defined workflow is represented as a graph network composed of two types of nodes: “tool” and “data”.

A “tool” node implements operations one can run on data, such as a filter (e.g. filtering a document from text values that follow a specific regular expression), a text analysis (e.g. corpus tokenizer), or a terminal operator (e.g. charts and web-based visualizations). Instead, a “data” node represents a single textual file or a collection of textual files (in plain text, PDF, or textual tabular format).

In this paper, we focus on the features of MITAO strictly related to the Topic Modelling analysis.

## 2.1 Building a topic modelling

A standard Topic Modelling workflow can be defined according to three main steps: (a) tokenization, (b) building the corpus and dictionary, and (c) building the topic model. In MITAO, these three steps can be defined as shown in Figure 1. The workflow starts with the two “data” nodes, which represents a collection of documents (i.e. docs) and a list of stopwords (i.e. stopwords). Both such nodes are specified as input to the “tool” node tokenizer which converts the texts into a list of terms with no stopwords in it. Then, the workflow creates the corpus and dictionary (i.e. “tool” node corpus and dict builder) to be further used as input in the creation of the topic model (i.e. “tool” node lda topic modelling).

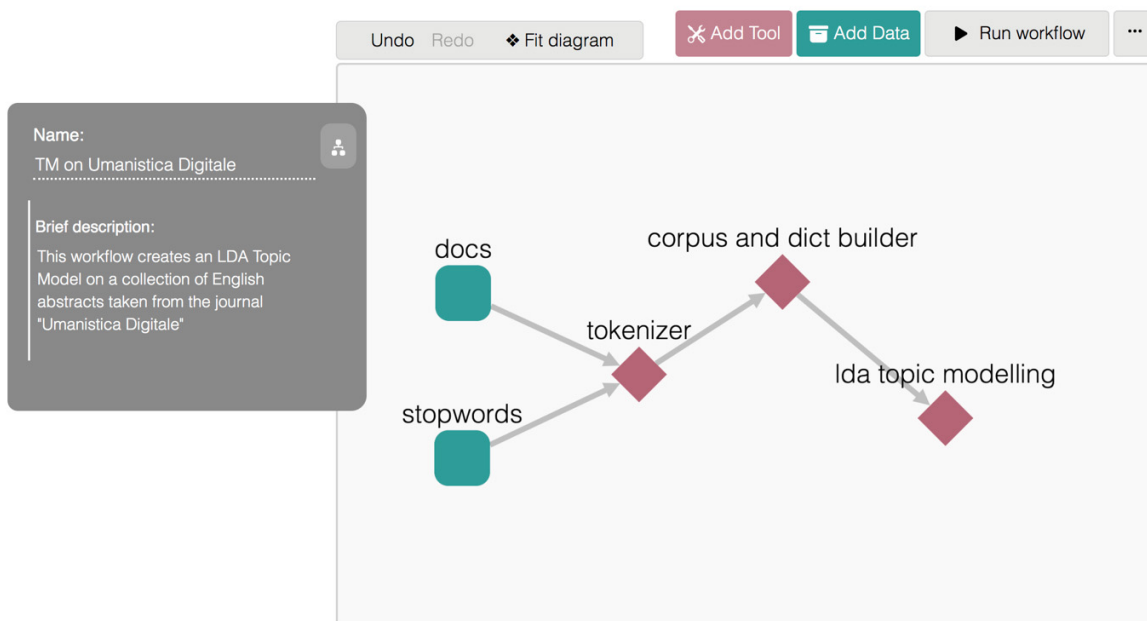


Figure 1: a Topic Modelling workflow defined in MITAO. Starting from a collection of documents (i.e. docs) and a list of stopwords (i.e. stopwords), the workflow goes through three different steps: (a) tokenization (i.e. tokenizer), (b) building the corpus and dictionary (i.e. corpus and dict builder), and finally (c) the creation of the topic model (i.e. lda topic modelling).

## 2.2 Topic modelling results

Using MITAO we can generate two important tabular datasets: (a) `termsXtopics`, i.e. the 30 terms that better characterise each topic, and (b) `docsXtopics`, i.e. a list of all the documents of the corpus with their corresponding representativeness for each topic in the topic model we built. Along with the tabular datasets, we can use MITAO to generate two other web-based dynamic visualisations: `LDavis` and `MTMvis`. `LDavis` [14] provides a graphical overview of the topics of our topic model.

Such topics are shown in a two-dimensional plane whose centres are determined by computing the distance between topics. `MTMvis` has been built for MITAO and shows the topic representativeness in the document corpus based on a metadata attribute of such documents.

These visualisations enable us to visually investigate the document corpus. In the next section, we present a real application of MITAO and demonstrate the potentials of these visualisations.

## 3 AN APPLICATION

In this section, we show an example of a Topic Modelling analysis performed using MITAO, accompanied by the datasets and visualizations we obtained by running the workflow. The topic modelling is done on a collection of 51 abstracts in English of the articles published in “Umanistica Digitale”. The MITAO workflow we developed, the datasets and the visualizations we obtained are available in [7]. We choose to create a topic model with five different topics. The number of topics should be given as input of the `lda` topic modelling step, along with the dictionary and the corpus expressed as a bag of vectors, created in the `corpus` and `dict builder` step. The process of choosing this number of topics is out-of-scope of the present paper and have been calculated with the help of a tool MITAO makes available to compute the coherence score of several topic models.

In Figure 2, we show the `LDavis` generated as a result of Topic Modelling activity. The chart shows five circles (topics) and, when selecting one of the topics, it shows the 30 most recurrent terms of such topic on the right side of the visualization. Figure 3 shows `MTMvis`, which plots the distribution of the five topics in time, considering the year of publication of the articles in our corpus. The combination of both these visualizations can let us come up with some initial insights. From the `MTMvis`, we see that topic-1 (in blue colour) appeared only in 2019. Moving the cursor over such a slice, `MTMvis` shows that 15.63% of the documents published in 2019 had this as the dominant topic. If we check topic-1 on `LDavis`, we see that it has many terms related to the second World War and the Holocaust, such as: “jew”, “holocaust”, “social”, “testimoni”, “oorlogsbronnen”, “state” etc.

Through `LDavis`, we can clearly see an intersection between topic-4 and topic-2. While looking at the most recurrent words of these two topics, we notice that both mention words close to Italian literature. More precisely we can see that topic-4 has a strong relation with the Italian poetry, especially with Dante Alighieri, with words such as “alighieri”, “literatur”, “librari”, “philolog”, “poetri”, “poet”, etc. `MTMvis` shows that topic-4 (in brown color) had a strong relevance in 2017, and gained less relevance in the following years. Another emerging fact is that topic-5 (in orange color) had constant relevancy throughout the years (average value of 25%). From `LDavis`, we can observe that topic-5 contained words such as “visual”, “document”, “model”, “corpora”, “ontolog”, etc. From these bags of words, we might infer that topic-5 is related to works dealing with data analysis involving in some cases the definition of a model or dealing with corpora. The constant relevancy of topic-5 along the years make us believe that these subjects are a regular and important part of “Umanistica Digitale” publications.



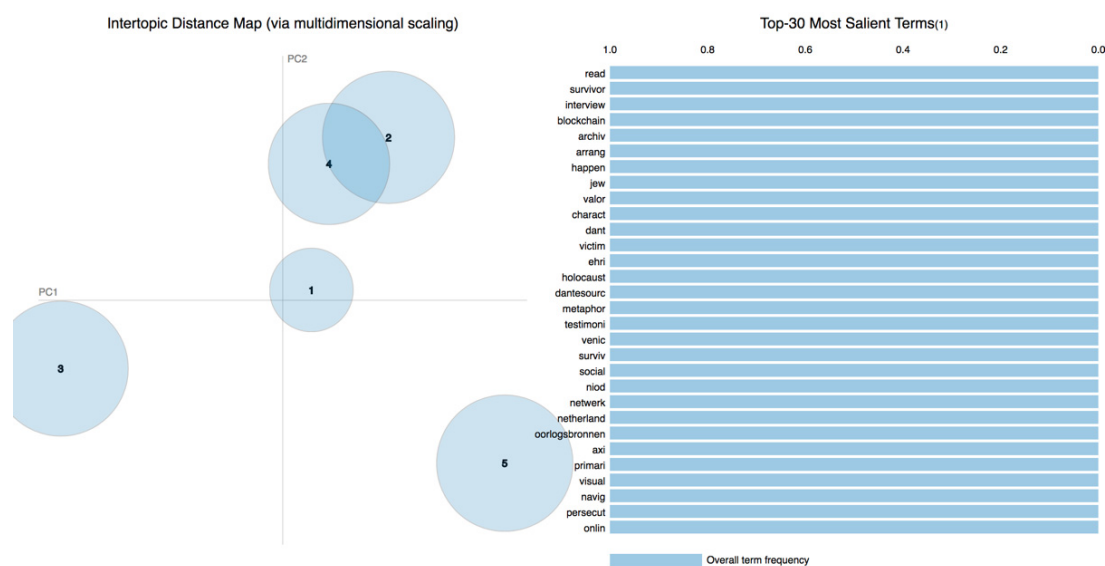


Figure 2: the LDAvis visualization of the topic model built over a collection of abstracts of articles published in “Umanistica Digitale”. This view has been generated using MITAO and it’s available in [7].

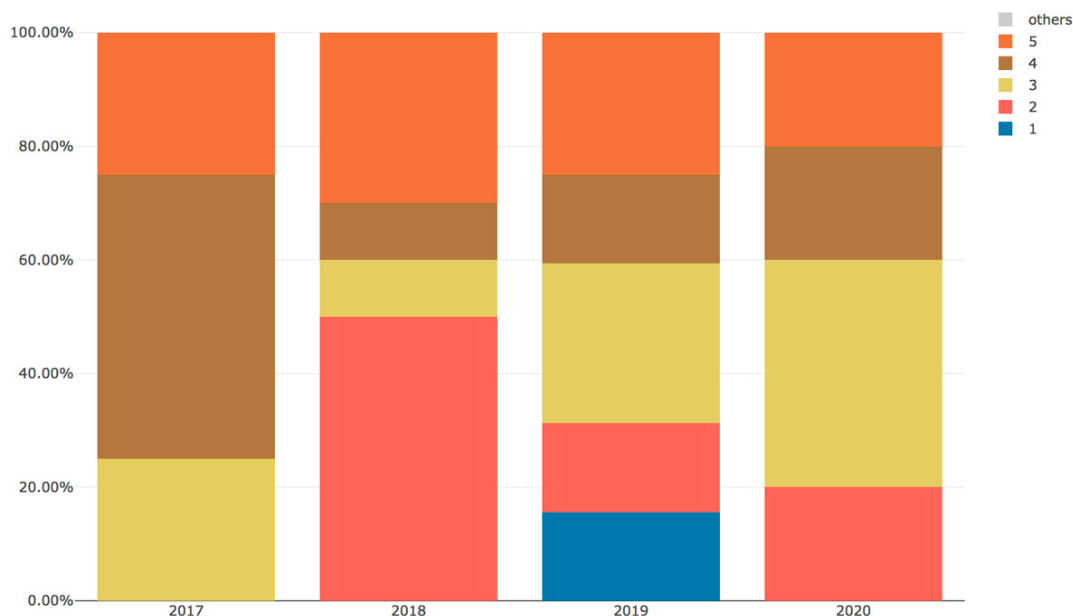


Figure 3: the MTMvis visualization of the topic model built over a collection of abstracts of articles published in “Umanistica Digitale”. We plotted the distribution of the topics according to the year of publication of the articles in the corpus.

## 4 CONCLUSIONS

In this article, we have claimed the need of supporting scholars in the Humanities having no or limited skills in coding in the use of computational tools for automated textual analysis. In particular, we have presented MITAO, a web-based tool that allows the definition of a visual workflow which can embed several textual analysis activities and methods such as Topic Modelling.

MITAO enables the integration of text analysis operations without needing a strong knowledge about their technical implementation and that it builds a visually comprehensive workflow which might be shared with other colleagues and foster the reproducibility of the results obtained.

While there is still room for further improvements, the first release of MITAO was already presented and tested during a symposium organised within the EURAM 2019 Conference – Exploring the future of management, which took place in Lisbon (Portugal) in June 2019<sup>4</sup>.

In the future, we plan to organise user testing sessions, dedicated workshops, and tutorials on the use of MITAO. This will help us promote our tool in different disciplines, to improve its usability, and to add new relevant features to address particular studies.

## REFERENCES

1. Blei, David M., Andrew Y. Ng, and Michael I. Jordan. ‘Latent Dirichlet Allocation’. *Journal of Machine Learning Research* 3, no. Jan (2003): 993–1022.
2. Ciotti, Fabio. ‘Modelli e metodi computazionali per la critica letteraria: lo stato dell’arte’, 2017.
3. DiMaggio, Paul, Manish Nag, and David Blei. ‘Exploiting Affinities between Topic Modeling and the Sociological Perspective on Culture: Application to Newspaper Coverage of U.S. Government Arts Funding’. *Poetics* 41, no. 6 (2013): 570–606. <https://doi.org/10.1016/j.poetic.2013.08.004>.
4. Ferri, Paolo, Maria Lusiani, and Luca Pareschi. ‘Accounting for *Accounting History* : A Topic Modeling Approach (1996–2015)’. *Accounting History* 23, no. 1–2 (2018): 173–205. <https://doi.org/10.1177/1032373217740707>.
5. Gamson, William A. *Talking Politics*. Cambridge [England] ; New York, NY, USA: Cambridge University Press, 1992. HN90.P8 G36 1992.
6. Heibi, Ivan, Silvio Peroni, Paolo Ferri, and Luca Pareschi. *Catarsi/Mitao: MITAO First Release* (version v1.1-beta). Zenodo, 2019. <https://doi.org/10.5281/ZENODO.3258327>.
7. Heibi, Ivan, Peroni, Silvio, Pareschi, Luca, and Ferri, Paolo. ‘MITAO: A Tool for Enabling Scholars in the Humanities to Use Topic Modelling in Their Studies (Data and Results of MITAO)’, 1 October 2020. <https://doi.org/10.5281/ZENODO.4061760>.
8. Jänicke, S., G. Franzini, M. F. Cheema, and G. Scheuermann. ‘Visual Text Analysis in Digital Humanities: Visual Text Analysis in Digital Humanities’. *Computer Graphics Forum* 36, no. 6 (2017): 226–50. <https://doi.org/10.1111/cgf.12873>.
9. Jelodar, Hamed, Yongli Wang, Chi Yuan, Xia Feng, Xiahui Jiang, Yanchao Li, and Liang Zhao. ‘Latent Dirichlet Allocation (LDA) and Topic Modeling: Models, Applications, a Survey’. *Multimedia Tools and Applications* 78, no. 11 (2019): 15169–211. <https://doi.org/10.1007/s11042-018-6894-4>.

---

4 <http://pastconferences.euram.academy/programme2019/symposia.html>

10. Jockers, Matthew L., and Ted Underwood. 'Text-Mining the Humanities'. *A New Companion to Digital Humanities*, 2015, 291–306.
11. Meeks, Elijah, and Scott B. Weingart. 'The Digital Humanities Contribution to Topic Modeling'. *Journal of Digital Humanities* 2, no. 1 (2012): 1–6.
12. Mohr, John W., and Petko Bogdanov. 'Introduction—Topic Models: What They Are and Why They Matter'. *Poetics* 41, no. 6 (2013): 545–69. <https://doi.org/10.1016/j.poetic.2013.10.001>.
13. Newman, David, Jey Han Lau, Karl Grieser, and Timothy Baldwin. 'Automatic Evaluation of Topic Coherence'. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, 100–108, 2010.
14. Sievert, Carson, and Kenneth E Shirley. 'LDAvis: A Method for Visualizing and Interpreting Topics', 2014. <https://doi.org/10.13140/2.1.1394.3043>.
15. Sumpter, Randall S. 'News about News: John G. Speed and the First Newspaper Content Analysis'. *Journalism History* 27, no. 2 (2001): 64–72.
16. Vayansky, Ike, and Sathish AP Kumar. 'A Review of Topic Modeling Methods'. *Information Systems* 94 (2020): 101582.

# On the Reusability of Terminological Data

Giorgio Maria Di Nunzio<sup>1</sup>, Federica Vezzani<sup>2</sup>

<sup>1</sup> Università di Padova – giorgiomaria.dinunzio [at] unipd.it

<sup>2</sup> Università di Padova – federica.vezzani [at] unipd.it

## ABSTRACT

Reusability of data is one of the most important practices in science, and investments in this (underestimated) operation may have positive long-term consequences in research (Pasquetto et al. 2017). In this paper, we discuss the benefits of this approach in terminology management by presenting a methodology for the preservation of multilingual terminological records and the practice of standardization as a fundamental step towards reusability. We present a case study to show the effectiveness of this methodology on an obsolete Website containing a multilingual medical glossary, and we share the source code as well as the standardized dataset.

## KEYWORDS

terminology re-usability, terminology management, medical termbases, open science, standardization

## TALK

### 1 INTRODUCTION

The design and implementation of high-quality terminological resources is expensive and time-consuming (Warburton, 2015); nevertheless, this costly process is mandatory if the resource has to be reusable, integrated, and useful for a variety of tasks [6]. Reusability is a keystone in the management of scientific data and, more generally, fits into a wider context of “data curation”, as a good practice for organizing, preserving and enhancing research data [5]. In this regard, the FAIR principles supported by the European Open Science Cloud (EOSC)<sup>1</sup> stress the needs for reusability of data and metadata by means of well-described and standardized approaches [10].

In this context, reusability in terminology implies that terminological databases should comply with a standard format, such as the TermBase eXchange (TBX) format, promoted by the ISO standard 30042: 2019, so that they can be interoperable and reusable [2; 3].

In this paper, we focus on a methodology for preserving and enhancing terminological research data in order to open the possibility of developing cross-lingual and multi-lingual applications.

---

1 <https://www.eosc-portal.eu>

We present a case study on 1) the recovery of medical terminological data stored in obsolete HTML pages of a European project; 2) the transformation of these data into the most recent ISO TBX standard; 3) the design and implementation of a new interactive Website to show the potential of the reusability of these data.

## 2 TERMINOLOGY REUSABILITY

In this section, we describe the methodology for transforming a semi-structured data stored in an obsolete Website into a reusable and standardized multilingual terminological termbase. Our case study is the “Multilingual Lemma Collection” website, which was part of the Multilingual Glossary of technical and popular medical terms.<sup>2</sup> The website, updated the last time on the 3rd of June 2000, contains about 1,800 medical concepts translated in eight<sup>3</sup> of the twenty-four official European languages: English, Dutch, French, German, Italian, Spanish, Portuguese and Danish, both in their technical and popular variants.

### SCRAPING DATA

In Figure 1, we show a portion of the first page of the English monolingual glossary and the corresponding HTML code. The terminological records of the original Web site are stored as items of an unnumbered list, where each item contains the technical term and the popular term separated by a comma, both identified by the value in the “name” tag (for example, in Figure 1, 0609 is the identifier for the term *epithelioma*). We use a pipelined R “tidyverse” approach<sup>4</sup> to analyze and extract the data from each webpage, to store and reorganize them into standard XML. The output of this process (source code available online<sup>5</sup>) produces a set of 14,471 terminological entries.

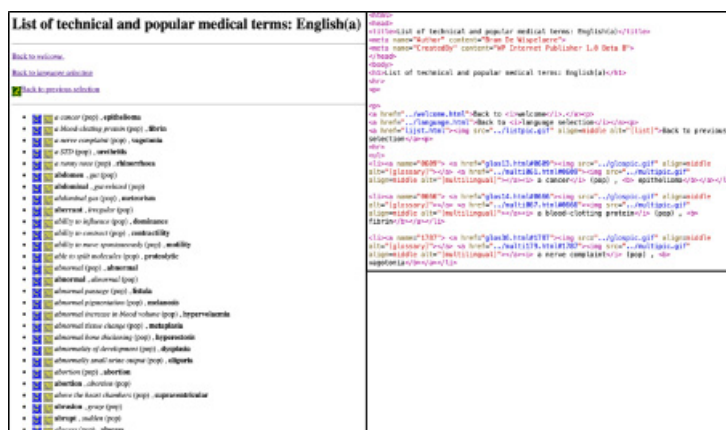


Figure 1: partial view of the Multilingual Glossary Website (left) with corresponding HTML code (right)

2 <https://users.ugent.be/~rvdstich/eugloss/welcome.html>

3 The project also declares the presence of the Greek language as a working language, but the terms are not available.

4 <https://www.tidyverse.org>

5 <https://github.com/gmdn>

## STANDARDIZATION AND VISUALIZATION

In order to provide a structured and standardized terminological version of the data recovered from the original HTML, we chose the structure of the terminological record of the TriMED multilingual medical termbase [7] which refers to the latest ISO standard 30042 (TBX) format [1] and therefore can be reused in any terminology management systems.

In Figure 2, we show an excerpt of the terminological record generated for the term *hypervolemia* recovered from the Multilingual Glossary. Most pieces of information are missing since the original glossary was a simple one; nevertheless, the missing items can now be filled-in semi-automatically by means of medical ontologies, such as the Unified Medical Language System (UMLS),<sup>6</sup> or manually by translators during their work.

The standardization of the terminological data allows not only the interoperability among different systems, but also an easier reuse in other applications that need these data. As an example, we designed and implemented an interactive Web page<sup>7</sup> that uses the standardized data, based on the following requirements:

- Group all the main features of the original Web page into a single interactive page
- Show the standardized XML version of the data
- Allow the download of the terminological record

Instead of a static browsing, in this Web application the user can search a term in any of the (source) languages using a keyword completion live search text area. If requested, the user can also choose one of the other (target) languages and see the terminological record as a bilingual glossary with both the technical and popular variants.

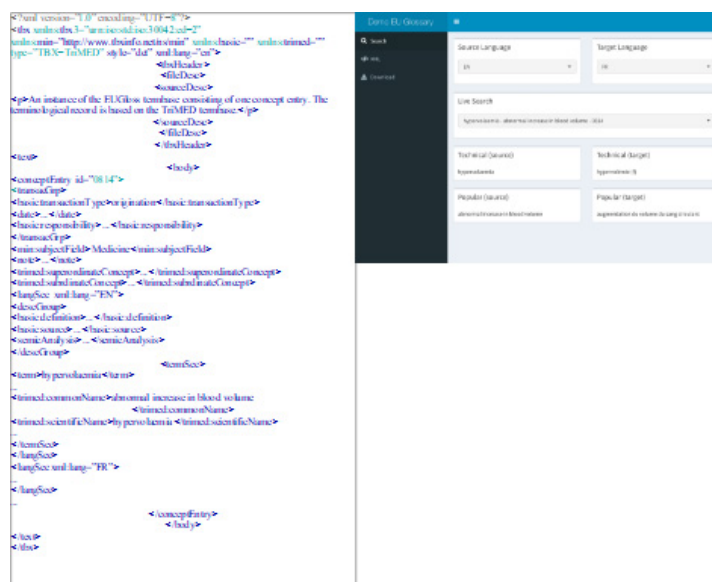


Figure 2: interactive web page and XML standard record

6 <https://www.nlm.nih.gov/research/umls/index.html>

7 <http://shiny.dei.unipd.it/glossary/>

### 3 CONCLUSIONS

The design and implementation of multilingual terminological resources open the possibility for developing cross-lingual and multi-lingual applications. In this paper, we addressed the problem of “terminology reusability” as a specific instance of the problem of the implementation of multilingual terminological resources.

We presented a case study and a pipeline for data retrieval, manipulation, and disposal as well as the process of standardization of these records according to the ISO standard in force (TBX) for terminology management based on the TriMED multilingual medical termbase. The standardized machine-readable format allows the reusability of terminological data for multilingual applications as well as the re-design of an interactive Web page that significantly improves user experience for lay-people while looking for medical information.

### REFERENCES

1. ISO-30042. *Management of terminology resources – TermBase eXchange (TBX)*. 2nd ed. Geneva,CH: Standard, International Organization for Standardization, 2019.
2. Melby, Alan K. ‘TBX: A Terminology Exchange Format for the Translation and Localization Industry’. In *Handbook of Terminology*, edited by Hendrik J. Kockaert and Frieda Steurs, Vol. 1. John Benjamins Publishing Company, 2015.
3. Melby, Alan K., Claus-Dirk Schmitz, and Sue Ellen Wright. *Handbook of Terminology Management: Application-Oriented Terminology Management*. Edited by Sue Ellen Wright and Gerhard Budin. Vol. 2. John Benjamins Publishing, 2001.
4. Paschetto, Irene V., Bernadette M. Randles, and Christine L. Borgman. ‘On the Reuse of Scientific Data’. *Data Science Journal* 16 (2017): 8. <http://doi.org/10.5334/dsj-2017-008>.
5. Poole, Alex H. ‘Now Is the Future Now? The Urgency of Digital Curation in the Digital Humanities.’ *DHQ: Digital Humanities Quarterly* 7, no. 2 (2013). <http://www.digitalhumanities.org/dhq/vol/7/2/000163/000163.html>.
6. Schmitz, K. D. ‘Using International Standards for Terminology Exchange’. *Terminologija* 19 (2012): 33–38. [https://ilias.th-koeln.de/goto.php?target=file\\_929759\\_download&client\\_id=ILIAS\\_FH\\_Koeln](https://ilias.th-koeln.de/goto.php?target=file_929759_download&client_id=ILIAS_FH_Koeln).
7. Vezzani, Federica, and Giorgio Maria Di Nunzio. ‘Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020)’. Marseille, France, 2020.
8. Wachsmuth, Henning. *Text Analysis Pipelines: Towards Ad-Hoc Large-Scale Text Mining*. Vol. 9383. Springer, 2015. <https://doi.org/10.1007/978-3-319-25741-9>.
9. Warburton, Kara. *Managing Terminology in Commercial Environments*. Edited by Sue Ellen Wright and Gerhard Budin. Vol. 1. John Benjamins Publishing, 2015.
10. Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, and Gabrielle Appleton. ‘Scientific Data, 3 (1) (2016)’. *Scientific Data* 3, no. 1 (2016). <https://doi.org/10.1038/sdata.2016.18>.

# **Sessione VI**

## **Aldo Manuzio**





# Fonti archivistiche medievali nel digitale. La sfida di trattare e visualizzare dati semi-strutturati

Roberto Rosselli Del Turco<sup>1</sup>, Enrica Salvatori<sup>2</sup>, Andrea Nanetti<sup>3</sup>, Marco Giacchetto<sup>4</sup>,  
Vera Isabell Schwarz-Ricci<sup>5</sup>, Antonella Ambrosio<sup>6</sup>

<sup>1</sup> Università di Torino – roberto.rossellidelturco [at] unito.it

<sup>2</sup> Università di Pisa – enrica.salvatori [at] unipi.it

<sup>3</sup> Nanyang Technological University – andrea.nanetti [at] ntu.edu.sg

<sup>4</sup> Università di Siena – giacchettomarco [at] hotmail.it

<sup>5</sup> Università di Napoli Federico II – vera.schwarz [at] monasterium-italia.net

<sup>6</sup> Università di Napoli Federico II – antonella.ambrosio [at] unina.it

## SINTESI

Nei giorni 22-23 giugno 2020 si è tenuto il webinar “Fonti archivistiche medievali nel digitale. La sfida di trattare e visualizzare dati semi-strutturati”, organizzato dal Laboratorio di Cultura Digitale<sup>1</sup> e dal responsabile del progetto Engineering Historical Memory<sup>2</sup>. Questo intervento vuole offrire al pubblico di AIUCD 2021 un breve resoconto dei risultati del workshop tenuto su piattaforma virtuale, con particolare attenzione per quanto riguarda future iniziative di collaborazione e sperimentazione nel campo dell’elaborazione ed estrazione di informazioni da edizioni semi-strutturate di documenti archivistici. Funzionalità avanzate di questo tipo sono infatti essenziali per ampliare la fruizione delle edizioni digitali di fonti archivistiche e di altri documenti scritti, rendendole strumenti di ricerca più flessibili ed aperti alla lettura da parte di algoritmi. Il problema centrale riguarda il rapporto tra la codifica e la creazione di interfacce quanto più possibile semplici ed immediate, in cui l’utente (umano o artificiale) possa accedere intuitivamente.

## PAROLE CHIAVE

semplificazione digitale, filologia digitale, storia digitale, edizioni digitali, codifica XML/TEI, codifica XML/CEI, information retrieval, digital public humanities, open culture, digital diplomatics

---

1 [www.labcd.unipi.it](http://www.labcd.unipi.it)

2 <https://engineeringhistoricalmemory.com/>

# INTERVENTO

## 1 INTRODUZIONE

Uno dei vantaggi delle edizioni scientifiche digitali rispetto alle tradizionali versioni a stampa è il fatto che si tratta di strumenti dinamici, all'interno dei quali i testi sono rappresentati in un formato che permette la loro elaborazione per fini che vanno al di là della semplice lettura. Grazie all'uso di linguaggi formali come l'XML, in particolare nella versione TEI P5, i dati testuali di un'edizione possono essere interrogati per ricavare rapidamente molti tipi di informazione.

Questa caratteristica risulta quindi particolarmente interessante non solo sul piano strettamente filologico (ad esempio per definire e visualizzare una tipologia di varianti testuali o collegare specifiche lezioni alle scansioni dei manoscritti), ma anche per l'esegesi di fonti primarie, fondamentali per gli studi storici. La prassi più frequente in quest'ambito, tuttavia, consiste nell'uso di database per archiviare e analizzare i dati, e questo porta in alcuni casi a uno sdoppiamento di funzioni tra le edizioni critiche digitali, utilizzate soprattutto per la visualizzazione di testi e immagini, e il ri-uso dei dati di tali edizioni in ambienti in cui si possono effettuare operazioni di ricerca e data mining.

La domanda che ha animato l'incontro è stata la seguente: è possibile individuare metodi alternativi che permettano di elaborare dati testuali in formato semi-strutturato all'interno di una edizione critica digitale, e fornire almeno parte del processing utile agli storici? Come è possibile creare edizioni digitali di fonti storiche che non rispondano a un'unica categoria di utenti, di solito coincidente con quella da cui origina l'edizione stessa? È possibile, infine, applicare la stessa metodologia a documenti di altro tipo, ad esempio opere letterarie? Quali sono i casi d'uso possibili?

## 2 CASI DI STUDIO

Nel corso dell'incontro sono stati esaminati, nell'ottica descritta, alcuni progetti operativi.

**Il Codice Pelavicino Edizione Digitale dentro e fuori EVT.** Progetto ormai giunto al termine di edizione digitale del liber iurium del vescovo di Luni (fine XIII secolo). Nel corso del tempo il progetto si è arricchito nella codifica e negli strumenti messi a disposizione degli utenti per la ricerca, ma è arrivato a una impasse sulla gestione di alcuni dati storicamente rilevanti e semistrutturati contenuti nei documenti, come ad esempio le liste dei censi. Come trattarli? Come consentire l'analisi allo studioso? Internamente all'edizione o esternamente, tramite la semplice messa a disposizione dei file XML? Parallelamente il Codice Pelavicino ha riscosso un insperato successo presso un pubblico di studiosi locali e appassionati, che contribuiscono giornalmente alla individuazione di toponimi e alla correzione di errori. Questo apre altri problemi: in che modo preservare o migliorare la navigabilità dell'edizione anche per un pubblico di non addetti ai lavori? L'implementazione di funzioni rischia di complicare la fruibilità del servizio? E in ultimo, aprendo ai contributi del pubblico l'edizione digitale diventa una "neverending edition" sempre mutevole: come gestire tale mutevolezza?

**L'originale del privilegio Religiosam vitam di papa Gregorio X per il Monte Sinai (1274).** Il privilegio Religiosam vitam fu concesso da papa Gregorio X in favore della comunità monastica di Santa Caterina del Monte Sinai nel 1274, insediamento monastico fortificato da Giustiniano e sopravvissuto alla conquista araba, alla controversia iconoclasta e alle crociate. Tra 1211 e 1276 la comunità monastica del Monte Sinai ottenne numerose lettere e concessioni sia dalla Chiesa di Roma sia dalla Repubblica di Venezia, a causa degli interessi in larga parte coincidenti delle due istituzioni. Roma cercava il sostegno della comunità per ogni progetto di crociata diretto al Mediterraneo orientale e nei tentativi di unione delle chiese latina ed orientale. Dal XIII secolo, con il passaggio dei territori di lingua greca dell'impero bizantino all'amministrazione veneziana (Stato da mar), Venezia vide nei monaci del Sinai un supporto per perseguire la pacificazione sociale con i sudditi greco-ortodossi principalmente a Creta. Nonostante questo, l'analisi dei testi conservati dimostra che spesso concessioni e privilegi erano più fittizi che reali.

La pubblicazione online MountSinai<sup>3</sup> offre alla ricerca e alla didattica un sistema interattivo per

- 1) leggere il documento
- 2) visualizzare e conoscere attraverso un grafico ad albero il tenor formularis del documento
- 3) esplorare l'aggregazione del patrimonio di informazione multidisciplinare prodotto da discipline quali la paleografia, la diplomatica e la storia e utile alla interpretazione del documento in esame
- 4) condividere in anteprima testi destinati alla rielaborazione per la pubblicazione accademica commerciale;
- 5) aggregare in tempo reale la letteratura secondaria (e.g., Europeana, Scopus Elsevier, Taylor and Francis) eventualmente utile alla discussione degli elementi via via presi in esame dall'utente.

Questa metodologia permette l'avanzamento della conoscenza in un sistema di pubblicazione aperto e partecipativo.

**I documenti dell'abbazia di S. Maria della Grotta.** L'edizione critica digitale collaborativa dei documenti dell'abbazia di S. Maria della Grotta (d'ora in poi SMG) nasce nel 2018 come una sperimentazione dell'editor della piattaforma Monasterium.Net. Essa, nata dagli sforzi di un gruppo internazionale, ha avuto la funzione precipua di testare le reali e attuali possibilità di produrre edizioni critiche di documenti online e di affrontare il problema del pieno riconoscimento di prodotti digitali di questo tipo nell'ambito della comunità scientifica di riferimento. Naturalmente a questa funzione sono sottese la ricerca di una soluzione pratica per interrogare e interpretare i dati in essa contenuta.

---

3 <https://engineeringhistoricalmemory.com/MountSinai.php>

Questo è il motivo per cui l'edizione ha attualmente diverse versioni:

- 1) l'edizione digitale nativa su Monasterium.Net, disponibile come visualizzazione user-friendly e anche come file utili al download in XML<sup>4</sup>;
- 2) il PDF dei soli testi, generati con una trasformazione dedicata delle edizioni su Monasterium.Net (XSLT, XSL-FO e un processore FO<sup>5</sup>);
- 3) l'edizione cartacea (2018) che contiene il testo e propone materiali aggiuntivi disponibili in Open Access su FEDOA<sup>6</sup>.

Uno degli aspetti sui quali riflettere per quanto riguarda SMG, e che anche Manfred Thaller ha sottolineato recentemente, è il fatto che essa, o qualsiasi altra edizione che verrà prodotta in futuro su Monasterium.Net, ha diverse potenzialità: si va a posizionare in un più ampio framework di fonti storiche digitalizzate, in continua espansione – il portale conta attualmente online più di novecentomila immagini di documenti e di dati ad essi connessi – man mano che vengono messi a disposizione da archivi, istituzioni di ricerca, università in 15 Paesi europei. Il lavoro di edizione di documenti medievali, tradizionalmente legato alla fruizione di un pubblico specialistico, si colloca in questo modo in un contesto più ampio che è quello rappresentato dalla progettualità delle sedi di conservazione europee: non solo la comunicazione online del materiale documentario (spesso attraverso registi o dati minimi), ma anche le strategie per la valorizzazione e la sua conservazione a lungo termine in ambiente digitale. In tal modo tende ad accorciarsi lo iato di iniziative che è esistito fin dall'inizio dello sviluppo della digitalizzazione dei documenti storici (ma anche dei libri) tra coloro che producono edizioni di documenti e coloro che li conservano. Inoltre, un'edizione, calata in questo contesto, può avvantaggiarsi dello sviluppo tecnologico costante di MOM-CA, l'ambiente collaborativo disponibile sul portale (Icarus - l'International Centre for Archival Research è responsabile della piattaforma, l'Università di Graz, nella persona di Georg Vogeler, di MOM-CA) e nello stesso tempo fornire agli altri gruppi di ricerca, che gravitano intorno a Monasterium, una mole di dati che possono essere utili in altri progetti (la base di dati di SMG, per esempio, viene attualmente utilizzata nell'ambito di un progetto internazionale basato sul riconoscimento automatico della scrittura, HIMANIS).

Infine, un'edizione collocata sul portale Monasterium permette a ciascun utente, da remoto, liberamente, il download delle riproduzioni fotografiche ad alta risoluzione dei documenti e dei dati in vari formati, tra i quali l'XML, per un eventuale riutilizzo, nonché di perfezionare e di inserire, sempre da remoto, ulteriori informazioni, con il filtro della pubblicazione online da parte dei responsabili del progetto, consentendo un vero e proprio crowdsourcing. Quindi un'edizione collocata sul portale, anche in forma di work in progress, ha molte possibilità di essere comunicata ad un pubblico vasto al quale si può chiedere finanche di collaborare ai dati.

---

4 [www.monasterium.net/mom/SMG1200-1250/collection](http://www.monasterium.net/mom/SMG1200-1250/collection)

5 <https://github.com/GVogeler/cei2pdf-SMG>

6 <http://www.fedoa.unina.it/12003/>

### **Elaborazione di informazioni paleografiche, named entities e ontologie nei testi medievali.**

Estendendo infine lo sguardo ad ambiti diversi da quello prettamente storico, l'evoluzione di EVT come strumento inizialmente concepito per la pubblicazione del Vercelli Book Digitale<sup>7</sup>, e di altre edizioni similmente incentrate su testi medievali, si caratterizza per una radicale trasformazione della base di codice, ma molto poco è cambiato per quanto riguarda il trattamento dei dati. Anche una funzionalità aggiunta di recente, il supporto per i fogli di stile VisColl<sup>8</sup>, è incentrata sul duplice obiettivo di visualizzazione e navigazione dei dati dell'edizione. La codifica utilizzata, tuttavia, rappresenta un potenziale tesoro di informazioni che attende solo di essere interrogato e messo a disposizione dell'utente.

I tre casi d'uso individuati, che non rappresentano certo la totalità di quanto possibile con un sistema più avanzato dell'attuale, sono:

- 1) l'elaborazione dei dati relativi ai caratteri speciali inseriti per mezzo della <charDecl> e <g>: questi caratteri rappresentano una preziosa fonte di informazione per individuare le abitudini e idiosincrasie scribali, in casi di studio precedenti questi calcoli sono stati fatti a mano (Conner 1986), mentre in un'edizione digitale potrebbero essere automatizzati;
- 2) per le named entities si applica lo stesso paradigma di elaborazione finalizzata a visualizzazione (nel contesto dell'edizione) e navigazione (grazie a liste separate, collegate al testo), ma sarebbe possibile aggiungere un livello di interrogazione dei dati del tutto nuovo, incrociando le informazioni disponibili in maniera tale da rispondere a interrogazioni complesse (v. le domande di ricerca relative alla Lira senese del XV sec.);
- 3) questo tipo di funzionalità potrebbe essere implementato grazie all'uso di ontologie, direttamente nella marcatura TEI se possibile, altrimenti usando gli strumenti più appropriati come il linguaggio RDF.

**La Lira senese del XV secolo.** A Siena negli ultimi anni sono state avviate alcune esperienze di ricerca che, mosse da un condiviso approccio di storia sociale, sono finite per convergere verso una metodologia di analisi prosopografica applicata alla storia cittadina. La serie di database realizzati nel corso del tempo ha creato l'esigenza di far interagire tra loro ed implementare queste banche dati facendo dialogare fonti diverse, di periodi diversi. Si tratta quindi di iniziative che non hanno guardato in primo luogo a progetti di edizione ma delle vere e proprie esperienze di lavoro e di schedatura condotte sui dati storici ricavati da estimi e catasti senesi Tre-Quattrocenteschi. L'aumento della massa dei dati ha permesso certo di ampliare la ricerca in termini cronologici e qualitativi ma ha evidenziato talvolta le criticità nel far comunicare fra loro informazioni diverse.

Ultimo in ordine di tempo è la schedatura degli imponibili fiscali e delle denunce autografe dei contribuenti del 1453. Tale lavoro permette adesso di collegare immediatamente sia l'imponibile che le denunce di 3.745 nuclei familiari senesi. A questo primo database è stato affiancato un altro relativo al contenuto delle denunce espandendo i contenuti anagrafici, patrimoniali e sociali di ogni individuo.

---

7 <https://www.collane.unito.it/oa/items/show/11>

8 <https://viscoll.org/>

Tuttavia l'elaborazione delle informazioni in banche dati non permette al momento di creare in maniera agile interrogazioni complesse mantenendo, al contempo, un rapporto diretto con i circa 9.000 documenti originali. Questi infatti, in assenza di trascrizioni integrali, rimangono colmi di informazioni paleografiche, linguistiche e storiche difficilmente frazionabili in database.

In altre parole risulta ancora deficitaria la possibilità di effettuare calcoli complessi sui dati raccolti. Quanti imprenditori, originari di un'altra regione, residenti in un'area specifica della città, investirono in società laniere con un capitale superiore ai 500 fiorini? Quante di queste avevano almeno due soci? Quanti individui, all'interno di nuclei formati da 5 componenti, contribuivano al reddito familiare? Quali erano gli immobili di maggior valore in un determinato quartiere e da chi erano posseduti? Quanti di questi avevano al proprio interno un'attività produttiva?

Queste sono solo alcune delle domande possibili che al momento trovano, invero, ugualmente risposta a discapito però del tempo necessario dettato dalla consueta analisi da parte dello storico.

Se quindi da una parte le competenze informatiche 'pure' degli addetti ai lavori rimangono indispensabili per la risoluzione di tali questioni, d'altra parte è doveroso chiedersi quanto e come gli storici possano ricorrere ai linguaggi di markup già nelle fasi iniziali di schedatura e/o trascrizione delle fonti storiche al fine di agevolare i successivi sviluppi.

### **3 SPERIMENTAZIONE**

Grazie a questo incontro è stato possibile discutere le tematiche relative ai singoli casi di studio, e arrivare a conclusioni preliminari utili a implementare questa funzionalità nel software EVT<sup>9</sup> e in un applicativo di tipo stand-alone da sviluppare all'interno del sistema interattivo EHM<sup>10</sup>. Una prima sperimentazione in tal senso si svolgerà nella prima metà del 2021 con l'obiettivo di rendere più sofisticato il motore di ricerca di EVT, in modo da rendere possibile una interrogazione dei dati del tipo sopra descritto. Caratteristica precipua dello sviluppo di questo strumento è sempre stata l'attenzione all'interfaccia utente e alla facilità d'uso, obiettivo specifico sarà l'individuazione di una interfaccia sufficientemente flessibile da prestarsi a esigenze diverse.

I metodi e gli algoritmi di interrogazione implementati in EVT potranno servire come modello, se non sarà possibile un impiego diretto, per le altre piattaforme citate.

### **4 CONCLUSIONI**

Quelle presentate e discusse sono esperienze certamente diverse ma che, dal punto di vista del trattamento dei dati che offrono agli studiosi, presentano alcune problematiche comuni.

Innanzitutto la problematica del legame tra interfaccia, strumenti di ricerca e dati. Chi fa un'edizione digitale si deve porre il problema di presentare il testo in forma tale da poter servire o da vera propria base di ricerca per quanto riguarda il testo stesso, o da punto di partenza per ricerche interdisciplinari: filologiche, storiche, geografiche e economiche, tanto per partire dalle principali domande che si possono porre a una fonte scritta.

---

9 <http://evt.labcd.unipi.it/>

10 [www.engineeringhistoricalmemory.com](http://www.engineeringhistoricalmemory.com)

L'altra problematica, condivisa pur nelle specificità di ciascun progetto di ricerca, è quella dell'uso limitato che si fa di dati spesso marcati in maniera complessa e sofisticata. Riuscire a sfruttare tali dati per rispondere alle domande fondamentali del ricercatore, oppure per sollevarlo da un'indagine manuale che risulterebbe del tutto anacronistica al momento in cui si dispone di mezzi di calcolo così potenti, costituirebbe un importante servizio reso alla comunità accademica.

I metodi possono ovviamente essere diversi, come diversi i pubblici a cui si pensa di rivolgersi, ma si deve resistere anche alla tentazione dell'originalità e dell'esaustività a ogni costo e valutare con cautela rapporto tra qualità scientifica e costo di fabbricazione, tra preziosità del tessuto e vestibilità per tutti.

Discutere di interfaccia e di strumenti di ricerca significa quindi di fatto discutere di pubblico, quindi di utenti, quindi di accessibilità e di pratiche di partecipazione (crowdsourcing). Vediamo troppo di frequente emergere progetti di Storia o Filologia digitali che riescono a parlare solo alla ristretta comunità scientifica di riferimento. Lo sforzo, che questo gruppo di ricerca ha iniziato a fare, è quello di ragionare su come le DH possono risolvere il problema della scarsa comunicazione tra le discipline e tra un progetto editoriale e il suo potenziale pubblico e questo passa attraverso una ricerca vera e propria e una modifica degli strumenti e metodi dell'edizione digitale.



## BIBLIOGRAFIA

1. Afferni, R., A. Borgna, M. Lana, P. Monella, and T. Tambassi. “... But What Should I Put in a Digital Apparatus?” A Not-so-Obvious Choice. New Types of Digital Scholarly Editions’. In *Advances in Digital Scholarly Editing. Papers Presented at the DiXiT Conferences in The Hague, Cologne, and Antwerp*. Leiden: Sidestone Press, 2017.
2. Ambrosio, Antonella. ‘Can Online European Documents Be a Valuable Resource for the Informal Learning of Adults? The Cases of the Monasterium.Net and Topotheque Portals in the Coop Project’. *Research Trends in Humanities Education & Philosophy (Quaderno Think Tank Immersive University Education, a Cura Di Flavia Santoianni)* 6 (2019): 22–30.
3. ———. ‘Digital Critical Editions of Medieval Documents on Monasterium.Net. The Digital Edition of the Documents of the Abbey S. Maria Della Grotta’. *Digitizing Medieval Sources. L’édition En Ligne de Documents d’archives Médiévaux: Enjeux, Méthodologie et Défis, Colloque International-Centre de Recherche Universitaire Lorrain d’Histoire (CRULH) de Nancy (Nancy, 9-10 June 2016)* 27 (2020): 60–71.
4. ———. ‘La Diplomatica e Il Digitale. Il Fondo Della Biblioteca Della Società Napoletana Di Storia Patria Online’. In *Digitizing Medieval Sources. L’édition En Ligne de Documents d’archives Médiévaux: Enjeux, Méthodologie et Défis*. Atelier de Recherche Sur Les Textes Médiévaux 27. Turnhout: Brepols, 2020.
5. ———. ‘L’edizione Critica Digitale Dei Documenti Medievali. Le Forme Degli Atti Di Octavianus Notarius’. In *Ingenita Curiositas. Studi Sull’Italia Medievale per Giovanni Vitolo*. Battipaglia: Laveglia & Carlone, 2018.
6. Ambrosio, Antonella, Vera Isabell Schwarz-Ricci, and Georg Vogeler. ‘Tra Analogico e Digitale. L’edizione Dei Documenti Dell’abbazia Di Santa Maria Della Grotta’. In *I Documenti Dell’abbazia Di S. Maria Della Grotta Di Vitulano (BN), 1200–1250*. Battipaglia: Laveglia & Carlone, 2018.
7. Ambrosio, Antonella, Georg Vogeler, Vera Isabell Schwarz-Ricci, and Georg Vogeler. *I Documenti Dell’abbazia Di S. Maria Della Grotta Di Vitulano (BN)*. Fonti per La Storia Del Mezzogiorno Medievale 26. Battipaglia: Laveglia & Carlone, 2018.
8. Conner, Patrick-W. ‘The Structure of the Exeter Book Codex (Exeter, Cathedral Library, MS. 3501)’. *Scriptorium* 40, no. 2 (1986): 233–42. <https://doi.org/10.3406/scrip.1986.1448>.
9. Di Pietro, Chiara, and Roberto Rosselli del Turco. ‘La Visualizzazione Di Edizioni Digitali Con EVT: Una Soluzione per Edizioni Diplomatiche e Critiche’. *Ecdotica* 17, no. 1 (2020): 148–173.
10. Jackson, Cornell. ‘Using Social Network Analysis to Reveal Unseen Relationships in Medieval Scotland’. *Digital Scholarship in the Humanities* 32, no. 2 (1 June 2017): 336–43. <https://doi.org/10.1093/llc/fqv070>.
11. Nanetti, Andrea. ‘L’originale Del Privilegio “Religiosam Vitam” Di Papa Gregorio X per Il Monte Sinai (1274): Edizione e Commento’. *Studi Di Storia Del Cristianesimo. Per Alba Maria Orselli*, 2008, 235–68.

12. Nanetti, Andrea, and Siew Ann Cheong. 'Computational History: From Big Data to Big Simulations'. In *Big Data in Computational Social Science and Humanities*, edited by Shu-Heng Chen, 337–63. Computational Social Sciences. Cham: Springer International Publishing, 2018. [https://doi.org/10.1007/978-3-319-95465-3\\_18](https://doi.org/10.1007/978-3-319-95465-3_18).
13. Rosselli Del Turco, Roberto. 'Designing an Advanced Software Tool for Digital Scholarly Editions': *Textual Cultures* 12, no. 2 (29 August 2019): 91–111. <https://doi.org/10.14434/textual.v12i2.27690>.
14. Rosselli Del Turco, Roberto, Giulia Cacioli, Angelo Maria Del Grosso, Chiara Di Pietro, Chiara Martignano, and Jurgen Memaj. *Edition Visualization Technology 2 Beta 2*. Pisa: Laboratorio di Cultura Digitale, 2020. <https://github.com/evt-project/evt-viewer>.
15. Salvatori, Enrica. 'Digital (Public) History: La Nuova Strada Di Una Antica Disciplina'. *RiMe. Rivista Dell'Istituto Di Storia Dell'Europa Mediterranea*, no. Vol. 1/I n.s. (December 2017) (Dicembre 2017): 57–94. <https://doi.org/10.7410/1291>.
16. ———. 'L'edizione Digitale Di Fonti Storiche: Nuove Opportunità, Nuovi Problemi, Nuove Figure'. In *Edito, Inedito, Riedito. Saggi Dall'XI Congresso Degli Italianisti Scandinavi, Università Del Dalarna–Falun, 9-11 Giugno 2016*, 2017.
17. Salvatori, Enrica, Roberto Rosselli Del Turco, Chiara Alzetta, Chiara Di Pietro, Chiara Mannari, and Alessio Miaschi. 'Il Codice Pelavicino tra edizione digitale e Public History'. *Umanistica Digitale*, 1 October 2017, No 1 (2017). <https://doi.org/10.6092/ISSN.2532-8816/7232>.
18. Siemens, Ray, Meagan Timney, Cara Leitch, Corina Koolen, Alex Garnett, and INKE with the ETCL and PKP Research Groups. 'Toward Modeling the Social Edition: An Approach to Understanding the Electronic Scholarly Edition in the Context of New and Emerging Social Media\*'. *Literary and Linguistic Computing* 27, no. 4 (1 December 2012): 445–61. <https://doi.org/10.1093/lc/fqs013>.
19. Thaller, Manfred. 'Archivi Ed Edizioni: Alcune Promesse Mantenute. Un Punto Di Vista Personale'. In *I Documenti Dell'abbazia Di S. Maria Della Grotta Di Vitulano (BN) 1200-1250*. Battipaglia: Laveglia & Carlone, 2018.
20. Vogeler, Georg. 'The "Assertive Edition"'. *International Journal of Digital Humanities* 1, no. 2 (1 July 2019): 309–22. <https://doi.org/10.1007/s42803-019-00025-5>.

## ULTERIORI RIFERIMENTI

Edizione digitale del Codice Pelavicino

<http://pelavicino.labcd.unipi.it/>

Monasterium.net:

<https://www.monasterium.net/mom/home? lang=ita>

Edizione digitale dei documenti dell'abbazia di S. Maria della Grotta di Vitulano (BN), 1200-1250

<https://www.monasterium.net/mom/SMG1200-1250/collection>

XSLT-Transformazione per *I documenti dell'abbazia di S. Maria della Grotta*

<https://github.com/GVogeler/cei2pdf-SMG>

Ambrosio, Antonella, Vera Isabell Schwarz-Ricci, e Georg Vogeler, a c. di. *I documenti dell'abbazia di S. Maria della Grotta di Vitulano (BN). 1200-1250*. Fonti per la storia del Mezzogiorno medievale 26.

Battipaglia (Sa): Laveglia & Carlone, 2018

<http://www.fedoa.unina.it/12003>

Nanetti, Andrea (a cura di), *Engineering Historical Memory*

<https://engineeringhistoricalmemory.com>

# Who wrote the erotic novel Josefine Mutzenbacher? Speculations, theories, and stylometric analyses

Simone Rebora<sup>1</sup>, Massimo Salgaro<sup>2</sup>

<sup>1</sup> Università di Verona – simone.rebora [at] univr.it

<sup>2</sup> Università di Verona - massimo.salgaro [at] univr.it

## ABSTRACT

The paper proposes a series of stylometric analyses aimed at attributing the erotic novel Josefine Mutzenbacher (1906) to one of its possible authors: Felix Salten, Arthur Schnitzler, Hermann Bahr, Peter Altenberg, Hugo von Hofmannsthal, Ernst Klein, and Willi Handl. After reviewing the discussion among literary scholars, two analyses have been performed: one that combines 480 different methods to compare the seven candidate authors; one that verifies the attribution using the “impostors” method. Results show how the most probable author is Felix Salten, more commonly known for his children’s book Bambi.

## KEYWORDS

Josefine Mutzenbacher, Austrian literature, stylometry, authorship attribution.

## TALK

### 1 INTRODUCTION

*Josefine Mutzenbacher oder Die Geschichte einer Wienerischen Dirne von ihr selbst erzählt* (engl. Josefine Mutzenbacher or the story of a Viennese whore as told by herself), published in Vienna in 1906, has surely constituted one of the most fascinating cases of authorship attribution in German literature. First of all, because this novel is “the only so-called pornographic novel by a German writer that has achieved the rank of world literature” [17]. This singular *Bildungsroman* (novel of development) depicts the childhood of the prostitute Mutzenbacher from the age of five to fourteenth years. Due to its massive success, it was followed by two sequels, widely translated, reprinted, and adapted into several films.

Secondly, because it has generally been attributed to Felix Salten, the author of the world-famous *Bambi*, but his authorship has never been proven. Other authors have been suspected of writing this erotic masterpiece. Among them is also Arthur Schnitzler who, in 1902, had published *Reigen*, a cycle of amorous adventures that caused scandal in the Viennese society. In addition, both *Reigen* and *Josephine Mutzenbacher* were published by the same publisher, Wiener Verlag. Some scholars maintain that this scandalous text was the result of a “bet” by the Jung-Wiener writers Hermann Bahr, Peter Altenberg, and Hugo von Hofmannsthal [14]; others state that it was written by less-known writers such as Ernst Klein [16; 10] and Willi Handl [6]; even the publisher Fritz Freund has been included in the list of suspects [6; 15].

Thirdly, this controversy on authorship attribution became a juridical case. In 1988, a German court rejected Salten’s heirs claim for royalties, since the documents and witness accounts they delivered were unable to demonstrate without fail Salten’s authorship of the *Mutzenbacher* novel [8].

This unsolved quest for the author of the novel is the starting point of our analysis.

## 2 STYLOMETRIC ANALYSIS

### CORPUS CONSTRUCTION

In order to test all theories and speculations proposed so far, we built a corpus of digitized texts, by collecting material from multiple sources such as Project Gutenberg<sup>1</sup>, Projekt Gutenberg-DE<sup>2</sup>, and Kolimo<sup>3</sup>. In four cases, we manually scanned and OCR-ed documents using the *Transkribus* platform [12]: manual evaluation of randomly-selected passages showed a character-error rate of about 2%, much below the minimum threshold for reliable authorship attribution [4; 9].

To compare the candidates for authorship in a genre-coherent setup, we collected novels and short stories by seven authors: Felix Salten, Arthur Schnitzler, Hermann Bahr, Peter Altenberg, Hugo von Hofmannsthal, Ernst Klein, and Willi Handl. The only candidate we had to leave out of the analysis is Fritz Freund, of which no novels or short stories (but just some lyrics and a collection of letters) are available.

Finally, we normalized any possible typographic discrepancy in the texts by transforming into digraphs all Umlauts and the Eszett (ä -> ae; ö -> oe; ü -> ue; ß -> ss).

---

1 <https://www.gutenberg.org/>

2 <https://www.projekt-gutenberg.org/>

3 <https://www.projekt-gutenberg.org/>

An overview on the composition of the corpus is provided by Table 1.

| Author                | Number of texts | Number of words | Source                      |
|-----------------------|-----------------|-----------------|-----------------------------|
| Arthur Schnitzler     | 40              | 777,234         | <i>Kolimo</i>               |
| Felix Salten          | 12              | 421,000         | <i>Projekt Gutenberg-DE</i> |
| Hugo von Hofmannsthal | 18              | 203,656         | <i>Kolimo</i>               |
| Hermann Bahr          | 4               | 203,482         | <i>Kolimo</i>               |
| Ernst Klein           | 3               | 169,377         | OCR                         |
| Peter Altenberg       | 195             | 92,276          | <i>Kolimo</i>               |
| Willi Handl           | 1               | 52,441          | OCR                         |

Table 1. Corpus composition<sup>4</sup>

## ANALYSIS SETUP

All stylometric analyses<sup>5</sup> were performed by using the functions of the R package *Stylo* [5], which enables an efficient application of Burrows’s Delta [1] and its numerous variants. To maximize the reliability of results through multi-methodology [11], a total of 480 different analyses were performed on the basis of word frequency, modifying respectively:

- 1 the number of most frequent words (MFW) to be taken into consideration (from 50 to 2,000, with 50-word steps);
- 2 the measure to calculate the distances between texts (Burrows’s Delta, Cosine Delta, Eder’s Delta, and Canberra);
- 3 the culling percentage (0%, 20%, and 50%), to tentatively counter-balance the possible influence of thematic similarities.

For each of the 480 analyses, then, a quality metrics was produced by adopting the “simple score” described by [7]: this, in order to preserve the potential of multi-methodology, acknowledging at the same time the varying efficiency of different approaches.

Before running the analyses, all texts of the candidate authors were split into 5,000-word-long chunks (being it the minimum text length for reliable authorship attribution, cf. [3]): this produced between 10 and 155 measurements for each author. Of all these measurements (i.e. stylometric distances between Josefine Mutzenbacher and the other texts in the corpus), only the shortest three were compared for each candidate author. This approach (inspired by the “rolling.classify” function in the *Stylo* package) offered the possibility to identify with the maximum level of precision the passages with a style closest to that of the *Mutzenbacher*.

---

4 The text of Josefine Mutzenbacher was downloaded from Project Gutenberg.

5 All scripts and corpora used for the analyses are available on Github.  
[https://github.com/SimoneRebora/Josfine\\_Mutzenbacher\\_stylometry](https://github.com/SimoneRebora/Josfine_Mutzenbacher_stylometry)

## RESULTS

Results of the analyses are presented in Table 2, which shows the percentage of attributions for each candidate, balanced through the “simple score” quality metrics.

| <b>Author</b>         | <b>Percentage of attributions</b> |
|-----------------------|-----------------------------------|
| Felix Salten          | 78.41 %                           |
| Arthur Schnitzler     | 12.26 %                           |
| Ernst Klein           | 9.33 %                            |
| Peter Altenberg       | 0 %                               |
| Hermann Bahr          | 0 %                               |
| Willi Handl           | 0 %                               |
| Hugo von Hofmannsthal | 0 %                               |

Table 2: final results of the analysis

Felix Salten clearly dominates in the attribution, followed at a long distance by Arthur Schnitzler and Ernst Klein. This stylometric analysis seems to confirm the dominant (but never-confirmed) theory among scholars, that the author of the scandalous *Mutzenbacher* is also the future inspirer of Disney’s children’s classic.

This is confirmed also by a curious coincidence, when *Bambi* (in a separated analysis ran on the non-split corpus) clusters closely to the *Mutzenbacher* (see Figure 1).

Josefine\_Mutzenbacher\_stylometry  
Cluster Analysis

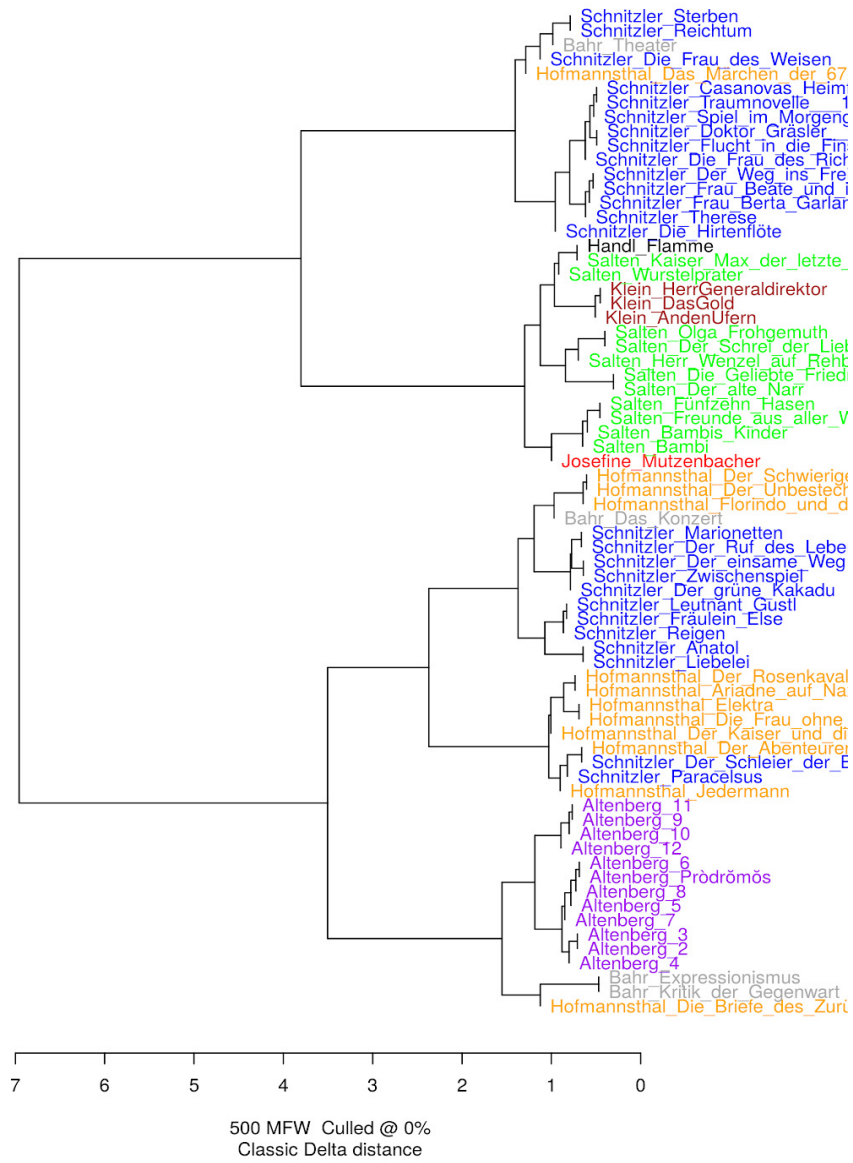
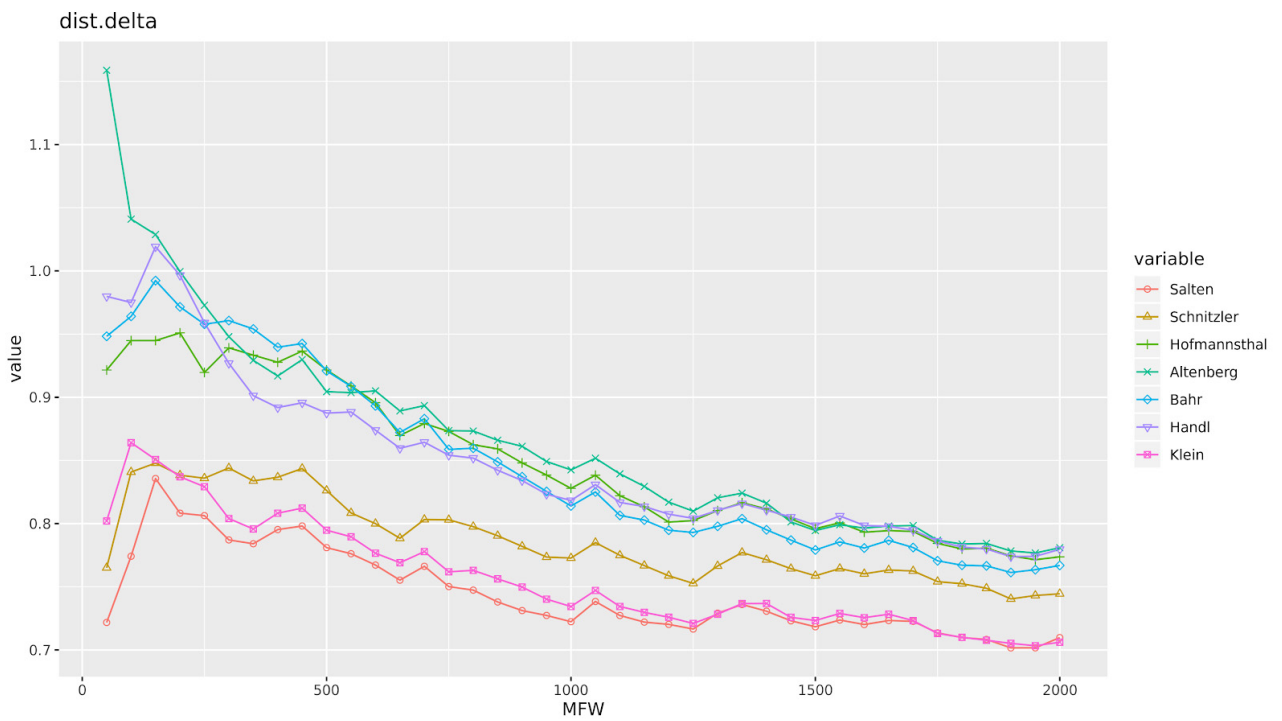


Figure 1: dendrogram showing a close connection between Bambi and Mutzenbacher





clearly excluding that they could be the authors of the *Mutzenbacher*.

| Author                | Authorship probability |
|-----------------------|------------------------|
| Felix Salten          | 53 %                   |
| Peter Altenberg       | 0 %                    |
| Hermann Bahr          | 0 %                    |
| Willi Handl           | 0 %                    |
| Hugo von Hofmannsthal | 0 %                    |
| Ernst Klein           | 0 %                    |
| Arthur Schnitzler     | 0 %                    |

Table 3: authorship probability verified with the “impostors” method.

Each percentage indicate the probability of the candidate author against the impostors

## 4 CONCLUSION

Through the combination of two stylometric approaches, our analysis confirmed how Felix Salten is the most probable author of the erotic novel *Josefine Mutzenbacher*. In methodological terms, we have shown how computational methods can be fruitfully paired with traditional research practices, by building on top of their critical theories and proposals; by testing, verifying, and eventually disproving them.

However, we cannot but consider the conclusion of this work as provisional. Future research should be devoted for example to finding ways to include the publisher Fritz Freund in the analysis; or to verify theories according to which the ending of the *Mutzenbacher* was written by a different hand [15; 6].

All these possibilities can still open new scenarios for this attributive problem. Still, with our work for the first time we have provided scientific evidence in support of a hypothesis that, despite having been debated for a century, had never found an actual confirmation.

## REFERENCES

1. Burrows, John. “Delta”: A Measure of Stylistic Difference and a Guide to Likely Authorship’. *Literary and Linguistic Computing* 17, no. 3 (2002): 267–287. <https://doi.org/10.1093/lcl/17.3.267>.
2. Eder, Maciej. ‘Authorship Verification with the Package Stylo’. *Computational Stylistics Group Blog* (blog), 2018. <https://computationalstylistics.github.io/blog/imposters/>.
3. ———. ‘Does Size Matter? Authorship Attribution, Small Samples, Big Problem’. *Digital Scholarship in the Humanities* 30, no. 2 (2015): 167–182. <https://doi.org/10.1093/lcl/fqt066>.
4. ———. ‘Mind Your Corpus: Systematic Errors in Authorship Attribution’. In *Procediings DH 2012*. Hamburg, 2012. [https://sites.google.com/site/computationalstylistics/preprints/m-eder\\_mind\\_your\\_corpus.pdf?attredirects=0](https://sites.google.com/site/computationalstylistics/preprints/m-eder_mind_your_corpus.pdf?attredirects=0).
5. Eder, Maciej, Jan Rybicki, and Mike Kestemont. ‘Stylometry with R: A Package for Computational Text Analysis’. *The R Journal* 8, no. 1 (2016): 107–21.

6. Englisch, Paul. 'Geschichte Der Erotischen Literatur (1927)'. In *Josefine Mutzenbacher Oder Die Geschichte Einer Wienerischen Dirne von Ihr Selbst Erzählt*, by Felix Salten, edited by Michael Farin, 507–8. München: Schneekluth, 1990.
7. Evert, Stefan, Thomas Proisl, Fotis Jannidis, Isabella Reger, Steffen Pielström, Christof Schöch, and Thorsten Vitt. 'Understanding and Explaining Delta Measures for Authorship Attribution'. *Digital Scholarship in the Humanities* 32, no. suppl\_2 (2017): ii4–ii16. <https://doi.org/10.1093/llc/fqx023>.
8. Farin, Michael. 'Geschichte Der Erotischen Literatur (1927)'. In *Die Letzten Illusionen. Josefine Mutzenbacher Vor Gericht*, 511–45. München: Schneekluth, 1990.
9. Franzini, Greta, Mike Kestemont, Gabriela Rotari, Melina Jander, Jeremi K. Ochab, Emily Franzini, Joanna Byszuk, and Jan Rybicki. 'Attributing Authorship in the Noisy Digitized Correspondence of Jacob and Wilhelm Grimm'. *Frontiers in Digital Humanities* 5 (2018): 4. <https://doi.org/10.3389/fdigh.2018.00004>.
10. Hall, Murray. 'Josef Und Josefine Mutzenbacher. Oder Recycling Der Pornographie'. In *Erotisch-Pornografische Lesestoffe: Das Geschäft Mit Erotik Und Pornografie Im Deutschen Sprachraum Vom 18 Jahrhundert Bis Zur Gegenwart*, edited by Christine Haug, Frimmel Johannes, Vogel Anke, Teresa Lang, and Adam Franz, 88:159–81. Buchwissenschaftliche Beiträge. Wiesbaden: Harrassowitz Verlag, 2015.
11. Juola, Patrick. 'The Rowling Case: A Proposed Standard Analytic Protocol for Authorship Questions'. *Digital Scholarship in the Humanities* 30, no. suppl\_1 (2015): i100–i113. <https://doi.org/10.1093/llc/fqv040>.
12. Kahle, Philip, Sebastian Colutto, Günter Hackl, and Günter Mühlberger. 'Transkribus-A Service Platform for Transcription, Recognition and Retrieval of Historical Documents'. In *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, 4:19–24. IEEE, 2017. <https://doi.org/https://doi.org/10.1109/ICDAR.2017.307>.
13. Koppel, Moshe, and Yaron Winter. 'Determining If Two Documents Are Written by the Same Author'. *Journal of the Association for Information Science and Technology* 65, no. 1 (2014): 178–187. <https://doi.org/10.1002/asi.22954>.
14. Liebrand, Claudia. 'Josefine Mutzenbacher Die Komödie Der Sexualität.' In *Felix Salten. Schriftsteller – Journalist – Exilant*, by Felix Salten, 82–89. edited by Siegfried Matzl and Werner Michael Schwarzer. Wien: Holzhausen, 2006.
15. Schidrowitz, Leo. 'Geschichte Der Erotischen Literatur (1927)'. In *Bilderlexikon Der Erotik (1929), Band 2, S. 692, 508.510*. München: Schneekluth, 1990.
16. Weichinger, Robert. 'Schluß Mit Genuß! Wer Schrieb Die "Mutzenbacher"? Wer Kriegt Die Tantiemen? Felix Saltens Erben Wohl Nicht'. *Die Presse*, Autumn 1991.
17. Wiener, Oswald. 'Geschichte Der Erotischen Literatur (1927)'. In *Der Obszöne Wortschatz Wiens. Beiträge Zu Einer Ädöologie Des Wienerischen (1969)*, 361–462. München: Schneekluth, 1990.

# Pirandello Nazionale: per un nuovo modello di edizione digitale, collaborativa e integrata

Milena Giuffrida<sup>1</sup>, Christian D'Agata<sup>2</sup>, Laura Giurdanella<sup>3</sup>, Pietro Sichera<sup>4</sup>

<sup>1</sup> Università di Catania – milenagiuffrida [at] gmail.com

<sup>2</sup> Università di Catania – christian.dagata [at] gmail.com

<sup>3</sup> Università di Catania – laura.giurdanella [at] unict.it

<sup>4</sup> ISTC-CNR – pietro.sichera [at] istc.cnr.it

## SINTESI

Il seguente contributo si propone di presentare il portale *Pirandello Nazionale*, edizione digitale dell'Opera Omnia di Luigi Pirandello. Il portale integra prodotti scientifici e supporti alla didattica, al fine di promuovere una fruizione innovativa dei testi dello scrittore agrigentino. La sinergia tra curatori ed editori dell'edizione cartacea e il gruppo di ricerca che si occupa dell'ambiente digitale permette di annoverare *Pirandello Nazionale* tra gli esempi virtuosi di edizione collaborativa, attenta alle esigenze di un'utenza particolarmente variegata, quale quella dei lettori di uno dei più importanti autori novecenteschi.

## PAROLE CHIAVE

Luigi Pirandello, didattica della letteratura, filologia digitale, concordanze, public humanities

## INTERVENTO

### 1 INTRODUZIONE

*Pirandello Nazionale*<sup>1</sup> è un portale dedicato all'Opera Omnia di Luigi Pirandello, scrittore agrigentino, vincitore del premio Nobel nel 1934, le cui opere sono tra le più lette e rappresentate nel mondo. *Pirandello Nazionale* nasce da un progetto ideato dall'équipe del CINUM (Centro di Informatica Umanistica dell'Università), coordinata da Antonio Sichera e Antonio Di Silvestro, composta da filologi, italianisti e informatici afferenti all'Università di Catania.

Il portale è patrocinato dal Comitato per l'Edizione Nazionale, nominata dal MIBACT, presieduta da Angelo Pupino (Unior), che dal 2017 ha avviato i lavori per la ripubblicazione delle opere dello scrittore in collaborazione con l'editore Mondadori.

---

1 <https://www.pirandellonazionale.it/>

*Pirandello Nazionale* svolge quindi una funzione di integrazione ai risultati prodotti dai curatori dell'Edizione Nazionale nella tradizionale modalità cartacea, proponendosi quale «spazio di irradiazione e di approfondimento (filologico, linguistico, critico, didattico e culturale) della grande letteratura, oltrepassando i modelli usuali della rete quale luogo virtuale della pura archiviazione di testi e di materiale»<sup>2</sup>.

## 2 UN APPROCCIO READER-ORIENTED

Il criterio adottato nella costruzione di *Pirandello Nazionale* è stato di tipo *top-down*: muove cioè da una riflessione sui mutamenti nell'accesso al patrimonio testuale, determinati dalla pervasiva diffusione del digitale, per sviluppare strumenti che possano ovviare ad alcune delle criticità rilevate. Infatti, lo straordinario successo e la versatilità delle opere di Pirandello – unite alla libertà dei suoi testi dai vincoli di copyright –, attraendo un pubblico molto vasto ed eterogeneo di lettori, hanno favorito negli ultimi anni la proliferazione di edizioni digitali, organizzate anche sotto forma di portali. Ma questi prodotti, sebbene abbiano l'indubbio merito di contribuire alla divulgazione dell'opera dello scrittore, non risultano sempre particolarmente curati o attendibili sul piano filologico. Inoltre, spesso ci si limita alla digitalizzazione di testi già pubblicati in formato cartaceo e/o, nel migliore dei casi, di letteratura critica precedentemente apparsa in rivista o in volume.

La rilevazione di questi dati pone di conseguenza due interrogativi strettamente correlati:

1. ciò che leggiamo con estrema facilità in rete è ciò che Pirandello ha davvero scritto?
2. quanto questi testi risultano comprensibili in tutta la loro portata?

Il primo quesito ha trovato una immediata risposta nella stretta connessione tra il portale e l'Edizione Nazionale, per antonomasia luogo volto a stabilire, restituire e rendere imperiture le opere letterarie in una forma filologicamente validata. In *Pirandello Nazionale* è così possibile leggere, direttamente online e in modalità *open access* (sotto vari formati dei quali si parlerà più avanti), i lavori dello scrittore le cui edizioni sono state curate da alcuni tra i più importanti studiosi di Pirandello. La sfida più impegnativa è stata invece quella di produrre strumenti nell'ambito dell'analisi del testo e della critica letteraria; strumenti che accompagnino le opere e permettano di analizzarle ed esplorarle a diversi livelli, al fine di favorirne il processo di interpretazione da parte degli utenti.

L'adozione di un approccio che fosse il più possibile *reader-oriented* non può prescindere dall'identificazione e classificazione delle effettive esigenze dei lettori, riconoscibili principalmente in studenti, studiosi e appassionati.

La progettazione degli strumenti ospitati da *Pirandello Nazionale* ha pertanto cercato di rispondere ai bisogni specifici di queste categorie. Ispirandosi allo stesso principio, si è scelto di dotare il portale di un'interfaccia estremamente intuitiva che permetta una duplice modalità di interrogazione: per sezioni di interesse (*Manoscritti, Vocabolari, Edizioni, Strumenti per la scuola*) e per singola opera (ogni opera raccogliendo nella propria pagina i prodotti delle sopracitate sezioni che la riguardano). A seconda quindi del proprio livello di competenza e curiosità, il lettore può muoversi agilmente attraverso i numerosi materiali ospitati dal sito, individuando con estrema facilità ciò che più lo interessa.

---

2 <https://www.pirandellonazionale.it/presentazione/>

### 3 UN KNOWLEDGE SITE

Più che edizione o archivio, *Pirandello Nazionale* può essere definito come un *knowledge site*: un portale nel quale la rappresentazione dei problemi del testo e del documento, caratteristica di molte edizioni digitali, convive alla pari con strumenti di indagine lessicografica ed ermeneutica.

Questi sono stati declinati a diversi livelli, privilegiando la componente didattica. Non bisogna dimenticare infatti che Pirandello è l'autore del Novecento al quale viene dedicato maggiore spazio all'interno dei programmi scolastici di secondo grado; successo derivatogli soprattutto dall'attualità dei contenuti delle sue opere, riassumibili in concetti quali – tra i più celebri – l'opposizione vita-forma, la maschera, il doppio. La facilità di lettura di romanzi e novelle, favorita senza dubbio dall'utilizzo di una lingua media e di una prosa moderna, nasconde però una densità e complessità filosofica, nonché riferimenti culturali e letterari difficili spesso anche solo da individuare per il lettore non specialistico. Si è quindi constatata la necessità di dare forma ad una sorta di guida che permetta ai docenti (ma anche agli studenti più curiosi, che hanno la capacità di muoversi in autonomia) di avventurarsi nei meandri del sistema di pensiero pirandelliano con maggiore consapevolezza. La sezione del portale dedicata agli strumenti didattici, intitolata *e-Pirandello*, si suddivide in due aree: *Risorse multimediali* e *FuoriClasse*. Nel primo segmento si trovano raccolte tutte le risorse multimediali disponibili sul sito, suddivise sulla base delle opere alle quali fanno riferimento: documenti, audio, video, grafici e statistiche (*data visualization*) elaborati sulla base della frequenza dei lemmi, alberi genealogici relativi ai personaggi dei romanzi, topografie letterarie e *timeline* degli eventi storici che fungono da cornice alle diverse opere. I materiali di questo repertorio, vastissimo e strutturato, sono spesso presentati come pptx navigabili in maniera personalizzata a seconda dell'intenzione di lettura.

*FuoriClasse* contiene invece risorse multimediali organizzate in percorsi didattici di tipo tematico, costruiti con l'obiettivo di:

- favorire una corretta comprensione del testo;
- proporre una vivace attualizzazione dei testi;
- immaginare una visita virtuale nel laboratorio pirandelliano, nell'officina dello scrittore;
- rendere fruibili dati lessicografici mediante l'utilizzo di Data visualization;
- proporre un approccio critico-ermeneutico dell'opera, pensandola come un «prospettarsi sul mondo» (Sichera 2019) con il fine ultimo della divulgazione;
- rispondere alla “Terza Missione” dell'Università, dunque, dialogare con la scuola e la società, farsi promotrice di uno scambio culturale, sociale ed educativo.

I diversi percorsi didattici sono stati creati tenendo bene a mente i principi psicodidattici dell'approccio umanistico-affettivo, in cui ‘umanistico’ – opposto per sua natura ai metodi meccanicistici dello strutturalismo – comporta la centralità dell'affettività, della fisicità, della relazionalità proprie dell'individuo.

I percorsi didattici guardano alla struttura messa a punto dalle più recenti teorie di glottodidattica [1] ovvero, 'M + GAS', in cui M sta per 'motivazione', G per 'globalità', A per 'analisi', S per 'sintesi'. Tali fasi trovano corrispondenza in diverse sezioni:

- Pirandello e noi (M), intesa come primo luogo e momento di incontro propizio tra lo studente-lettore e l'opera. Il contatto è previsto mediante la presentazione di documenti audio-visivi che si pongono quali strumenti necessari per accedere alla chiave di lettura privilegiata del testo, determinata dai bisogni emotivi e/o linguistici dell'apprendente;
- Pirandello commenta Pirandello (M) che vorrebbe favorire ora l'incontro diretto con l'autore mediante fonti extra-testuali di mano pirandelliana, quali interviste, saggi, corrispondenze, conferenze radiofoniche ecc.;
- Raccontare Pirandello (G e A) in cui vengono proposti i percorsi di senso che consentono l'esplorazione ermeneutica del significato globale dell'opera e, successivamente, la penetrazione nella "grammatica" del testo per mezzo di un approccio più analitico grazie all'utilizzo di strumenti, quali, ad esempio, Data Visualization; infine, Attività (S) che, lungi dal presentare meri esercizi di comprensione, costituisce piuttosto un'ultima occasione di interazione dello studente con il testo. Anzi, più che mai questa fase cerca «la collaborazione del lettore come interprete attivo e datore di senso» [9].

Tale sequenza si ispira in particolare alla psicologia della Gestalt [2], secondo la quale il percorso della percezione – seguendo l'ordine naturale di apprendimento (ascoltare > parlare > leggere > scrivere) – è composto essenzialmente di tre fasi (una globale, una analitica ed una sintetica) che trasformano ciò che è 'percepito' in elementi 'recepiti' dalla nostra mente e, nella migliore delle ipotesi, se vi sono le condizioni, in elementi 'acquisiti'. Tale modello potrebbe configurarsi come una sorta di *format* didattico, altrettanto efficace per favorire un'esperienza del testo che si voglia, da un lato, filologica, dunque autentica e saldamente legata al testo, e dall'altro interessante, accattivante e soprattutto dinamica e interattiva.

Le risorse didattiche per scuola e università sono costruite sfruttando prodotti scientifici, anch'essi ospitati dal portale, destinati soprattutto ad un'utenza iper-specializzata (gli studiosi di Pirandello). In questo ambito, in *Pirandello Nazionale* si distinguono tre sezioni: *Manoscritti*, *Edizioni* e *Vocabolario*. In *Manoscritti* vengono ospitate le riproduzioni digitali di manoscritti e dattiloscritti di opere pirandelliane, con relative trascrizioni a fronte. Si è finora preferito presentare questo tipo di materiali attraverso le applicazioni più diffuse (come il pdf sfogliabile in flip book), per garantire la maggiore accessibilità possibile ad ogni tipo di utente, consapevoli però della poca attenzione prestata alla componente estetica. Questa scelta è stata in parte dettata anche dalla predilezione di un profilo didattico. Infatti, studenti non universitari e lettori occasionali sono tendenzialmente poco interessati alla semplice presentazione del documento, che rappresenta certo un *surplus* e – se ben congegnata – ha un potere di attrazione determinato dalla fascinazione che questo tipo di prodotti produce, destinata però a svanire dopo una prima visualizzazione, perché poco proficua per l'arricchimento del lettore non specialistico.



La scelta di un accesso facilitato e aperto è stata perseguita anche per la sezione *Edizioni*. In questa, infatti, è possibile leggere le edizioni critiche delle opere di Pirandello, realizzate, per venire incontro alle aspettative dei filologi tradizionali, con il consueto apparato a piè di pagina e proposte in formato digitalizzato.

In alternativa è possibile accedere alle edizioni anche in versione sinottica, cioè con l'accostamento dei diversi testimoni dell'opera e l'evidenziazione delle porzioni di testo coinvolte in variante. Per ottenere questo tipo di visualizzazione si è proceduto con una collazione automatica – realizzata inizialmente attraverso l'algoritmo di allineamento di *Juxta*<sup>3</sup>, successivamente attraverso quello di *Variance Viewer*<sup>4</sup> – poi ampliata per permettere l'accostamento simultaneo di più di due testi.

Uno dei punti di forza di *Pirandello Nazionale* è senza dubbio la sezione relativa al vocabolario di Pirandello. Il *Vocabolario*<sup>5</sup> è stato realizzato grazie al *Liotro Concord v2*, software sviluppato da Pietro Sichera appositamente per il CINUM sfruttando la tecnologia del database relazionale 4D.

*Liotro Concord* è stato progettato per mettere insieme due diversi aspetti tecnici relativi alla lingua: 1) il modello Savoca per la realizzazione della concordanza, basato sull'associazione di ogni singola parola del testo a un record del vocabolario; 2) l'incremento dell'automatizzazione del processo di concordanza, dovuto a un sistema di autoapprendimento che permette di associare automaticamente occorrenze e record. Il processo di lemmatizzazione viene eseguito dal software in maniera automatica grazie ad un procedimento di auto-apprendimento della lingua italiana, a partire da una base dati di lemmi (e relative declinazioni) che costituiscono il dizionario di macchina. L'intervento umano permette di correggere i limitati errori del processo di lemmatizzazione automatica o gli eventuali errori nel testo, tramite combo box e menu contestuali che rendono il procedimento rapido e privo di errori procedurali. Il testo in input viene spezzato in token attraverso l'uso di alcune espressioni regolari (tramite un parser di espressioni regolari sviluppato in proprio denominato RegExLab) e l'interfaccia permette di navigare tra i token, offrendo la possibilità di intervenire sulla categorizzazione di ogni singola forma, modificandola o creando un nuovo lemma per identificarla. A conclusione del processo di revisione (durante il quale i singoli paragrafi vengono inoltre associati alle pagine dell'edizione nazionale cartacea, che costituisce sempre il testo di riferimento) la concordanza viene integrata nel database del portale, disponibile per un'interrogazione per forme o lemmi che attraversi la singola opera o tutto il *corpus* a disposizione. Se invece si volesse avere una panoramica completa della concordanza di una singola opera, o avere accesso alle liste di frequenza, queste possono essere visualizzate in un pdf consultabile mediante un flip book, ottenuto in output dallo stesso *Liotro Concord* che conferisce alle informazioni del database l'aspetto di una classica concordanza cartacea.

---

3 <https://www.juxtaoftware.org/>

4 <http://variance-viewer.informatik.uni-wuerzburg.de/Variance-Viewer/>

5 <http://vocabolari.pirandellonazionale.it/>

## 4 CRITICITÀ E OBIETTIVI FUTURI

Accessibilità, scientificità dei contenuti, approccio *reader-oriented* sono certamente i punti di forza di *Pirandello Nazionale*. Molto però rimane da fare, soprattutto in termini di condivisione, diffusione e replicabilità dei risultati effettivi della ricerca. Al momento infatti non è possibile visualizzare i codici delle applicazioni utilizzate all'interno del sito e i testi presenti non sono stati sottoposti a codifica XML. Negli ultimi mesi però è stato avviato un progetto per superare queste particolari criticità e permettere una maggiore interazione tra tutte le sezioni del portale.

La grande varietà dei testi pirandelliani (tragedie, romanzi, novelle, saggi, poesie) ha suggerito lo sviluppo di una codifica del testo che rispettasse le specificità di ogni genere letterario, permettendo una rappresentazione di molteplici livelli testuali. La codifica del testo avrà dunque due fondamentali obiettivi:

- rappresentare il livello strutturale del testo (divisione in atti, capitoli, paragrafi), ancorando il livello testuale al numero di pagina dell'edizione nazionale cartacea;
- individuare per ogni testo le entità nominate (luoghi e personaggi) e descrivere in modo esaustivo le occorrenze dei dialoghi così da permettere ricerche testuali avanzate.

Attraverso la descrizione degli interlocutori presenti in ogni scambio comunicativo si potrà usufruire in modo nuovo anche della concordanza, che non si legherà più all'intero testo, ma si potrà orientare all'analisi del lessico dei diversi personaggi ottenendo anche gli indici di frequenza relativi al singolo personaggio. Infatti, mediante la GUI di *Liotro Concord*, il testo di riferimento di ogni opera verrà taggato in TEI e lemmatizzato. Il processo di lemmatizzazione automatica e la taggatura TEI verranno entrambe eseguite in maniera rapida, attraverso la stessa interfaccia grafica che mostrerà un *preset* di tag individuati all'inizio della codifica (ma sempre ulteriormente implementabili con un menu a parte). Il marcatore potrà accedere ad una serie di funzionalità che permetteranno di rendere più accurato e veloce il processo di taggatura (come l'associazione automatica di alcuni lemmi a un `<persName>`). I tag verranno segnalati all'utente tramite una descrizione in linguaggio naturale che proporrà all'utente anche la possibilità di inserire determinati attributi per quel tag, attributi scelti in fase di modellizzazione della codifica da parte del direttore scientifico della stessa. In questo modo, ogni occorrenza di parola verrà associata sia al lemma (nella fase della lemmatizzazione) che a uno o più tag (nella fase della codifica) all'interno del Database relazionale che soggiace a tutta l'architettura. Successivamente, attraverso un algoritmo di *merge* verranno combinate le informazioni della lemmatizzazione (ad es. pos e lemma `< fingere >`) e quelle presenti nella codifica (ad es. `<q who="Enrico IV">`), rendendo così possibile la ricerca di tutte le occorrenze di un lemma usato da un personaggio e la creazione di vocabolari selettivi per personaggio.

```
<sp who="#andolfo" toWhom="#bertoldo"><speaker>Landolfo</speaker><stage>(a Bertoldo come seguitando una spiegazione).</stage><p>E questa è la sala del trono!</p></sp>
<sp who="#arialdo" toWhom="#bertoldo"><speaker>Arialdo</speaker><p>A <placeName ref="#goslar">Goslar!</placeName></p></sp>
<sp who="#ordulfo" toWhom="#bertoldo"><speaker>Ordulfo</speaker><p>O anche, se vuoi, nel Castello dell'<placeName ref="#hartz">Hartz</placeName>!</p></sp>
<sp who="#arialdo" toWhom="#bertoldo"><speaker>Arialdo</speaker><p>O a <placeName ref="#worms">Worms</placeName>. </p></sp>
<sp who="#andolfo" toWhom="#bertoldo"><speaker>Landolfo</speaker><p>Secondo la vicenda che rappresentiamo, balza con noi, ora qua, ora là.</p></sp>
<sp who="#ordulfo" toWhom="#bertoldo"><speaker>Ordulfo</speaker><p>In <placeName ref="#sassonia">Sassonia</placeName>!</p></sp>
<sp who="#arialdo" toWhom="#bertoldo"><speaker>Arialdo</speaker><p>In <placeName ref="#lombardia">Lombardia</placeName>!</p></sp>
<sp who="#andolfo" toWhom="#bertoldo"><speaker>Landolfo</speaker><p>Sul <placeName ref="#reno">Reno</placeName>!</p></sp>
<sp who="#valletto1" toWhom="#ordulfo #arialdo #bertoldo"><speaker>Uno dei valletti</speaker><stage>(senza scomporsi, muovendo appena le labbra).</stage><p>Ps! Ps!</p></sp>
```

Figura 1: esempio di marcatura TEI dell'Enrico IV

Infine, la creazione di divisioni testuali (<div>) verrà eseguita attraverso un'interfaccia rapida che offrirà una visione logica del testo (strutturata per paragrafi). L'utente quindi non scriverà mai codice XML puro, ma verrà guidato alla realizzazione dello stesso. Tale metodo impedirà overlapping di tag. Per una visualizzazione immediata e analitica dell'output, un processo parallelo all'interfaccia di codifica mostrerà il TEI risultante sia come testo XML che come lista gerarchica multilivello.

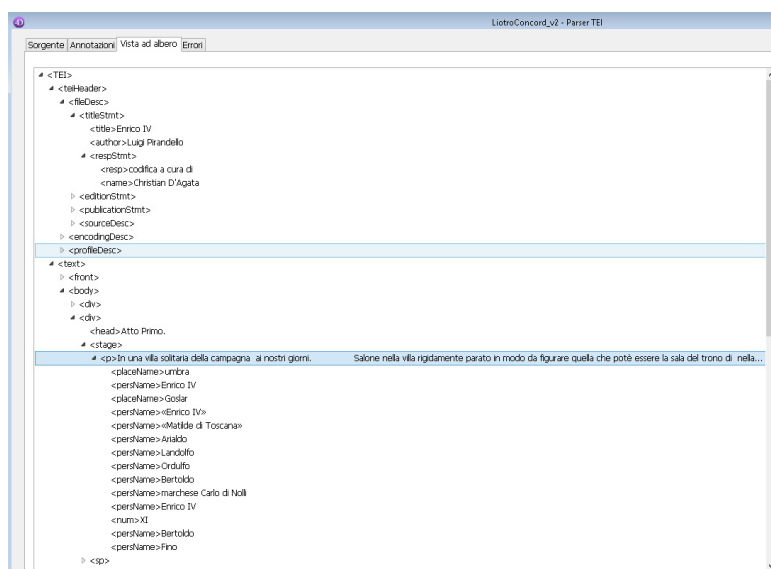


Figura 2: visualizzazione lista gerarchica output TEI Liotro Concord

Infine, ruolo preminente verrà affidato allo sviluppo di un'interfaccia che integri tutte le risorse presenti in Pirandello Nazionale con l'obiettivo di costruire un'edizione 'umentata', che proponga un dialogo continuo tra testi (con relativa codifica), concordanza, edizione critica, risorse didattiche e multimediali, manoscritto (permettendo un miglioramento anche nell'ambito della rappresentazione degli autografi pirandelliani).

Le principali caratteristiche di questa interfaccia si potranno individuare nel suo essere:

- dinamica (restituire i dati a seguito dell'interrogazione del DB);
- asincrona (permettere la navigazione mentre carica i diversi contenuti);
- responsive (adattarsi automaticamente alla pagina e al device usato).

Salone nella villa rigidamente parato in modo da figurare quella che poté essere la sala del trono di Enrico IV nella casa imperiale di Goslar. Ma in mezzo agli antichi arredi due grandi ritratti a olio moderni, di grandezza naturale, avventano dalla parete di fondo, collocati a poca altezza dal suolo su uno zoccolo di legno lavorato che corre lungo tutta la parete (largo e sporgente in modo da poterlisi mettere a sedere come su una lunga panconata), uno a destra e uno a sinistra del trono che, nel mezzo della parete, interrompe lo zoccolo e vi si inserisce coi suo seggio imperiale e il suo basso baldacchino. I due ritratti rappresentano un signore e una signora, giovani entrambi, camuffati in costume carnevalesco, l'uno da «Enrico IV» e l'altra da «Matilde di Toscana». Usci a destra e a sinistra.

Al levarsi della tela, i due valletti, come sorpresi, balzano dallo zoccolo su cui stanno sdraiati, e vanno a impostarsi come statue, uno di qua e uno di là ai piedi del trono, con le loro alabarde. Poco dopo dal secondo uscio a destra entrano Arialdo, Landolfo, Ordulfo e Bertoldo, giovani stipendiati dal marchese Carlo di Nollì perché fingano le parti di «Consiglieri Segreti», vassalli regali della bassa aristocrazia della Corte di Enrico IV. Vestono perciò in costume di cavalieri tedeschi del secolo XI. L'ultimo, Bertoldo, di nome Fino, assume ora per la prima volta il servizio. I tre compagni lo ragguagliano pigliandoselo a godere. Tutta la scena va recitata con estrosa vivacità.

LANDOLFO (a Bertoldo come seguitando una spiegazione) È questa è la sala del trono!

ARIALDO A Goslar!

ORDULFO O anche, se vuoi, nel Castello dell'Hartz!

ARIALDO O a Worms.

LANDOLFO Secondo la vicenda che rappresentiamo, balza con noi, ora qua, ora là.

ORDULFO In Sassonia!

ARIALDO In Lombardia!

fingere - verbo - Occorrenze: 7 - Percentuale: 0,03574%

| Chiave       | Contesto SX                    | Occorrenza | Contesto DX                   |
|--------------|--------------------------------|------------|-------------------------------|
| EN AI 65 3   | marchese Carlo di Nollì perché | fingano    | le parti di «Consiglieri      |
| EN AI 74 7   | sua amica ha bisogno di        | fingere    | di non saperlo; ridere sempre |
| EN AI 76 14  | (con le spalle voltate,        | fingendo   | di richiamarlo di nascosto)   |
| EN AI 97 2   | - Nessuno di noi mente o       | finge      | ! C'è poco da dire: ci siamo  |
| EN AI 112 12 | accosta ai due giovani e       | finge      | di dar loro spiegazioni.      |
| EN AI 113 4  | DONNA MATILDE Ma dobbiamo      | fingere    | prima di licenziarli!         |
| EN AI 125 6  | nella sala del trono? -        | Fingo      | di non vederla...             |

Figura 3: prototipo della nuova interfaccia

Essa, infatti, si basa su un modello a due finestre [6; 7] e permetterà sul lato sinistro dello schermo la possibilità di scorrere il testo (seguendo il numero di pagina dell'edizione cartacea dell'Opera Omnia di Luigi Pirandello) e sul lato destro renderà possibile la visualizzazione di tutte le risorse associate presenti nel database (inserite in sede di codifica e lemmatizzazione). In questo modo l'utente potrà contestualmente leggere il testo e passare alla concordanza, accedere alle risorse didattiche e multimediali, visualizzare il pdf dell'edizione cartacea e l'immagine del manoscritto.

## 5 CONCLUSIONI

*Pirandello Nazionale*, come progetto generale ma anche in ogni sua singola sezione, è il prodotto di un lavoro di collaborazione costante tra diverse istituzioni e studiosi, ognuno dei quali si occupa di una specifica mansione, tenendo però sempre presente gli obiettivi generali dell'edizione digitale (accessibilità, scientificità, attenzione al lettore). Il portale può essere quindi annoverato tra gli esempi virtuosi di edizione collaborativa, proprio in virtù della costante e fruttuosa interazione tra le diverse anime degli studi testuali: quella filologica e tradizionale dell'Edizione Nazionale, quella innovativa della filologia digitale e quella ermeneutica degli strumenti interpretativi, rigorosi sul piano scientifico e al contempo attenti agli articolati bisogni dei lettori dell'opera pirandelliana.

## BIBLIOGRAFIA

1. Balboni, Paolo E. *Le sfide di Babele: Insegnare le lingue nelle società complesse*. UTET università, 2008.
2. Becchi, E. *La pedagogia della "Gestalt"*. Firenze: La Nuova Italia, 1961.
3. Ciotti, Fabio. «Il testo e l'automa: saggi di teoria e critica computazionale dei testi letterari». Aracne, 2007.
4. Coronato, Rocco, e Sara Gesuato. «The Literature-Linguistics Interface–Bridging the Gap Between Qualitative and Quantitative Approaches to Literary Texts». *Umanistica Digitale* 3, n. 6 (2019). <https://umanisticadigitale.unibo.it/article/view/9601>.
5. Italia, Paola. *Editing Duemila*. Roma: Salerno, 2000.
6. Kiernan, Kevin. «Electronic Beowulf». *Electronic Beowulf*, 2016. <http://ebeowulf.uky.edu/ebeo4.0/CD/main.html>.
7. Rosselli Del Turco, Roberto, Chiara Di Pietro, e Chiara Martignano. «Progettazione e implementazione di nuove funzionalità per EVT 2: lo stato attuale dello sviluppo». *Umanistica Digitale*, n. 7 (2019). <https://doi.org/https://doi.org/10.6092/issn.2532-8816/9322>.
8. Russo, Ada, e Cristina Marras. «Progettare, realizzare e promuovere contenuti culturali digitali. Un esempio di didattica integrata: il progetto "Filosofia & Migrazioni"». *Umanistica Digitale* 4, n. 8 (2020).
9. Sichera, Antonio. *Ermeneutiche. Punti di vista sul confine*. Leonforte: Euno Edizioni, 2019.
10. Sichera, Antonio, Antonio Di Silvestro, e Michael Subialka. «The future of Pirandello: in the new Digital Edition of Pirandello's Opera Omnia». *The Journal of The Pirandello Society of America* XXXII (2019): 107–16.
11. Stella, Francesco. *Testi letterari e analisi digitale*. Carocci, 2018.
12. Tomasi, Francesca. *Edizioni o archivi digitali? Knowledge sites e apporti disciplinari*. Roma: Sapienza Università Editrice, 2016.

# **Editing (and publishing) medieval vernacular inscriptions in a digital environment: potential and limitations**

Nadia Cannata  
Università di Roma La Sapienza – [nadia.cannata \[at\] uniroma1.it](mailto:nadia.cannata@uniroma1.it)

## **ABSTRACT**

I should like to discuss briefly the advantages of an integrated system of publication – digital and traditional – in relation to the critical editing and linguistic analysis of medieval texts, with particular reference to texts of a non-literary nature, composed in the numerous Romance varieties in use in Italy in the late Middle Ages. The digital publication constitutes a ground-breaking innovation in the methods of textual and linguistic analysis, and ultimately for the historical interpretation of our linguistic past. This is linked both to the potential for interrogation of marked up texts, and the invaluable information about the ‘grammar of writing’ and the spoken languages underlying written records it yields, and to the fact that important datasets, acquired independently by different teams of scholars, across time and space, are expanding very significantly the corpora of texts available to scholars and open to digital interrogation. Therefore we can engage in analysis and cross-referencing of textual and linguistic (sometimes even material) features of medieval Romance texts in the numerous languages in which they were composed and gauge our findings in a much wider context than traditional philological methods would allow – a fact that does not only make more data available and brought to scholarly attention, but actually has the potential to influence our perception of history very significantly. A change in the means of documenting the past cannot but result in an important reassessing of its significance and of our abilities to interpret it.

## **KEYWORDS**

cultural history, orality and writing, multilingualism, digital editions, medieval vernaculars, digital epigraphy

# TALK

## 1 INTRODUCTION

I have devoted the best part of ten years to a major digital project: *EDV: Epigraphic Database of Inscriptions in Italian Vernaculars (IXth-XVth c.)* now nearing conclusion. It records the corpus of all extant vernacular inscriptions produced in Italy from the middle of the 9th to the end of the 15th century, provided they were meant to be displayed publicly, are still extant, and were composed in any language other than Latin. The database currently records around 570 inscriptions nicely spread across the whole period.

The aim of the study is fundamentally threefold: to preserve a record of a significant and hitherto largely unknown corpus of medieval texts of a non-literary nature; to provide an archive of documentary evidence of the uses of language(s) other than Latin in public script in late Medieval and Early Modern Italy; and to devise a powerful research tool to study such languages.

It is my contention that the first two aims are best served by traditional forms of publication, in print and in book form, whilst the research potential unleashed by the possibility of interrogating texts marked up in XML is such that it should become an integral part of the research methods applied to the study of the linguistic heritage of the past. We are planning to publish the complete catalogue of inscriptions and its study in book form, as well as to keep the website open to update and add materials as appropriate. The two forms of publication are devised independently but designed to support one another. Whilst the updating of data is an open-ended task, which benefits greatly from the flexibility of digital means, scholarship needs books and libraries for which have provided for thousands of years the most reliable and durable form of cultural transmission.

Moreover, digital projects are very costly and do not guarantee sustainability over time. The issue of the long term preservation of the outcomes of such projects is, therefore, a pressing one. An integration between digital and traditional means of publication might help address such issues, and allow wide dissemination and accuracy of data, as well as durability in time, thus ensuring that the very significant investment of time and resources that normally goes into big digital projects reaps its expected rewards.

I shall devote the next two paragraphs of this contribution to illustrate the theoretical issues and research questions that I believe digital resources enable us to address in a way not possible before; then I shall say something more about how this is working for the EDV project.

## 2 RESEARCH QUESTIONS AND EXPECTED RESULTS

At the time of writing, we have recorded for the entire corpus the general features (date, place of origin, location, type of object, material, language(s) used, function, script) and the complete bibliography available to date for each inscription, all openly available and easily browsable on our website.

We have also completed the diplomatic transcription and critical edition of one hundred and twenty inscriptions from Central and Southern Italy and marked up the text in Epidoc, which is an established implementation of TEI, used widely by epigraphers working on inscriptions from the most diverse cultures, times, and languages.

For these, we were, therefore, able to generate complex indexing of spelling systems and linguistic features, which is allowing us to catalogue important information for the interpretation of the linguistic data gathered in the corpus. It is a known fact to sociolinguistics that any sociolinguistic investigation depends upon the ability to engage with a comprehensive dataset and to contextualize whatever the linguistic document under scrutiny is into the wider picture of general language use. When studying documents written in languages which were not standardized, the relationship between spelling and pronunciation, or the influence of the standard on common usage, and of the usage on the perceived standard, as well as the possibility of comparing such data with a contemporary production of a different nature yields information that can prove very interesting to understand linguistic heritage. It is particularly useful to identify language uses which – because of their marginality or lack of prestige – may well have escaped our attention. As a philologist and textual critic mostly engaged in the study of languages and literary traditions in Romance Europe, with particular reference to Italy and its vernaculars, I have often wondered through what methods – other than those which historical linguistics and Romance Philology, in particular, have passed down to us – might we be able to add to our knowledge of the languages actually spoken in Medieval and Early Modern Europe, given that the available documentation testifying to the linguistic culture of that time and place is constituted by written records. Attempting to engage in historical sociolinguistics is so much harder since it is very questionable if the languages whose use was solely, or mostly, oral and confined to everyday exchanges are somehow accounted for in writing, or rather if – as one would be inclined to conclude – they have mostly fallen through the net of historical memory, due to the lack of positive means apt to secure their preservation.

Interestingly, such veil of haze surrounding the language varieties in most frequent use in the period and places under consideration is mirrored by the lack of appropriate terms we have at our disposal to describe them. When considering the languages in use in Italy in the period 5th – 13th c. one has to contend, as a preliminary consideration, with the fact that our description of the languages of Italy in the Middle Ages is based on categories which were not shared by the communities whose languages are the focus of our attention.



There are around five centuries of linguistic usage from the 5th to at least the 9th c. which appear unaccounted for, since we would be quite unable to answer the simple question “What language(s) were spoken in Italy in the 5th, 6th, 7th, and 8th centuries?” We define languages in use in Italy after the 9th c. by using as our pole of attraction no longer Latin, but the dominant literary vernaculars at the time: Sicilian, Tuscan, Provençal, Old French, Catalan and so on, which does not help with the issue of how to account for languages not graced with a written tradition, nonetheless largely in use by local communities. But even when we move forward in time, to consider the centuries between the 10th and the 15th we are at pains at finding names for the languages recorded by epigraphy or documents (which are widely understood to record ‘natural’ everyday language more faithfully than other kinds of text) and those supposedly spoken other than as Tuscan, Central Italian, Sicilian etc. vernaculars.

Roger Wright and Michel Banniard, among others, have argued extensively about the need to devise a new paradigm to understand language change, its periodization and to deal with the issues that the rigid protocols of 19th c. philology seems to have passed to us and which now seem to hinder, to an extent, our understanding of linguistic and cultural change in Neo-Latin Europe.

In this context I believe that digital interrogation of marked-up texts might contribute to our finding an answer to two important research questions, closely related to one another, which have rather escaped us so far: 1) Is there a way for us to retrieve the voices of languages used in Romance Europe before the birth of ‘Italian’, ‘French’, ‘Spanish’, or indeed of ‘Catalan’ ‘Occitan’ ‘Galego portuguese’ etc. and after the demise of ‘Latin’, thence restoring them as far as possible to memory?

2) Might it be possible to develop a methodology which will allow us to better understand the grey zone connecting written and spoken language, and the hitherto scarcely charted territory from which Italian, as other modern Romance languages, issued and was subsequently circulated and preserved?

These issues may seem of a ‘pure’ philological and linguistic nature, but they can only be addressed if we look at language questions in the wider perspective of cultural history. For nearly a millennium in most of Europe Latin obliterated to posterity the wide variety of languages spoken in the 1000 years or so of the Roman domination in Europe. In the early modern period more languages graduated to writing, and we date their birth based on their written records. They do not document, however, the birth of new languages (which were alive long before any of them came to be written), but rather, more crucially, the development of a culture of writing which enabled them to acquire a written form and therefore to transfer a memory of their existence in a form which is quintessentially different from the features we would like to believe are preserved. The creation of a new written culture also entails the birth of a new shared written culture and a common language to serve the purpose. Modern Europe was born multilingual, indeed as a consequence of a process through which new voices, many issuing from Latin, established themselves as independent from one another.

Writing has had the formidable task of constituting the fundamental means for the composition, reading, transmission and preservation of texts. All written documents are also linguistic documents and survive thanks to material support. It follows that any critical approach to the study of text and language cannot ignore either textual transmission, or, indeed, palaeography and the study of the history of writing in its material aspects.

It is my contention that it would be impossible to study and understand fully any written message if we were unable to decode not just the components that form it but the grammar of writing presiding over the way in which that language is written. Any critical approach to the study of texts cannot ignore either textual transmission or, indeed, the study of the material aspects (scripts, writing surfaces, actors) of linguistic documents.

There appears, therefore, to be a third research question which begs to be answered:

3) Is here a way to separate, in as far as possible, the issues of ‘how they spoke’ from ‘how they wrote’ and to define, accordingly, language use in previously uncharted areas?

### 3 IN SEARCH OF A NEW METHODOLOGY

Digital publication allows us to assemble online and interactive libraries of texts, marked up through TEI P5 XML to interrogate phonology, spelling vs pronunciation, morphology, syntax, lexis, writing materials, scripts. We have applied this to the edition of epigraphic texts in the medieval vernaculars in use in Italy up to the 15th c., and are proceeding to cross-reference this body of texts with other corpora of texts located and dated with a reasonable degree of accuracy, in order to establish the social context in which the documents were produced and their areas of circulation, and trace chains of influence between Latin and modern vernaculars, and across the vernaculars. This is helping to promote the appreciation of the sociolinguistics and the sociology of writing in late Medieval and Early Modern Italy, and the multilingual reality in communities in which Latin and vernaculars coexisted and nurtured each other; and it guides us in a better understanding of the components of the *scripta* by relating the linguistic evidence of the documents ‘horizontally’ to each other, rather than to ideal linguistic standards which had not yet been established.

To begin with, we were thus able to identify 15 different vernaculars and produce indexes of the morphological and syntactic features through which we have categorized them.

As regards devising a methodology which would allow us at least to attempt to answer our research questions, here again, digital interrogation is essential, since digital tools can support an innovative interaction between historical linguistics and the history of writing by allowing the cross-referencing of huge amounts of textual data which, through their sheer volume, may give us better grounds to infer the relation between sign and sound in recording languages with a very limited written tradition.

An area of particular interest in this respect is constituted by spelling, for example in the number of graphic solutions elaborated to account for the development of palatals (which varied greatly in time from area to area and according to the script adopted, albeit not as haphazardly as one may think). The possibility of interrogating thousands of texts to look for the ways in which palatal and affricate consonants are represented in script would allow us to acquire new information about the varying distance over time of sound and sign. As an example: many Christian inscriptions display COIUX for ‘spouse’. More often than not the form is considered as simply lacking the required *titulus*. However, the frequency with which the *titulus* slipped from the chisel, when seen alongside the use of ‘I’ for the palatal /j/ in later Latin and vernacular documents (capital ‘I’ used when small ‘i’ would be expected in Beneventan script to signal in an unequivocal way a pronunciation of the palatal /j/) invites the inference that the palatal had developed much earlier than current scholarship allows and that written culture strove to accommodate it.

Similarly the differing graphic solutions for /ʃ/, /ʎ/, /ɲ/ provide some indication on the pronunciation of sounds which failed to be preserved as phoneme in Italian (but constitute phonemes in Italian Romance vernacular varieties) because they did not find an accepted graphic representation; whereas the wealth of solutions adopted to represent /tʃ/ /dʒ/, /ts/, /dz/ suggests that for certain areas from the Latin consonants coupled with a glide (S+J, L+J, N+J, D+J, T+J) the development of an affricate rather than the palatal and thus give indications as to the pronunciation which has previously escaped our radar. We have noticed, for example, developments of spellings specific of certain regions: in texts written in Southern Italy, the grapheme <cz> as opposed to <ç> or <z> is generally used for /ts/. This use has proven to be systematic and contradicts common wisdom about the development of affricates in southern dialects of the Neapolitan type.

Similarly, the graphic solutions adopted in Christian inscriptions, both in Latin and in Greek alphabets, once studied systematically, have yielded very precious information about language development in Rome 5th-9th c. by allowing relations which have resulted in the identification of phenomena belonging to Romance linguistic varieties at a very early stage. They, however, were not recorded in script and therefore belong for linguists to the realm of late or vernacular Latin, or else Protoromance, without acquiring the identity of a modern language.

Such detailed analysis of spelling features is fundamental to try and chart a sociolinguistic of lost language varieties and to understand matters relating to the history of linguistic culture which cannot be studied independently from the history of written culture. The complex indexing we have elaborated for our data which can be perused freely at [www.edv.uniroma1.it/it](http://www.edv.uniroma1.it/it) testify to it.

## 4 CONCLUSIONS

Digital publication allows for the interrogation of the materials in ways traditional publication does not countenance. Its limitations consist mainly of the volatility of the repository of the data. This, consequently, poses a tangible risk for the dissemination and impact in the long term of the results achieved. Scholars must therefore make sure they elaborate a system integrating digital and more traditional means of publication so that both can support and enhance each others' potential. Only a solid institutional network constituting an international platform into which the data are fed and to which institutions are committed can counterbalance such volatility and help unleash the research potential of connecting a great wealth of data that may normally not 'communicate' among each other and help to interrogate them with the systematic approach with which only texts in TEI and machine-readable forms can afford as well as ensure their preservation to memory in the long run.

It might also be added that digital means allow for a wealth of new forms of dissemination, designed to encourage the general public to develop an interest in matters traditionally confined to the close cloisters of scholarship, by helping 'exhibit' and share whichever progress we make in learning through the visualization in websites (or even in actual physical spaces) of maps for the geo-localization of the documents, sound and image-files to hear the voices and bring back to life lost everyday language, or helping the viewer to elaborate some personal itinerary across the documents [14].

## REFERENCES

1. Banniard, Michel. «Viva Voce. Communication écrite et communication orale du IVe au IXe siècle en Occident latin». *Vita Latina* 115, n. 1 (1989): 42–46.
2. Bruni, Francesco. *Italia. Vita e avventure di un'idea*. Bologna: Il Mulino, 2010.
3. Cannata, Nadia. «Le parole sono pietre. Lingua communis e lingua literata in alcune epigrafi romane (secc. IV VI)» *Studi Romanzi*, n. 10 (2014).
4. ———. «Lontano da dove? Tradizioni culturali e coscienza linguistica in Europa». *Studi Romanzi* 12 (2016): 2016.
5. Cannata, Nadia, Margaret J.-M. Sönmez, e Maia Wellington Gahtan. *Museums of Language and the Display of Intangible Cultural Heritage*. Routledge, 2019.
6. Ciociola, Claudio, a c. di. *Visibile parlare: le scritture esposte nei volgari italiani dal Medioevo al Rinascimento*. Napoli: ESI, 1997.
7. Ingrand-Varenne, Estelle. *Langues de bois, de pierre et de verre. Latin et français dans les inscriptions médiévales*. Paris: Classiques Garnier, 2017.
8. Lepschy, Giulio C. «Ethnicity and language in Italy». In *L'Italia allo specchio*. Venezia: Marsilio, 2015.
9. Lori Sanfilippo, Isa, e Giuliano Pinto. *Comunicare nel Medioevo. La conoscenza e l'uso delle lingue nei secoli XII-XV*. Roma: Istituto Storico Italiano per il Medioevo, 2015.
10. Maiden, Martin, John Charles Smith, e Adam Ledgeway. *The Cambridge History of the Romance Languages: Volume 1, Structures*. Cambridge: Cambridge University Press, 2010.
11. Signorini, Maddalena. «Il ritmo cassinese: cultura grafico-libraria e qualche proposta di correzione». In *Scrivere il volgare fra Medioevo e Rinascimento*. Pisa: Pacini, 2009.
12. Vincent, Nigel. «Languages in Contact in Renaissance Italy». In *Rethinking Languages in Contact. The Case of Italian*. Oxford: Oxford University Press, 2006.
13. Wright, Roger. *A SocioPhilological Study of Late Latin*. Turnhout: Brepols, 2002.
14. ———, a c. di. *Latin and the Romance Languages*. London: Routledge, 1990.



# The digital Gazetteer of Ancient Arabia: an example of reuse and exploitation of annotated textual corpora

Annamaria De Santis<sup>1</sup>, Matteo Gallo<sup>2</sup>, Irene Rossi<sup>3</sup>, Jérémie Schiettecatte<sup>4</sup>

<sup>1</sup>Independent researcher – annamaria.desantis [at] unipi.it

<sup>2</sup>Senior developer – matteogal [at] gmail.com

<sup>3</sup>SPC-CNR – irene.rossi [at] cnr.it

<sup>4</sup>CNRS – UMR8167 Orient & Méditerranée – jeremie.schiettecatte [at] cnrs.fr

## ABSTRACT

This paper aims at presenting the results of the early adhesion to the principles afterwards codified as Open Science and FAIR principles in the frame of digital epigraphic projects in a niche area of research such as the pre-Islamic Arabian studies, to show how annotated corpora, provided that they adopt international standards and best practice, and expose data in an open format, have many more chances to be easily exploited and reused for different objectives than traditional, analogue corpora. The case study analysed in this paper is the Digital Archive for the Study of pre-Islamic Arabian inscriptions – DASI, an online annotated corpus of the textual sources from Ancient Arabia, which also exposes its records in standard formats (oai\_dc, EpiDoc, EDM) in an OAI-PMH repository. The initiatives of reuse of DASI open data in the frame of the recently ANR-funded project MAPARABIA (CNRS-CNR) are discussed in the paper, focusing on the exploitation of DASI's onomastic and geographic data in a new reference tool, the Gazetteer of Ancient Arabia.

## KEYWORDS

ancient Arabia; annotated textual corpora; digital epigraphy; digital gazetteers; digital reference tools; open data; reuse and exploitation of textual content; semantic web.

## TALK

### 1 INTRODUCTION

Building large, annotated textual corpora is hard-working, time-consuming and expensive. Moreover, work and skills gained by researchers committed to this activity are still scarcely acknowledged by the academic community. However, annotated corpora provided that they adopt international standards and best practice, and expose data in an open format, have many more chances to be easily exploited and reused for different objectives than traditional, analogue corpora.

This paper illustrates this thesis by focusing on the potential of the open-access archive of pre-Islamic Arabian inscriptions DASI, which is the main source of data for a different and further reference tool, the MAPARABIA digital Gazetteer of Ancient Arabia.

## **2 DASI ARCHIVE: THE SOURCE OF ONOMASTICS AND GEOGRAPHIC DATA**

Over the past years, considerable advancements have been made in the research on Ancient Arabia, leading to the production of a mass of archaeological and textual data. The epigraphic databases have played an outstanding role in disseminating this knowledge through the web publication of tens of thousands of annotated inscriptions and graffiti, spanning 1,500 years of history and covering a region that had long remained at the margins of the Near Eastern studies.

DASI - Digital Archive for the Study of pre-Islamic Arabian Inscriptions<sup>1</sup> is an online archive publishing at present the curated edition of nearly 8,500 inscriptions from Ancient Arabia<sup>2</sup>. The information encoded in DASI primarily regards the epigraphic texts. Annotation of the textual phenomena is made according to the TEI-EpiDoc schema [3] in an XML encoding module embedded into the database. The annotation concerns transcriptional, philological, and linguistic information – with a focus on onomastics. A general index of the names by specific onomastic type is automatically generated in DASI, also allowing to retrieve all their occurrences. It currently indexes more than 7,000 name-forms, comprising: personal names; divine names; names of months and decades; names of objects; names of buildings; names of social, political, and geographical entities. The latter two categories presently include about 3,300 names.

DASI epigraphic records are also provided with a series of textual and contextual metadata, such as information on script and language, chronology, text genre, type of support and iconographic elements, archaeological and geographical context. The sites that are places of provenance (production or discovery) of the epigraphs are catalogued in specific records. More than 400 sites are indexed in DASI, and provide information about: ancient and modern toponymy; location (country, geographic area and present governorate, coordinates and related accuracy); types of the findings, architectural structures and monuments; history and chronology; history of research; kingdoms, languages, deities and tribes attested at that site; general description; bibliography. Each site record may be linked to the other ones, thus representing the spatial relations between them.

---

1 DASI corpus is consultable through indexes and search tools [3]. DASI is the output of a five-year research project funded by the ERC from 2011 to 2016 (GA 269774; PI: Alessandra Avanzini, University of Pisa). It is currently maintained at the Consiglio Nazionale delle Ricerche.

<http://dasi.cnr.it/>

2 DASI corpus is for the most part made of the epigraphic heritage of the Ancient South Arabian civilization, which flourished since the early 1st millennium BCE until the advent of Islam, in the region corresponding to modern Yemen and neighbouring areas (the classical Arabia Felix). The Ancient South Arabian textual heritage is composed of over 10,000 inscriptions and graffiti, and hundreds of textual sources on perishable material. See [4] and [1].

### 3 THE MAPARABIA PROJECT AND THE GAZETTER OF ANCIENT ARABIA

The opportunity to exploit and enhance this wealth of geographic data and of onomastics having relation to a territory has occurred with the MAPARABIA project<sup>3</sup>. The objective of MAPARABIA is creating several tools to integrate and perform analysis, with the aim of finally producing synthesis, on the considerable amount of datasets collected in the last 50 years of intensive research on the pre-Islamic history of the Arabian Peninsula. Besides producing a web-GIS, a historical atlas and an online thematic dictionary of Ancient Arabia, MAPARABIA envisages the creation of a further reference tool, the Gazetteer of Ancient Arabia<sup>4</sup>.

This consists of a list of places, providing their identification, description, and semantic relationships among them. The MAPARABIA Gazetteer adopts the definition of “place” disseminated by the project Pleiades<sup>5</sup>, therefore it takes into consideration elements of the natural and anthropic landscape, entire settlements and individual artefacts, political, social and cultural entities related to the territory, “whether or not exactly locatable, whether or not their actual relation with the real world can be ascertained”.

The Gazetteer is designed to build upon several archaeological and geographical databases and textual archives, which have joined the project MAPARABIA or expose their data under open licenses. The aim is to organize and cross the information they include and stimulate study and reflection on fundamental research topics of Ancient Arabia that concern territorial dynamics, such as the settlement process and the man-environment relationship, the socio-political organization, the pre-Islamic linguistic and writing landscape, the evolution of pre-Islamic religion and the origins of Islam.

Indeed, the Gazetteer has been conceived to provide a complementary, semantic approach to the GIS, as it better points out the information about past geography provided by ancient texts, which is in the form of names, and allows to express cultural phenomena, such as political and administrative entities, which are not easily represented in their physical extension, and their numerous changes over time [6]. Moreover, as gazetteers enhance the name-based search of spatial information and the spatially-oriented search of textual information on the web, which has a semantic organization, it is expected to support the description, discovery, understanding, and process of data about Ancient Arabia on the web [5].

---

3 MAPARABIA is a four-year research project funded by the French National Research Agency (project ANR-18-CE27-0015). It is coordinated by Jérémie Schiettecatte (CNRS-UMR8167 - Orient & Méditerranée), with the partnership of Mounir Arbach (CNRS-UMR5133 - Archéorient), Irene Rossi (CNR-ISPC) and formerly Alessandra Avanzini (University of Pisa). It involves a large number of researchers belonging to both French and Italian institutions.

<http://www.orient-mediterranee.com/spip.php?article4002>

4 The Gazetteer of Ancient Arabia is not yet openly available. It will be made public from 2022 onwards at

<https://ancientarabia.huma-num.fr/gazetteer/>

Meanwhile, access to the database is allowed to authorized users at

<http://ancientarabia.cnr.it/gazetteer>

5 See the technical documentation available on the Pleiades website

<https://pleiades.stoa.org/help/concepts>



## 4 CONCEPTUAL MODEL AND SYSTEM ARCHITECTURE OF THE GAZETTER

The main entity of the Gazetteer of Ancient Arabia's conceptual model is the Place. Each Place record identifies univocally and persistently an ancient place and is related at least to one Location, that is its geographical expression, or one Name, that is its onomastic occurrence in an ancient written source. The relation with a Location or alternately to a Name is the condition of existence of a Place. A Place may have a physical or cultural-historical relation with another Place, that can be also chronologically qualified. Location may be provided with Bibliography; similarly, Name inherits a link with at least a Source which witnesses its existence (Fig. 1).

In this first phase of the project MAPARABIA, the Gazetteer is synchronized only with the system DASI. The web application consists of: a module importing data from the DASI web APIs, a database, a data entry and a module exporting data.

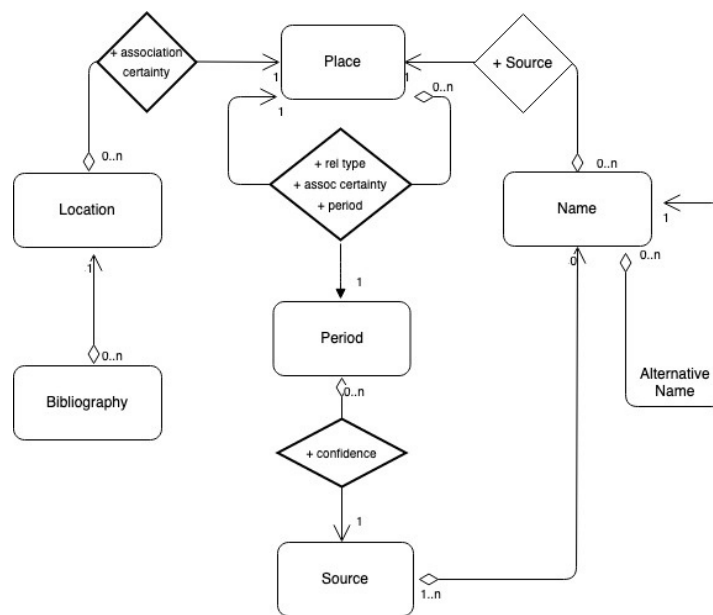


Figure 1: conceptual model of the MAPARABIA Gazetteer

The population of the Gazetteer, indeed, is automatically performed by importing on demand data from DASI, in particular from:

- site records (archaeological sites, which are provenance or place of production of inscriptions, and related monuments), implementing the Location records of the Gazetteer,
- word List, limited to onomastics (toponyms, nisbas, tribe names, and names of (sacred) buildings attested by the inscriptions encoded), implementing Name records,
- epigraph records, implementing Source records,
- and vocabularies, above all the Periods.

Each couple onomastic item + EpiDoc compliant tag of the DASI Word List is used to create one instance of the entity Name of the Gazetteer. It inherits also the relation with the DASI epigraphic source it is attested in, just as Source inherits the Period it is dated to in DASI.

Whereas Location records cannot be created or modified, Place, Name, Source and Period can be also entered manually in the Gazetteer and, when already existing, implemented with new metadata.

| DASI EpiDoc compliant tag                | name type          |
|--|--------------------|
| <placeName></placeName>                  | toponym            |
| <placeName type="sanctuary"></placeName> | religious building |
| <placeName type="building"></placeName>  | other building     |
| <rs type="nisba"></rs>                   | nisba              |
| <orgName type="tribe"></orgName>         | tribe              |

Table 1: values of the field “name type” populated from DASI

The example of as-Sawdā’ / Nashshān illustrates the conceptual model and the functioning of the Gazetteer. as-Sawdā’, as an archaeological site, is described by a Site record in DASI<sup>6</sup>. Data imported from this record (country, coordinates, coordinates accuracy, type of site, structures, location and toponymy, history of research, general description and chronology) and from the linked Bibliography records of DASI, populate the Location record of the Gazetteer, that cannot be manipulated.

This archaeological site corresponds to the ancient city of Nashshān, mentioned as *ns<sup>2</sup>n* <toponym> in 2 Minaic inscriptions and 27 Sabaic inscriptions encoded in DASI (Fig. 2). The Gazetteer is then provided with a Name record, related to the Source records that correspond to the Minaic and Sabaic inscriptions of DASI, which can be enriched with comments and information regarding language, accuracy and completeness of transcription, and links with alternative appellations, such as in different languages, different transliterations and so on. For instance, a new Name record can be created, dealing with the Latin toponym Nestum, which will be linked to *ns<sup>2</sup>n*, as Alternative Name, and to a new Source record dealing with Pliny the Elder’s *Naturalis Historia* (VI, 32, 160) (Fig. 3).

---

6 <http://dasi.cnr.it/sit-39>

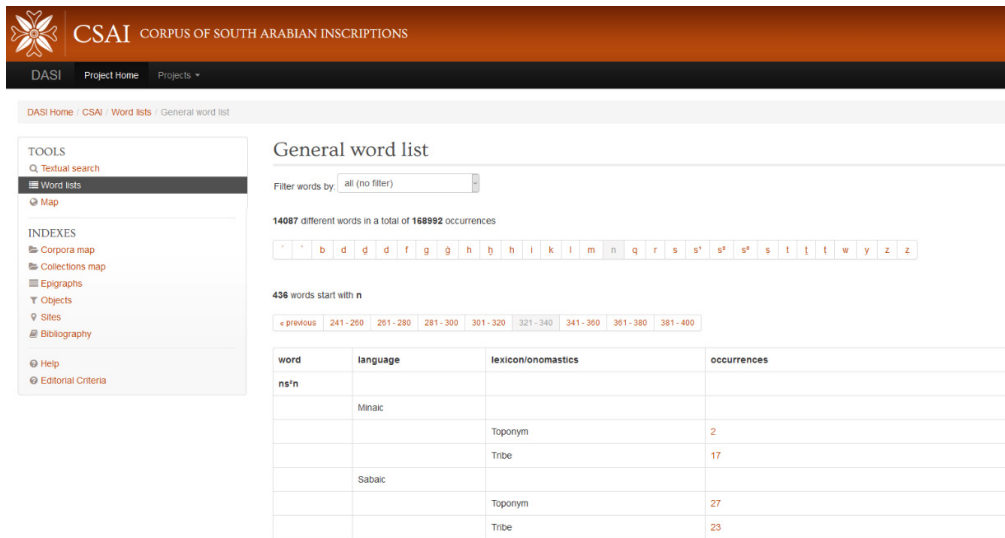


Figure 2: onomastic item  $ns^2n$  in the Word List of DASI

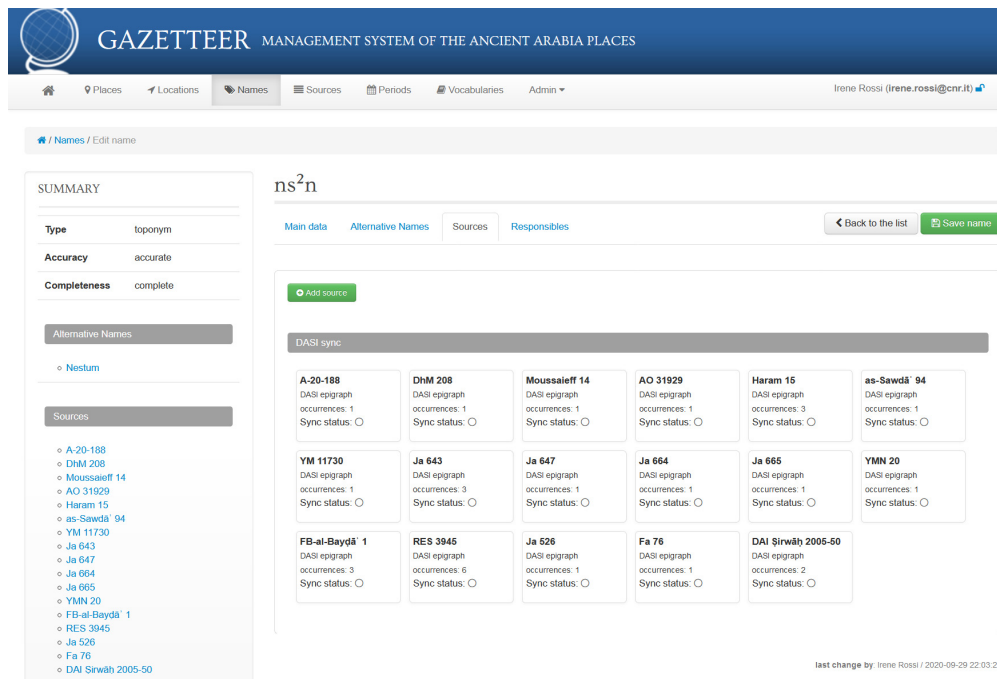


Figure 3:  $ns^2n$  Name record in the Gazetteer of Ancient Arabia, with related Source records

Both the Location and the Name records are connected to the Place record as-Sawdā' / Nashshān (archaeological site / ancient settlement). This can be implemented with a proper description and, in turn, can be related to the monuments located in its territory and to the near city of al-Bayḍā' / Nashq (Fig. 4).

Since  $ns^2n$  is also the appellation of a tribe (<orgName type="tribe"></orgName>; Fig. 2), further Name and Place records are created for this tribe; the Place record can be linked firstly to as-Sawdā' / Nashshān (archaeological site / ancient settlement), and to further ancient settlements it was connected with.

## 5 CONCLUSIONS

In conclusion, if the historical and cultural domain of the MAPARABIA Gazetteer – i.e. Ancient Arabia – is its main peculiarity, its added value, compared to the other gazetteers of the ancient world, is the direct bond with the annotated epigraphic corpus of DASI. The core data of the Gazetteer result automatically from the mass digitization of the direct written heritage of pre-Islamic Arabia conducted during the project DASI, according to guidelines that have established themselves as proper “standards” in the digital epigraphy field, and applying the best practices that were later formalized under the label of FAIR principles [7]. DASI records, provided with URIs, are exposed in standard formats (oai\_dc, EpiDoc, EDM) in an OAI-PMH repository, thus allowing different projects to access and use its data.

The screenshot displays the 'GAZETTEER MANAGEMENT SYSTEM OF THE ANCIENT ARABIA PLACES' interface. The top navigation bar includes a home icon, 'Places', 'Locations', 'Names', 'Sources', 'Periods', 'Vocabularies', and 'Admin'. The user 'Irene Rossi (irene.rossi@cnr.it)' is logged in. The main content area shows the 'Edit place' page for 'as-Sawdā' / Nashshān. On the left, a 'SUMMARY' sidebar lists the 'Type' as 'archaeological site settlement <ancient>', a 'Description' of 'The archaeological site of as-Sawdā' in north-west...', and lists 'Locations', 'Related Names' (including 'nsʾn (toponym)', 'nsʾnyfn (nisbe)', 'ʾsʾsʾn (nisbe)', and 'Nestum (toponym)'), and 'Related Places' (including 'as-Sawdā' - Aranyada' Temple (intra-muros) (certain) - Period: "Pre-Islamic kingdoms"', 'as-Sawdā' - City walls (certain) - Period: "Pre-Islamic kingdoms"', 'as-Sawdā' - Alhtar dhū-Risafum Temple (extra-muros) (certain) - Period: "Pre-Islamic kingdoms"', 'has attestation of Nsʾn (certain) - Period: "Pre-Islamic kingdoms"', and 'near al-Baydā' (certain) - Period: "Pre-Islamic kingdoms"'). The main form on the right includes tabs for 'Main data', 'Locations', 'Related Names', 'Related Places', 'Responsibles', and 'Rights'. It features a 'status' dropdown set to 'draft', a 'Type' dropdown set to 'archaeological site', a 'Title' field containing 'as-Sawdā' / Nashshān', and a 'Description' text area with a detailed paragraph about the site's history. Below the description are fields for 'Period' (set to '"Pre-Islamic kingdoms"') and 'Provider' (set to 'MAPARABIA gazetteer of pre-Islamic Arabian places'). Action buttons for 'Back to the list', 'Save place', and 'Delete place' are visible at the top right of the form.

Figure 4: place record as-Sawdā' / Nashshān in the Gazetteer of Ancient Arabia

Imported data can be enriched in the Gazetteer with additional details arising from the thorough study of the sources and the effort to systematize identification and relations. The instances of the main entity Place, in fact, are the only ones that must be created from scratch, requiring the editorial intervention to disambiguate, identify and circumscribe an ancient “place”. As well as the creation of the relation Place to Place, this is the step of the workflow the scientific reflection focuses on, and the editorial responsibility is more significant.

The export module of the Gazetteer allows to expose Place records, being the other entities nested within, in JSON-LD format. Each Place item is identified with a URI and is released under open license. This is consistent with the philosophy that has allowed the Gazetteer itself to be created and is expected to increase dissemination through aggregation and linking with further gazetteers, and therefore with archaeological, textual and geographic data, pertaining to different chronological and cultural contexts.

## REFERENCES

1. Avanzini, Alessandra. *By Land and by Sea. A History of South Arabia before Islam Recounted from Inscriptions*. Vol. 10. L'Erma di Bretschneider, 2016.
2. Avanzini, Alessandra, Annamaria De Santis, and Irene Rossi. '1 Encoding, Interoperability, Lexicography: Digital Epigraphy Through the Lens of DASI Experience', 2018.
3. Elliott, Tom, Gabriel Bodard, Elli Mylonas, Simona Stoyanova, Charlotte Tupman, and Scott Vanderbilt. 'EpiDoc Guidelines: Epigraphic Documents in TEI XML (Version 8)', 2017 2007. <http://www.stoa.org/epidoc/gl/latest/>.
4. Schiettecatte, Jérémie. *D'Aden à Zafar. Villes de l'Arabie Du Sud Préislamique*, 2011.
5. Shaw, Ryan. 'Gazetteers Enriched: A Conceptual Basis for Linking Gazetteers with Other Kinds of Information'. In *Placing Names: Enriching and Integrating Gazetteers*, edited by Merrick Lex Berman, Ruth Mostern, and Humphrey Southall. Indiana University Press, 2016. <https://doi.org/10.2307/j.ctt2005zq7>.
6. Southall, Humphrey, Ruth Mostern, and Merrick Lex Berman. 'On Historical Gazetteers'. *International Journal of Humanities and Arts Computing* 5, no. 2 (2011): 127–45. <https://doi.org/10.3366/ijhac.2011.0028>.
7. Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. 'The FAIR Guiding Principles for Scientific Data Management and Stewardship'. *Scientific Data* 3, no. 1 (2016): 160018. <https://doi.org/10.1038/sdata.2016.18>.

# **Sessione VII**

## **Maria Montessori**



# **Fare comprendere ai ragazzi il valore della propria privacy: esiti di una sperimentazione a scuola**

Fulvio Oscar Benussi  
MIUR – fulvio.benussi [at] gmail.com

## **SINTESI**

Presentazione di un percorso didattico sperimentato in un liceo milanese. Con l'obiettivo di contrastare l'abitudine oggi diffusa tra i giovani di tendere a considerare la privacy un privilegio anacronistico cui rinunciare senza rimpianti abbiamo costruito una "contro-narrazione" facendo lavorare gli studenti su simulazioni e casi concreti per "spiazzarli" rispetto alle loro convinzioni ingenue e portarli a modificare il senso delle credenze che avevano, fino a quel momento, accettato acriticamente. L'obiettivo perseguito è stato in parte raggiunto.

## **PAROLE CHIAVE**

privacy, atteggiamenti, consapevolezza, case history, simulazione, interdisciplinarietà

## **INTERVENTO**

### **1 INTRODUZIONE**

“La narrazione prevalente tende a considerare la privacy un privilegio quasi anacronistico nell'era della massima compenetrazione tra uomo e tecnologia” [1]. Partendo dal presupposto che tale assunto non può essere confutato con una narrazione scolastica fondata sulla consueta spiegazione frontale abbiamo autonomamente ideato e sperimentato la seguente proposta didattica.

### **2 LO SVILUPPO DEL PROGETTO – FASE PROPEDEUTICA**

Il percorso didattico, svolto durante le ore curricolari delle lezioni di diritto ed economia era rivolto a tre classi composte da 72 studentesse e 4 studenti del primo biennio di un liceo milanese. Con un veloce sondaggio abbiamo innanzitutto appurato che la maggioranza degli studenti aveva una spiccata preferenza per un particolare social network: Instagram. Abbiamo poi proposto agli studenti un'esercitazione di economia relativa a vari case history di prodotti lanciati sul mercato da imprese nazionali e internazionali.



L'esercitazione, collegata a un test di verifica, aveva lo scopo di evidenziare come alcuni prodotti avevano avuto successo mentre il lancio di altri si era risolto in un insuccesso. In classe abbiamo poi svolto una riflessione socializzata su quali strategie potessero favorire/ostacolare il successo di un prodotto lanciato sul mercato. Segnaliamo che nell'esercitazione abbiamo considerato i case history: Postal market (rivista); fashion on line; yogurt Muller; marmellata Santa rosa in confezione "squezzabile"; Google glass; Bus cinese che, grazie a grosse ruote e un interasse elevato, poteva viaggiare sopra due file di automobili; Apparecchio per la diagnostica computerizzata dei guasti delle automobili; Dropbox; caffettiera Bialetti; Televisione 3D; MMS; Renault Twingo; PayPal; WhatsApp; Iphone.

Gli obiettivi di apprendimento erano relativi all'analisi economica delle strategie seguite dalle aziende valutando situazioni reali di mercato cui gli studenti si accostavano tramite documenti multimediali reperiti in Internet o forniti dal docente<sup>1</sup>. Un secondo obiettivo, legato anche al tema della privacy, era quello di preparare gli studenti ad affrontare la successiva esercitazione in cui sarebbe stato chiesto loro di esprimere un parere sulle potenzialità economiche di nuovo social network pronto al lancio sul mercato. Il secondo obiettivo era quindi anche quello di evitare una focalizzazione eccessiva dell'attenzione degli studenti al tema della privacy presente nella successiva esercitazione favorendo così il realizzarsi dello spiazamento cognitivo di cui trattiamo nel successivo paragrafo 5.

### 3 LA SIMULAZIONE

Prima di procedere con la seconda esercitazione abbiamo predisposto una finta home page di un ipotetico nuovo social network da sottoporre agli studenti. Il social network, che abbiamo chiamato "Comics & games social", sarebbe apparso su Internet con una home page ottenuta elaborando con un editor di immagini una immagine che risultava, graficamente, abbastanza simile alla home page del social network preferito dagli studenti.

Successivamente abbiamo predisposto un test digitale con Google Forms da utilizzare nella fase di simulazione. Con il test gli studenti, in veste di consulenti della startup, erano chiamati ad esprimere un parere sulle possibilità di successo/insuccesso sul mercato del nuovo social. In particolare il test alterna domande relative alla valutazione che ogni studente dava delle condizioni necessarie affinché il nuovo social potesse ottenere un buon gradimento in Rete e conseguentemente un buon ritorno economico, a domande dove dovevano esprimere il loro parere sulla raccolta dei dati dell'utente relativamente alla tutela o violazione della sua privacy. "Ecco perché appare sempre più urgente promuovere e portare avanti lo sviluppo di quelle capacità tecniche e sociali, social networking skills, che possano stimolare il mondo dell'educativo nella direzione della formazione e condivisione di buone pratiche educative, tra queste tutelare la propria privacy imparando a muoversi in una dimensione etica che contempli anche, e soprattutto, il rispetto dell'altro" [2].

---

1 I materiali multimediali utili per essere proporre case history sono stati raccolti negli anni dal docente acquisendoli dalla televisione e da Internet e memorizzandoli su capienti supporti digitali in vista di possibili successivi usi didattici

## 4 LO SVILUPPO DEL PROGETTO – FASE DELL’ANALISI SOCIALIZZATA DEGLI ESITI

Finita la fase della simulazione, dopo una breve pausa, abbiamo analizzato gli esiti del questionario mediante la visualizzazione nella LIM del laboratorio delle infografiche generate da google forms e relative a come gli studenti della classe si erano orientati nel rispondere alle domande proposte. Ci siamo soffermati a commentare ogni singolo grafico e in particolare quelli che, rispetto a quanto la startup proponeva come regole del social network nel trattamento della privacy degli utenti, evidenziavano da parte della classe un forte dissenso; la risposta, che se barrata, evidenziava il dissenso era: “con una regola di questo tipo mi sentirei spiato, perciò non la accetterei”. Nei grafici le risposte che indicano un forte dissenso degli studenti nei confronti delle norme sulla privacy sono indicate in verde scuro (figura 1).

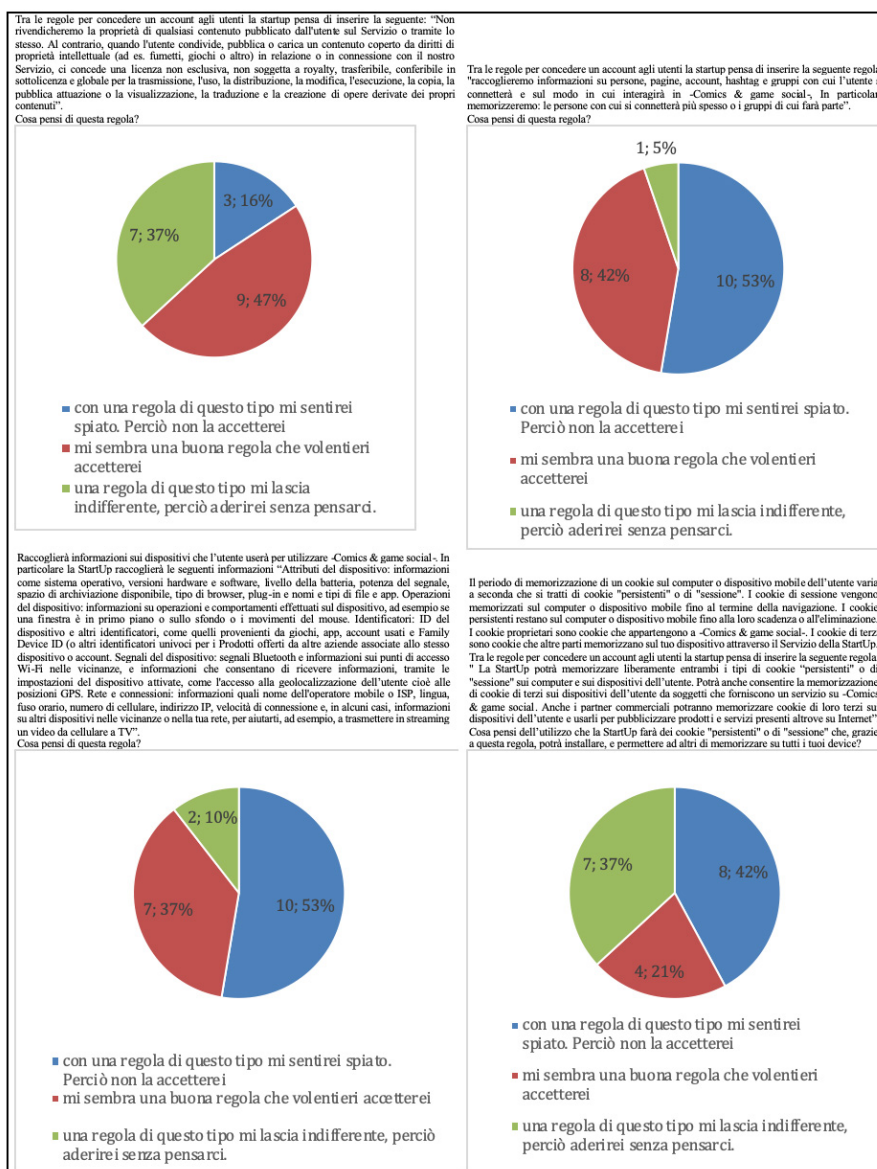


Figura 1

## **5 LO SVILUPPO DEL PROGETTO – FASE DEL DIBATTITO SOCIALIZZATO**

La maggior parte delle norme sulla privacy definite dalla start-up risultavano non gradite da un ampio numero di studenti e quindi gli stessi collegavano la possibilità di successo del nuovo social network proposto dalla startup ad un'accurata revisione delle norme sulla privacy.

A questo punto il docente ha svelato come aveva proceduto nel realizzare l'esercitazione. Ha esplicitato, in particolare, che le norme relative alla privacy erano state copiate dal social network maggiormente seguito dagli studenti e poi incollate per realizzare il test da proporre alla classe. Ovviamente nell'operazione di "copia-incolla" il docente aveva provveduto a togliere le parole e i riferimenti che avrebbero permesso agli studenti di riconoscere la fonte delle norme sulla privacy che venivano sottoposte al loro giudizio nel test.

Quest'ultima operazione ha prodotto un *déplacement* o spiazzamento cognitivo nei ragazzi. "Lo spiazzamento cognitivo è un passaggio obbligato per la messa in atto e lo studio dei processi cognitivi: attraverso questo è possibile disinnescare gli automatismi del proprio pensiero. Ha un effetto destabilizzante dal punto di vista cognitivo poiché rende evidente la trama dei presupposti che regolano in modo implicito la nostra vita e i nostri saperi consolidati" [1]. Infatti la maggioranza degli studenti hanno subito espresso fastidio per la sensazione di essere stati imbrogliati, altri invece si sono dichiarati non colpiti in quanto erano consci che la loro privacy è scarsamente tutelata quando frequentano i social network. Questi ultimi hanno anche dichiarato che la perdita della privacy nei social network era compensata dalle opportunità di relazione che i social stessi offrivano, inoltre hanno segnalato la loro difficoltà a rapportarsi con un documento la cui lunghezza e difficoltà non consentiva loro di comprendere appieno il contenuto di quanto riportato nell'accordo sulla privacy. Alcuni studenti hanno anche dichiarato che, avendo compreso il problema, avrebbero ridotto nei propri account, le autorizzazioni circa quanto erano disposti a cedere della propria privacy nei social.

## **6 CONCLUSIONI**

La tendenza diffusa a sottovalutare l'importanza della tutela della privacy personale da parte dei ragazzi è stata tematizzata a partire da un'esercitazione predisposta ad hoc.

L'esercitazione ha favorito l'assunzione di responsabilità degli studenti nel valutare le loro convinzioni e riconsiderare il senso delle credenze che avevano, fino a quel momento, accettato acriticamente. Alcuni studenti hanno anche modificato atteggiamento sulla tutela della propria privacy in ambiente digitale.

L'esercitazione che prevedeva che lo studente assumesse il ruolo di valutatore economico della possibilità di successo di un'iniziativa imprenditoriale ha permesso di approcciare il problema della privacy in termini di opportunità: quali caratteristiche e quali norme sulla privacy avrebbero permesso al nuovo social della startup di avere successo?

In questo modo si è evitato che i ragazzi vivessero la proposta didattica come invasiva del territorio dove liberamente si relazionano con i coetanei e costruiscono la loro identità.

Lo spiazzamento cognitivo avvenuto nella fase di debriefing, ha consentito di "aprire una breccia" nelle convinzioni assunte acriticamente dagli studenti portandoli a una riflessione più ponderata e in qualche caso a una modifica di atteggiamento nei confronti dell'importanza di tutelare la propria privacy.

## BIBLIOGRAFIA

1. Premoli, Silvio. *Verso l'autonomia. Percorsi di sostegno all'integrazione sociale di giovani*. Milano: Franco Angeli, 2009.
2. Ranieri, Maria, e Stefania Manca. *I social network nell'educazione. Basi teoriche, modelli applicativi e linee guida*. Trento: Erickson, 2013.
3. Redazione ANSA. «Soro, proteggere la vita digitale dei figli - Internet e Social». Agenzia ANSA, 23 ottobre 2019. [http://www.ansa.it/sito/notizie/tecnologia/internet\\_social/2019/10/23/soro-protectere-vita-digitale-dei-figli\\_a7f90690-422d-4701-8cf7-f9e0318991b6.html](http://www.ansa.it/sito/notizie/tecnologia/internet_social/2019/10/23/soro-protectere-vita-digitale-dei-figli_a7f90690-422d-4701-8cf7-f9e0318991b6.html).



# Un serious game per promuovere la cultura della salute e sicurezza nei giovani lavoratori

Rosina Bentivenga<sup>1</sup>, Sara Stabile<sup>2</sup>, Emma Pietrafesa<sup>3</sup>

1 Inail – r.bentivenga [at] inail.it

2 Inail – sa.stabile [at] inail.it

3 Inail – e.pietrafesa [at] inail.it

## SINTESI

Le tecnologie digitali permettono di progettare percorsi formativi attraverso formati di facile accesso e personalizzati e di monitorare il raggiungimento degli obiettivi anche attraverso strumenti di autovalutazione, l'apprendimento diventa così più efficace, multimodale, partecipato e collaborativo.

Nell'ambito dei contesti di apprendimento risulta, pertanto, importante incrementare l'uso delle tecnologie digitali per migliorare i processi di apprendimento e per sviluppare nuove competenze. La gamification, può dare un valido contributo in tal senso in quanto, attraverso il divertimento, aumenta la motivazione, il coinvolgimento e la collaborazione tra gli studenti.

In base alla normativa vigente in materia di salute e sicurezza sul lavoro i ragazzi che sono inseriti nei percorsi per le competenze trasversali e per l'orientamento vengono equiparati ai lavoratori e quindi devono ricevere un'adeguata formazione e informazione, finalizzata a prevenire infortuni e malattie professionali, che possono essere supportate dall'utilizzo di strumenti digitali e nuove modalità comunicative, come ad esempio i videogiochi.

## PAROLE CHIAVE

tecnologie digitali, serious game, salute e sicurezza sul lavoro, scuola, apprendimento

# INTERVENTO

## 1 INTRODUZIONE

L'European Agency for Safety and Health at Work (EU-OSHA) considera i giovani lavoratori tra i 18 e i 24 anni una categoria vulnerabile in quanto sono più esposti a rischio per una serie di fattori tra i quali la scarsa percezione del rischio e dei comportamenti pericolosi, la scarsa esperienza e la facilità ad accettare mansioni pericolose, la mancanza di informazione e formazione e la scarsa conoscenza dei propri diritti in materia di salute e sicurezza sul lavoro (SSL)<sup>1</sup>.

È molto importante, pertanto, garantire ai giovani un lavoro sano, sicuro e adeguato alle loro capacità e questo è attuabile oltre che attraverso la predisposizione di ambienti di lavoro salubri e sicuri, anche mediante l'erogazione di un'adeguata informazione e formazione al momento dell'assunzione e una continua supervisione. Per recepire la cultura della sicurezza sul luogo di lavoro in cui i ragazzi operano e per dare loro la possibilità di contribuire di fatto alla gestione della SSL è necessario che siano attivamente coinvolti, supportati e incoraggiati.

La normativa in materia di formazione sulla SSL, anche nell'ottica del life long learning, si sta orientando verso l'acquisizione da parte dei giovani lavoratori delle conoscenze di base in materia di prevenzione dei rischi a scuola e nelle università, ancor prima di entrare a fare parte dei contesti lavorativi.

Alcuni percorsi digitali possono essere utilizzati per aumentare la consapevolezza dei lavoratori in materia di salute e sicurezza, in quanto attraverso l'attivazione dell'attenzione e il coinvolgimento emotivo, favoriscono l'acquisizione di informazioni utili al lavoratore per pensare e agire in termini di sicurezza.

A tale proposito è stato realizzato un progetto di ricerca finalizzato alla costruzione di un serious game per accrescere la consapevolezza dei rischi e del valore della sicurezza negli ambienti di lavoro degli studenti coinvolti nei percorsi per le competenze trasversali e l'orientamento (PCTO).

I serious game, *giochi con un esplicito e ben definito scopo educativo, non pensati primariamente per il divertimento, senza però escluderlo* [1], negli ultimi anni hanno assunto il ruolo di strumenti innovativi in grado di promuovere la crescita personale e il benessere psicologico e potrebbero essere introdotti come tecnologie positive<sup>2</sup>. Tali giochi possono avere, infatti, un ruolo rilevante sia per le esperienze individuali, sia per le relazioni interpersonali in quanto alimentano emozioni positive, promuovono l'impegno e migliorano l'integrazione sociale e la connessione [3].

I serious game, dunque, non hanno esclusivamente o principalmente uno scopo di intrattenimento, ma contengono elementi educativi che integrano conoscenze dichiarative e procedurali, bilanciando la dimensione simulativa, ludica e formativa, che sono fondamentali per rendere più efficace l'apprendimento e il coinvolgimento [2].

---

1 <https://osha.europa.eu/en/themes/young-workers>

2 La tecnologia positiva si basa sulla ricerca teorica e applicata, e ha l'obiettivo di indagare come le tecnologie dell'informazione e della comunicazione possano essere utilizzate per potenziare la qualità dell'esperienza personale a livello di benessere edonico, eudaimonico e sociale.

Questi aspetti sono molto importanti nell'ambito dei PCTO, che hanno l'obiettivo di avvicinare gli studenti al mondo del lavoro aiutandoli a individuare i propri interessi professionali, attraverso l'integrazione della formazione d'aula con l'apprendimento esperienziale presso contesti pubblici o privati [6] e che sono centrati sull'auto-orientamento, che consente di adeguare le competenze alle richieste di un mercato del lavoro in rapida trasformazione.

Nell'ambito dei PCTO gli studenti acquisiscono lo status di lavoratori ai sensi dell'articolo 2, comma 1, lettera a), del D.Lgs. 81/2008<sup>3</sup> e sono soggetti agli adempimenti previsti in materia di tutela della SSL, come ad esempio la formazione e, se necessarie, la sorveglianza sanitaria e la dotazione di dispositivi di protezione individuale.

Il tipo di formazione che deve essere garantita agli studenti viene specificata dalla Carta dei diritti e dei doveri degli studenti impegnati nei percorsi, emanata con il Decreto 195 del 2017<sup>4</sup> e consiste in una formazione generale e una specifica in SSL, come stabilito dagli Accordi Stato-Regioni del 2011 e del 2016<sup>5,6</sup>. È importante, pertanto, progettare la formazione utilizzando metodologie che stimolino l'interesse dei giovani e li coinvolgano attivamente anche avvalendosi delle opportunità offerte dalle tecnologie dell'informazione e della comunicazione e dai nuovi strumenti e linguaggi digitali.

## 2 MATERIALI E METODI

La ricerca, che si è basata sulla partecipazione attiva finalizzata alla co-progettazione di un videogioco "Sicuri si diventa"<sup>7</sup>, ha coinvolto 7 istituti secondari di secondo grado, distribuiti in 3 Regioni, relativi a 3 settori lavorativi: agricoltura, costruzioni e manifatturiero. Hanno preso parte allo studio, 12 docenti e circa 300 studenti, delle classi III, IV e V degli Istituti-tecnici.

È stata realizzata un'indagine quali-quantitativa attraverso la somministrazione di questionari, interviste semi-strutturate individuali e discussioni di focus group sulle rappresentazioni dei rischi in generale sul luogo di lavoro e sulle esperienze di insegnamento e di apprendimento.

Ai docenti è stata somministrata un'intervista semi-strutturata e per i ragazzi sono stati realizzati 4 world café/focus group, che hanno portato alla creazione di gruppi di lavoro su piattaforma di messaggistica istantanea (WhatsApp). Questi gruppi di lavoro sono serviti a individuare un gruppo di volontari, composto da 12 studenti con il ruolo di "ambasciatori tra pari", che avevano il compito di intervistare i loro amici e altri conoscenti e di interagire con i ricercatori in merito alla progettazione del videogioco.

---

3 Decreto Legislativo 9 aprile 2008, n. 81. Attuazione dell'articolo 1 della legge 3 agosto 2007, n. 123, in materia di tutela della salute e della sicurezza nei luoghi di lavoro.

4 Decreto 3 novembre 2017, n. 195. Regolamento recante la Carta dei diritti e dei doveri degli studenti in alternanza scuola-lavoro e le modalità di applicazione della normativa per la tutela della salute e della sicurezza nei luoghi di lavoro agli studenti in regime di alternanza scuola-lavoro.

5 Accordo finalizzato alla individuazione della durata e dei contenuti minimi dei percorsi formativi per i responsabili e gli addetti dei servizi di prevenzione e protezione, ai sensi dell'articolo 32 del decreto legislativo 9 aprile 2008, n. 81 e successive modificazioni (GU Serie Generale n.193 del 19-08-2016).

6 Accordo tra il Ministro del lavoro e delle politiche sociali, il Ministro della salute, le Regioni e le Province autonome di Trento e Bolzano per la formazione dei lavoratori, ai sensi dell'articolo 37, comma 2, del decreto legislativo 9 aprile 2008, n. 81 (GU Serie Generale n.8 del 11-01-2012).

7 <https://site.unibo.it/sicuri-si-diventa/it>



Sulla base delle analisi, dei principali risultati e dal confronto con gli esperti di SSL, dei ricercatori e dei docenti, è stata sviluppata una versione preliminare del videogioco sulla salute e sicurezza sul lavoro, per la quale è stato effettuato un beta test. La versione completa del gioco è stata utilizzata per realizzare una sperimentazione pilota e un sondaggio relativo al livello di soddisfazione/utilità del videogioco con il coinvolgimento di 75 ragazzi suddivisi in squadre. La versione finale, adattata in base ai risultati qualitativi e quantitativi, è stata presentata nell'ambito di un contest, al quale hanno preso parte 39 studenti, anch'essi suddivisi in squadre.

### **3 RISULTATI**

Hanno partecipato alla ricerca 12 docenti e circa 300 studenti di 7 Istituti tecnici (3,6% agrario, 28,8% manifatturiero, 30,2% edile). Hanno preso parte all'intervista semi-strutturata 10 docenti del campione totale e 50 ragazzi che hanno partecipato a 4 world café/focus group, sono stati realizzati 3 gruppi WhatsApp che hanno visto il coinvolgimento di 12 studenti, infine è stato somministrato un questionario anonimo, realizzato ad hoc, a 277 studenti (63% ragazzi e 37% ragazze) degli ultimi tre anni degli Istituti coinvolti (terza 22,7%, quarta 44,1%, quinta 33,2). Il 92,5% dei ragazzi riferisce di avere già partecipato ai PCTO, ma in tale ambito solo il 27,1% ha seguito corsi di formazione sulla SSL. Tale formazione è ritenuta certamente utile dal 62,8% dei ragazzi, mentre il rispetto della normativa garantisce di lavorare in sicurezza per il 69,7% del campione. Per circa il 50% degli intervistati la formazione più efficace è quella sul campo, mentre solo per il 5% lo è quella realizzata attraverso lezioni frontali con la proiezione di slide.

Per quanto riguarda le rappresentazioni sociali del rischio, i ragazzi lo associano principalmente a pericolo (67,5%), danno (39%), farsi male (30,3%), imprevisto (28,9%) e disattenzione (27,4%). Secondo gli intervistati, le persone sul lavoro rischiano prevalentemente per disattenzione (61,3%), per fretta (56,8%), per comodità (36,5%) e per proprie abitudini (31,1%). Il 90% dei ragazzi ha gradito il videogioco nel suo complesso, il 55% lo ritiene molto utile per imparare divertendosi, il 43% per conoscere le regole della sicurezza sul lavoro, il 42% per diventare più consapevoli dei rischi sul lavoro e, infine, il 26% per ricordare le regole della sicurezza sul lavoro.

### **4 CONCLUSIONI**

Imparare giocando è sempre stata un'attività che ha permesso lo sviluppo di abilità personali e professionali, oggi le tecnologie e il gioco possono offrire elementi cognitivi e operativi adatti riconoscere e quindi prevenire i rischi professionali [4]. Questo studio conferma che gli elementi educativi possono essere integrati in un videogioco, per fare sì che vengano acquisiti dai giocatori durante il gioco stesso. Le modalità formative innovative, essendo più vicine ai linguaggi e ai processi cognitivi ed emotivi dei giovani assumono dunque un ruolo rilevante nel processo dell'apprendimento. Nello specifico, il videogioco creato propone una sfida in cui si ottengono punteggi in funzione dell'abilità nel tutelare e mettere in sicurezza i lavoratori, in modo da evitare gli infortuni. Lo scopo è quello di far emergere come l'applicazione delle regole non comporti una perdita di tempo o di guadagno, ma anzi favorisca l'efficienza e la crescita in termini economici, di salute e di sicurezza.

Usando lo spazio virtuale e sperimentando i percorsi necessari per raggiungere i vari obiettivi, il giocatore apprende le regole mettendole in pratica in contesti che a mano a mano diventano più complessi. La salute e la sicurezza sul lavoro possono, pertanto, essere inserite nelle attività esperienziali dei discenti per diventare parte integrante del loro orientamento culturale che ne influenza l'azione [5].

## 5 RINGRAZIAMENTI

Questo lavoro è stato realizzato nell'ambito del Bando di Ricerca in Collaborazione (BRIC) 2016 ID 48, finanziato dall'Inail, in collaborazione con il Dipartimento di scienze politiche e sociali dell'Università "Alma Mater Studiorum" di Bologna - Responsabile scientifico Prof.ssa Pina Lalli - che ha condotto la ricerca con il contributo della AUSL di Modena, AUSL di Viterbo e ASUR Marche A.V. 4 - Fermo.

## BIBLIOGRAFIA

1. Abt, Clark C. *Serious games*. Lanham: University Press of America, 1987.
2. Anolli, Luigi, e Fabrizia Mantovani. *Come funziona la nostra mente. Apprendimento, simulazione e Serious Games*. Bologna: il Mulino, 2011.
3. Argenton, Luca, Esther Schek, e Fabrizia Mantovani. «Serious games as positive technologies». In *International Conference on Virtual, Augmented and Mixed Reality*, 169–177. London: Springer, 2014.
4. Lalli, P, Farina F.G., Bentivegna R. Stabile S., e Bernardini M. Pietrafesa E. «Sicuri si diventa: il gaming per la formazione su salute e sicurezza sul lavoro in alternanza scuola-lavoro». In *Convegno Internazionale SIRD, SIPES, SIREM, SiEMENS su Le Società per la società: ricerca, scenari, emergenze*, 2019.
5. Schudson, Michael. «How culture works». *Theory and Society* 18, n. 2 (1989): 153–180.
6. Stabile, Sara, Cinzia Milana, Rosina Bentivenga, e Emma Pietrafesa. «Alternanza Scuola Lavoro: sviluppo di nuove competenze e tutela della salute e sicurezza sul lavoro». *Formazione e insegnamento. Rivista internazionale di scienze dell'educazione e della formazione*. 17, n. 1 (2019): 259–272.



# Istruzione e GAFAM: dalla coscienza alla responsabilità

Paolo Monella  
Università di Palermo – paolo.monella [at] unipa.it

## SINTESI

Durante la pandemia COVID-19, la didattica a distanza d'emergenza in Italia, sia a scuola sia all'università, è stata realizzata quasi esclusivamente tramite infrastrutture e piattaforme proprietarie appartenenti alle grandi multinazionali informatiche, soprattutto *Google (G-Suite for Education)*, *Microsoft (Teams)* e *Facebook (Zoom e Whatsapp)*. Questo apre problemi di tutela dei dati degli studenti ampiamente sottovalutati nel dibattito pubblico. In altri paesi, tra cui ad esempio la Francia, il Ministero dell'istruzione ha messo a disposizione infrastrutture pubbliche e fondate su protocolli aperti. La comunità scientifica ha la responsabilità di elaborare e proporre un bilancio ragionato di quanto è avvenuto nel fuoco dell'emergenza per costruire alternative realizzabili nell'immediato del secondo anno di pandemia, e per evitare che pratiche deteriori diventino sistematiche nella formazione scolastica e universitaria del futuro. D'altra parte, è auspicabile che le istituzioni formative (MIUR,USR,Atenei e loro consorzi) si assumano la responsabilità di costruire infrastrutture per la didattica digitale che siano pubbliche, condivise e fondate su tecnologie aperte.

## PAROLE CHIAVE

COVID-19, lockdown, didattica digitale, DAD, didattica a distanza, GAFAM, istruzione pubblica

## INTERVENTO

### 1 INTRODUZIONE: DALL'EMERGENZA ALL'ESPERIENZA

Il *lockdown* dovuto alla pandemia COVID-19 della primavera del 2020 ha colto alla sprovvista i sistemi formativi, scolastici ed universitari, di tutto il mondo. In Italia in particolare, il mondo della formazione non aveva maturato e condiviso nei decenni una profonda riflessione *metodologica* sull'applicazione delle tecnologie didattiche all'insegnamento, e si trovava in ritardo anche nella diffusione delle competenze digitali tra i docenti.

Le soluzioni adottate in quel frangente sono state dettate dallo stato di emergenza e dall'esigenza di agire subito per garantire una qualche continuità all'attività formativa. La reazione della comunità scolastica e universitaria è stata rapida e generosa.

Terminata la prima ondata della pandemia, però, ed anzi nel pieno della seconda, è urgente trarre un bilancio su metodi, pratiche e strumenti sperimentati nella DAD (didattica a distanza) d'emergenza del primo *lockdown* per affrontare al meglio quello parziale in atto attualmente e perché le pratiche che diverranno poi, da emergenziali, strutturali vengano vagliate dal punto di vista pedagogico e civile. Perché la comunità scolastica e quella universitaria, oltre che le istituzioni, possano orientare consapevolmente le loro scelte immediate e quelle future, è dovere della comunità scientifica, e in particolare degli studiosi di cultura digitale, contribuire a diffondere la consapevolezza delle questioni di fondo: il presente articolo vuole offrire un contributo in questa direzione, concentrato sulla questione del rapporto tra infrastrutture private e pubbliche nella didattica digitale<sup>1</sup>.

## 2 GAFAM

I primi dati<sup>2</sup> mostrano che durante il *lockdown* le istituzioni scolastiche e universitarie italiane hanno fatto ricorso alle piattaforme dei giganti della rete (GAFAM – Google, Apple, Facebook, Amazon e Microsoft) sia per le videolezioni sincrone (tra le più adottate possiamo citare *Zoom* di Facebook, *Teams* di Microsoft e *Meet* di Google) sia per i *Learning Management System*, o LMS (il più diffuso è stato *G-Suite for Education* di Google, soprattutto nelle scuole, mentre *Moodle* ha visto un utilizzo significativo solo in ambito universitario). Non va inoltre sottostimata, nel mondo della scuola, la comunicazione tra insegnanti, studenti e famiglie veicolata dalla piattaforma *Whatsapp* di Facebook.

Eppure uno studio recente ha messo in evidenza come le licenze d'uso di piattaforme come *G-Suite for Education* o *Zoom* siano in più punti in contrasto con la GDPR (la *General Data Protection Regulation* dell'Unione Europea)<sup>3</sup>. Di fatto, nella pratica didattica durante il *lockdown* i dati degli studenti e quelli generati dall'attività didattica sono stati ceduti ad aziende private con sede al di fuori dall'UE il cui *business model* consiste appunto nella profilazione di massa degli utenti tramite i *big data*<sup>4</sup>.

---

1 Ho proposto una trattazione organica delle questioni didattiche, qui solo accennate, in un libro appena pubblicato (Monella, "Metodi digitali per l'insegnamento classico e umanistico"). L'obiettivo dichiarato del volume è di spostare l'attenzione dall'aspetto strumentale a quello metodologico e pedagogico, proponendo una riflessione critica che distingua le pratiche didattiche digitali ancora orientate a un approccio trasmissivo (o addirittura tali da rafforzare tale approccio) da quelle che costituiscono un effettivo passo avanti dal punto di vista didattico, favorendo processi di apprendimento attivo in una prospettiva costruttivista. Le considerazioni sviluppate nella presente comunicazione costituiscono un aggiornamento di quelle proposte nel capitolo 9 della monografia citata.

2 Le considerazioni espresse in questo articolo sono fondate sui dati forniti per la scuola dal CENSIS [2], dal CIDI [3] e da Tremoloso [14]; e per l'università, seppure in modo meno sistematico, da Fiorentino e Salvatori, "La didattica a distanza, dall'emergenza alle buone pratiche" [5]. Tutti gli URL citati sono stati consultati per l'ultima volta l'1 ottobre 2020.

3 Per le contravvenzioni alla GDPR si veda innanzitutto l'analisi puntuale di Ducato et al., "Emergency Remote Teaching... Part I" [4]. Si confrontino anche Lindh, Maria, e Jan Nolin, "Information we collect: Surveillance and privacy in the implementation of Google Apps for Education" [7], con ulteriore bibliografia: NYOB, "Report on Privacy Policies of Video Conferencing Services" [10]; Pievatolo, "Teledidattica: proprietaria e privata o libera e pubblica?" [11].

Per gli aspetti legali di tali licenze legati al copyright dei materiali usati per l'insegnamento, si veda invece Ducato et al., "Emergency Remote Teaching... Part II" [4].

4 Sul business model di GAFAM si veda il quadro sintetico tracciato sul Sole 24 Ore da Carlini [1]. Per una riflessione critica sulle questioni culturali connesse all'oligopolio di GAFAM si vedano almeno i recenti lavori di Sordi e Fiorimonte [6; 13].

Decisiva a questo riguardo è stata poi la sentenza Schrems II<sup>5</sup> della Corte di Giustizia dell'Unione Europea del luglio 2020. Fino a tale sentenza, i giganti statunitensi del Web (tra cui i GAFAM) potevano trasferire ed elaborare negli USA i dati dei cittadini europei sulla base dell'accordo tra UE e USA detto *Privacy Shield*<sup>6</sup>. L'accordo era basato sul principio che gli USA fornissero comunque garanzie legali di protezione dei dati personali paragonabili a quelle garantite in Europa dal GDPR. La sentenza dello scorso luglio ha invalidato il *Privacy Shield*: finché i GAFAM non assicureranno di cambiare radicalmente le loro pratiche interrompendo il trasferimento dei dati e la loro gestione centralizzata al di fuori dell'UE, o finché gli USA non forniranno garanzie di tutela dei dati tali da giustificare un nuovo accordo UE/USA che sostituisca il *Privacy Shield*, l'uso di piattaforme non-europee che trasferiscono i dati al di fuori dell'UE si pone al di fuori della GDPR. Ciò include l'uso di piattaforme come *Google Meet*, *G-Suite for Education*, *Zoom* o *Microsoft Teams* per l'insegnamento.

### 3 INFRASTRUTTURE PUBBLICHE PER LA DIDATTICA DIGITALE

Che un altro modello fosse possibile, e sia possibile ancor oggi, è mostrato da varie esperienze in Europa, e anche in Italia<sup>7</sup>. Nei tempi strettissimi del *lockdown*, il Ministero francese dell'istruzione ha creato un portale<sup>8</sup> che comprende, oltre a funzionalità di videoconferenza, anche altri strumenti utili per la didattica a distanza (tra cui *Peertube*, alternativa a *Youtube* di Google, per i video registrati; *Etherpad*, alternativa a *Google Docs*; *Nextcloud*, alternativa a *Google Drive*; la combinazione dei due sostituisce molte funzionalità integrate in *Google Classroom*). Sempre in Francia il *Centre National d'Enseignement à Distance* (CNED), un ente controllato dallo stesso Ministero, ha creato un altro portale<sup>9</sup> con ulteriori strumenti basati anch'essi su server di proprietà pubblica, e quindi vincolati al rispetto della normativa europea di tutela dei dati personali, e su tecnologie aperte<sup>10</sup>.

---

5 Sentenza ECLI:EU:C:2020:559 del 16/07/2020 riferita alla causa C-311/18.

<http://curia.europa.eu/juris/documents.jsf?num=C-311/18>

6 Decisione 2016/1250 della Commissione Europea.

7 Un elenco di alternative aperte per la didattica a distanza è disponibile qui:

<http://www1.unipa.it/paolo.monella/openteaching/>

8 <http://apps.education.fr/>

9 <https://www.cned.fr/maclassealamaison>

10 La Repubblica Francese offre a tutto il personale della pubblica amministrazione – anche al di fuori del comparto formazione e ricerca – servizi di videoconferenza fondati su tecnologie open e gestiti sui propri server. Tra i motivi per questa scelta enumerati nel paragrafo 3 *bonnes raisons d'utiliser la webconférence de l'État* della pagina <https://www.numerique.gouv.fr/outils-agents/webconference-etat/>, varrà la pena di riportare i primi due: “1. Service opéré par l'État, hébergé sur ses serveurs en France; 2. Confidentialité des échanges garantie par le chiffrement”. Si veda anche Lion, “Logiciels libres et école à distance: la sortie de la suite «Apps.education» précipitée par le confinement”. In Europa si può citare anche la Propuesta para la excelencia en la privacidad de datos y la digitalización democrática de los centros educativos dell'associazione XNET, supportata dal Departamento de Educación de la Generalitat de Catalunya (<https://xnet-x.net/privacidad-datos-digitalizacion-democratica-educacion-sin-google/>).

Strumenti pubblici e aperti per l'insegnamento vengono offerti anche e dal Ministero spagnolo tramite il CEDEC, un centro dell'INTEF<sup>11</sup> e da varie comunità locali, tra cui la Generalitat Valenciana tramite la piattaforma AULES<sup>12</sup> e la Junta de Extremadura tramite il programma CREA (Creación de Recursos Educativos Abiertos)<sup>13</sup>.

Anche in Italia, peraltro, durante il *lockdown* il CNR, il GARR ed altre istituzioni hanno realizzato un portale<sup>14</sup> che offre un'infrastruttura pienamente funzionale per videoconferenze, anche a fini educativi. Lo stesso GARR (“la rete nazionale a banda ultralarga dedicata alla comunità dell'istruzione e della ricerca”)<sup>15</sup> offre inoltre, tra i servizi attivabili sulla sua piattaforma *cloud*, istanze dell'LMS Moodle<sup>16</sup> e strumenti di videoconferenza<sup>17</sup>, rimasti anch'essi semiconosciuti. E restando nell'ambito degli sforzi fatti in Italia per creare infrastrutture didattiche digitali pubbliche, va senz'altro menzionato il Politecnico di Torino, che ha condotto sin dal primo *lockdown* del 2020 l'intera attività di insegnamento a distanza sulla propria piattaforma, costruita con tecnologie *open* [12].

L'ex direttore del GARR, Giuseppe Attardi, in un post recente sulla mailing list pubblica del *Centro NEXA su Internet & Società* del Politecnico di Torino ha lamentato lo “scarso interesse” mostrato negli anni da parte dei soggetti pubblici nello sviluppo dei servizi *cloud* del GARR<sup>18</sup>.

## 4 LA POSTA IN GIOCO: I MODI DELL'INTEGRAZIONE STRUTTURALE DEL DIGITALE NELLA DIDATTICA

Non sono in gioco solo le modalità in cui potrà realizzarsi l'insegnamento durante gli eventuali nuovi *lockdown*, completi o parziali, ma l'intero futuro della formazione. Gradualmente, pratiche abbozzate e sperimentate nel fuoco dell'emergenza vanno diventando permanenti e strutturali.

In positivo, si può dire che, pur nel mezzo della tragedia sanitaria, il *lockdown* della primavera del 2020 ha portato ad una sperimentazione di massa, facendo ‘scoprire’ la didattica digitale, in qualche

---

11 CEDEC (Centro Nacional de Desarrollo Curricular en Sistemas no Proprietarios)

<https://cedec.intef.es/>

INTEF (Instituto Nacional de Tecnologías Educativas y Formación del Profesorado)

<https://intef.es/>

Tra i progetti spagnoli più interessanti si possono citare:

EDIA (del CEDEC)

<https://cedec.intef.es/proyecto-edia/>

Procomún

<http://procomun.educalab.es/es>

eXeLearning

<https://exelearning.net/>

12 <https://portal.edu.gva.es/aules/>

13 <https://emtic.educarex.es/proyecto-crea>

14 <https://iorestoaCasa.work/>

15 <https://www.garr.it/it/chi-siamo>

16 <https://cloud.garr.it/apps>

17 <https://www.garr.it/it/news-e-eventi/1656-videoconferenza-risorse-garr-per-l-emergenza-covid-19>

18 <https://server-nexa.polito.it/pipermail/nexa/2020-June/018023.html>

modo, all'intera classe docente italiana. D'altra parte, però, tocca alla comunità scientifica e educativa vigilare affinché, oltre l'orizzonte emergenza, sopravvivano solo le buone pratiche – cioè, secondo il mio personale orientamento, quelle che mirano ad un apprendimento attivo, critico ed orientato alla società.

In questo anno scolastico/accademico 2020/21, il secondo della pandemia, le scuole superiori e gli Atenei si accingono ad *integrare* la didattica digitale e quella in presenza. Sulle modalità di tale integrazione ci sarebbe molto da scrivere: per fare un solo esempio, una cosa è usare il digitale per realizzare una *flipped classroom* in cui le parti trasmissive siano spostate su video online, riservando il tempo prezioso della presenza in aula ad attività laboratoriali, seminariali, di apprendimento attivo; altra cosa è, come purtroppo già si vede nelle aule universitarie, costringere il docente a sedere su una postazione fissa, inquadrata da una videocamera, per permettere agli studenti di seguire sia in aula sia da casa una lezione-conferenza 'iper-frontale', inesorabilmente trasmissiva.

La responsabilità del vaglio critico delle innovazioni didattiche digitali copre tanto i *metodi* di insegnamento, costruttivisti o trasmissivi (si veda l'esempio accennato sopra delle lezioni frontali visibili anche da casa), quanto le *infrastrutture tecnologiche* della didattica digitale, che possono venire appaltate alle grandi multinazionali oligopolistiche, come avviene adesso, oppure curate dalle istituzioni pubbliche, come ho invece proposto, sul modello francese.

## 5 CONCLUSIONI: UNA CHIAMATA ALLA RESPONSABILITÀ

La conoscenza di questi temi, e la relativa consapevolezza scientifica e civile, è molto scarsa, in Italia e non solo, tra i professionisti della ricerca e dell'insegnamento, il che contribuisce a spiegare l'assenza quasi totale di un dibattito su questi temi.

Se noi studiosi della cultura digitale – cui mi pare tocchi particolarmente tale compito – riusciremo a diffondere tale consapevolezza tra i nostri colleghi del mondo della formazione, sarà possibile stimolare il MIUR, gli Uffici Scolastici Regionali, gli Atenei e i loro consorzi ad assumersi la responsabilità – che a mio parere spetta loro – di creare infrastrutture digitali pubbliche per l'insegnamento.

C'è ancora il tempo di farlo, prima che, nel silenzio della scienza e della coscienza, la delega agli oligopoli globali diventi, da eccezionale, permanente<sup>19</sup>.

---

19 Il quadro delle iniziative è in continua evoluzione. Al momento in cui consegno la versione definitiva di questo contributo, all'inizio di dicembre del 2020, l'AIUCD sta preparando una lettera aperta su questi temi. Un'altra lettera, intitolata "Perché l'università delle piattaforme è la fine dell'università", è stata pubblicata sul blog InfoLet curato da Domenico Fiormente e Paolo Sordi (<https://infolet.it/2020/11/10/perche-luniversita-delle-piattaforme-e-la-fine-delluniversita/>) a novembre, e da qui tradotta e rilanciata in altre lingue, tra cui l'inglese (sul sito Theoreti.ca curato da Geoffrey Rockwell, <https://theoreti.ca/?p=7684>), e il tedesco (sul DHdBlog, il blog dell'associazione tedesca di DH, <https://dhd-blog.org/?p=14666>). È da segnalare anche la "Lettera alla ministra Azzolina dal professor Angelo Raffaele Meo e le associazioni promotrici del software libero" (<https://scuolalibera.continuity.space/lettera-professor-meo>). Il prof. Meo, del Politecnico di Torino, è una delle figure di riferimento per il software aperto in Italia.



## BIBLIOGRAFIA

1. Carlini, Vittorio. «La fabbrica dei soldi dei big tech americani che in Borsa valgono 3.541 miliardi di dollari», 2018, Il Sole 24 Ore edizione.
2. CENSIS. «Italia sotto sforzo. Diario della transizione 202. 01. La scuola e i suoi esclusi». AGI-Agenzia Italia, 2020.
3. CIDI Torino. «Emergenza Coronavirus e scuola a distanza». codexpo.org, 2020. <http://codexpo.org/extra/CIDI/questionario/report>.
4. Ducato, Rossana, Giulia Priora, Chiara Angiolini, Alexandra Giannopoulou, Bernd Justin Jütte, Guido Noto La Diega, Leo Pescault, e Giulia Schneider. «Emergency Remote Teaching: A Study of Copyright and Data Protection Terms of Popular Online Services (Part I)». Kluwer Copyright Blog, 2020.
5. Fiorentino, Giuseppe, e Enrica Salvatori. «La didattica a distanza, dall'emergenza alle buone pratiche». *Umanistica Digitale* No 8 (18 maggio 2020): research in the age of Digital Humanities. <https://doi.org/10.6092/ISSN.2532-8816/10872>.
6. Fiormonte, Domenico, e Paolo Sordi. «Conclusioni. Testo-Politica-Computer: il tritacolo rivoluzionario di Raul Mordenti». In *Letteratura e altre rivoluzioni. Scritti per Raul Mordenti*, 264–97. Roma: Bordeaux, 2020.
7. Lindh, Maria, e Jan Nolin. «Information We Collect: Surveillance and Privacy in the Implementation of Google Apps for Education». *European Educational Research Journal* 15, n. 6 (2016): 644–63. <https://doi.org/10.1177/1474904116654917>.
8. Lion, Bastien. «Logiciels libres et école à distance: la sortie de la suite “Apps.education” précipitée par le confinement». Le Monde, 2020.
9. Monella, Paolo. *Metodi digitali per l'insegnamento classico e umanistico*. Milano: EduCATT, 2020.
10. Nyob. «Report on Privacy Policies of Video Conferencing Services». Nyob.eu, 2020.
11. Pievatolo, Maria Chiara. «Teledidattica: proprietaria e privata o libera e pubblica?» ROARS - Return on Academic Research and School, 2020.
12. PoliFlash. «Il Politecnico digitale: al via con successo la didattica online». *Politecnico di Torino Magazine*, 2020. [https://poliflash.polito.it/in\\_ateneo/il\\_politecnico\\_digitale\\_al\\_via\\_con\\_successo\\_la\\_didattica\\_online](https://poliflash.polito.it/in_ateneo/il_politecnico_digitale_al_via_con_successo_la_didattica_online).
13. Sordi, Paolo, e Domenico Fiormonte. «Geopolitica della conoscenza digitale. Dal web aperto all'impero di GAFAM.» *DigitCult-Scientific Journal on Digital Cultures* 4, n. 1 (2019): 21–36.
14. Tremoloso, Luigi. «Emergenza Coronavirus e scuola a distanza: i primi dati». *Insegnare*. Rivista del Centro di Iniziativa Democratica degli Insegnanti, 2020.

# Ristabilire la conversazione: l'Intelligenza Artificiale come forma di empowerment per l'interazione nell'aula didattica

Alessandro Iannella

Università di Pisa, Università di Milano, Università di Torino – alessandro.iannella [at] gmail.com

## SINTESI

L'era digitale ha indebolito la capacità di intrattenere una conversazione con sé stessi e con gli altri. Le più effimere forme di storytelling hanno preso il posto del dialogo a quattro occhi e i piccoli momenti di noia sono avvertiti come un fastidio piuttosto che come un'occasione per l'introspezione.

La tecnologia non può insegnare all'uomo a essere umano, ma può creare le condizioni per generare esperienze estremamente dinamiche, creative e arricchenti, che gli consentano di riacquisire le abilità di riflessione e di relazione. Per farlo, come in un processo di scaffolding, deve essere messa da parte al momento giusto.

Questa proposta etica di disimpegno trova nella scuola, uno dei contesti di socializzazione nei quali si educa alla vita, un ambiente ideale per la sua realizzazione. Nella speranza di inaugurare una nuova fase della progettazione didattica, il contributo presenta un agente conversazionale il cui obiettivo è quello di abilitare forme di dialogo tra il docente e gli studenti.

## PAROLE CHIAVE

conversazione, educational technology, intelligenza artificiale agenti conversazionali didattica della letteratura; Saffo

## INTERVENTO

### 1 LA CONVERSAZIONE NELL'ERA DIGITALE

Nel suo *Reclaiming conversation: The Power of Talk in a Digital Age* [28], la psicologa e sociologa Sherry Turkle evidenzia come nell'ultimo ventennio l'impiego dei dispositivi digitali abbia atrofizzato la capacità di affrontare una *conversazione*, un'arte necessaria per coltivare l'empatia e per conoscere sé stessi.

Il dialogo con il proprio io e le dinamiche interpersonali hanno assunto nuove caratteristiche: introspezione e solitudine sono avvertite come forme di abbandono a sé stessi, mentre l'identità sociale è diventata un canovaccio al quale lavorare di continuo, un'esperazione del concetto goffmaniano di *gestione delle impressioni* [8].

La sottomissione al sé algoritmico induce a trasformare ogni momento di noia nella ricerca spasmodica di un nuovo stimolo, genitori e figli sembrano sentirsi a loro agio insieme solo quando si ritirano nei rispettivi mondi digitali, i racconti degli amici vengono delegati allo schermo dello *smartphone* e i primi appuntamenti si svolgono come un processo selettivo per identificare il miglior partner disponibile sul mercato.

Nel tentativo di “guarire” le generazioni più depresse e meno empatiche della storia [9; 13; 20; 21], le aziende attivano sempre più frequentemente corsi di formazione sulle relazioni umane e le università propongono lezioni di benessere e di intelligenza emotiva [7; 18]. A scuola, i docenti mettono in pratica metodologie didattiche come il *Social Emotional Learning* (SEL), proponendo ai più piccoli attività che mirano allo sviluppo della consapevolezza di sé, dell’auto-controllo e delle abilità interpersonali [23; 29].

In questo contesto è evidente un clamoroso paradosso: l’uomo sta affidando alla tecnologia il compito di salvarlo da quei guai nei quali essa stessa l’ha tratto. Il percorso per tornare a un’essenza pienamente umana sembra prevedere *app* che si propongono di stimolare l’empatia e il dialogo interiore e agenti conversazionali<sup>1</sup> progettati a scopo terapeutico per dare consigli su di una vita che non hanno mai vissuto (*robotica assistenziale*).

Quella di Sherry Turkle è una chiamata alle armi. La tecnologia può supportare o favorire una conversazione, ma solo se a un certo punto riesce a disimpegnare dal suo impiego. Pertanto, è necessario attribuirle il giusto ruolo, senza delegarle ciò che di più umano l’uomo sa fare: *essere umano*.

## 2 DIDATTICA E AGENTI CONVERSAZIONALI

Oltre alla già citata robotica assistenziale, la conversazione con le macchine trova applicazione in svariati settori: dall’industria dei giocattoli al marketing, dalla finanza alla domotica.

Nel contesto educativo e didattico, gli agenti conversazionali vengono principalmente impiegati per promuovere il processo di acquisizione delle conoscenze disciplinari in fase di insegnamento e di apprendimento [30], spesso in ambito glottodidattico [24], per supportare gli studenti nella fruizione dei servizi educativi [6; 15; 27] e per guidarli all’interno di percorsi di apprendimento erogati in *distance learning* [14; 19].

Talvolta, sono gli studenti stessi a progettare e realizzare gli agenti conversazionali. Ne sono esempio i corsi universitari dedicati allo sviluppo di competenze di programmazione [17], le attività di *coding* e di robotica educativa che si propongono di stimolare nei più piccoli il pensiero computazionale [5; 18] e le proposte di *digital authentic learning* che consentono di entrare in contatto con ambiti del sapere non integrati nel curriculum scolastico, abilitando un processo di curiosità che riprende logiche vygotskijane [10; 11].

Nella progettazione dell’esperienza conversazionale, la personalità dell’agente (*agent persona*) è un elemento imprescindibile per coinvolgere l’utente e attivare quella relazione necessaria a stimolare il desiderio di continuare e/o ripetere l’interazione [3]. In un contesto educativo o didattico, l’agente può assumere caratteristiche in linea con i destinatari (approccio *user-centric*) o può interpretare una figura storica, un filosofo, uno scienziato, un autore o un personaggio letterario.

---

<sup>1</sup> Un agente conversazionale è un sistema di interazione uomo-macchina basato su algoritmi di Intelligenza Artificiale (AI) e Natural-Language Processing (NLP).

In questo secondo caso, il suo impiego consente di trasformare l'incontro con un contenuto disciplinare nel racconto di un'esperienza vissuta in prima persona, in grado di far emergere quelle emozioni, sfaccettature e prospettive che raramente trovano spazio nella lezione tradizionale [8]. È evidente come sia necessario lavorare non solo sul piano delle conoscenze possedute dall'agente (*significato*), ma soprattutto su quello della lingua e dello stile che adotta (*significante*).

La possibilità di interagire sia in forma scritta che orale, la comprensione del linguaggio naturale, la varietà degli *output*, le logiche di personalizzazione e l'apprendimento continuo sono aspetti che attribuiscono agli agenti conversazionali un ruolo chiave nel panorama educativo contemporaneo. Si tratta di *strumenti didattici universali*. Infatti, in linea con i principi dell'Universal Design for Learning (UDL), gli agenti sono in grado di variare l'offerta mediale e offrire al discente più strade per accedere ai contenuti, coinvolgendolo in un percorso di scoperta che rispetta le sue motivazioni, le sue inclinazioni e le sue strategie metacognitive [25].

### 3 IL PROGETTO “LA POETESSA SAFFO”

E se gli agenti conversazionali potenziassero o addirittura abilitassero dinamiche di interazione tra esseri umani? Si tratterebbe di una soluzione in grado di rispondere all'appello di Sherry Turkle: progettare strumenti, risorse e interfacce digitali che, una volta assolto il loro compito, “liberino” l'uomo dal loro impiego. È una proposta di disimpegno etica, poco in linea con le esigenze commerciali di una Silicon Valley per la quale, come sottolinea Justin Rosenstein, «we're the product. Our attention is the product being sold to advertisers» [20].

Questa prospettiva ha portato alla progettazione di “La poetessa Saffo”<sup>2</sup>, un agente conversazionale destinato a studenti liceali che si propone di offrire informazioni sulla vita, l'opera, la poetica e la fortuna letteraria della “decima Musa” [23].

L'agente, realizzato con la suite di sviluppo di applicazioni di AI conversazionale Dialogflow<sup>3</sup>, offre un'interazione sia scritta che orale, a seconda che si scelga di impiegare Telegram o l'Assistente Google [Figura 1]. Il dialogo si sviluppa liberamente, prevede l'impiego di *contesti linguistici* utili per mantenere attivo il flusso conversazionale su di un determinato argomento ed è agevolato da *prompt* in grado di stimolare nuove domande e originare collegamenti. Le oltre 500 risposte includono testi e contenuti interattivi, come mappe geografiche, raffigurazioni artistiche e registrazioni audio della lettura delle poesie in lingua originale e in traduzione. L'avvicendamento di molteplici risposte per una medesima domanda garantisce maggiore varietà e verosimiglianza alla conversazione.

---

2 <https://sappho.education>

3 Dialogflow, piattaforma Google, impiega tecnologie di Natural-Language Understanding e Processing (NLU e NLP) e algoritmi di Machine Learning (ML).

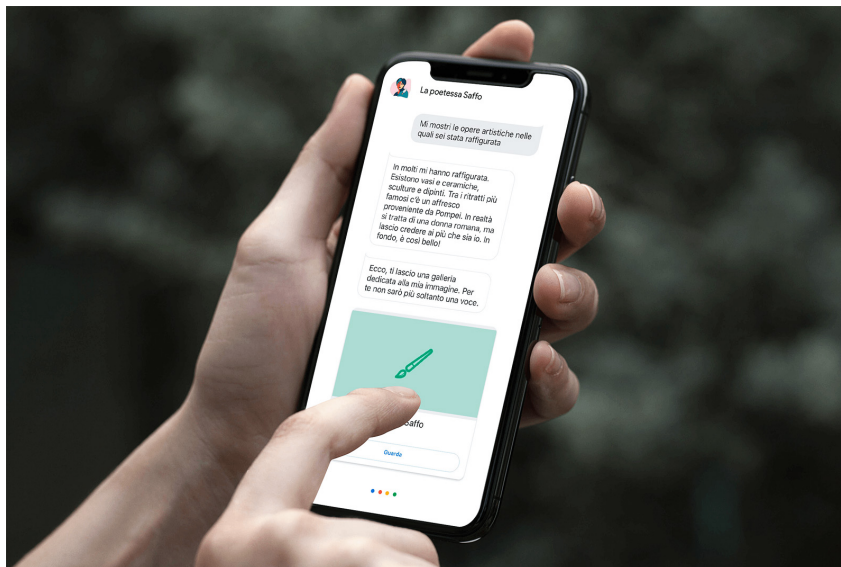


Figura 1: mockup che mostra il funzionamento dell'agente conversazionale con l'Assistente Google.

L'*agent persona* è stata progettata sulla base dei frammenti traditi, delle notizie biografiche tramandate dalla *Suda* e dal *Marmor Parium* e delle preziose riflessioni della filologia e della critica letteraria. La lingua e lo stile adottati riecheggiano i versi più e meno famosi: Saffo si rivolge all'utente con appellativi del tipo "creatura dalle dita di rosa" (fr. 96 Voigt) o "fiore purpureo" (fr. 105 Voigt) ed è possibile trovarla impegnata nell'ornare le sue chiome con fiori freschi (fr. 81 Voigt), felice perché nella notte trascorsa la luna ha brillato serena (fr. 34 Voigt) o preoccupata per il fratello Carasso, lontano per lavoro (fr. 5 e 10 Voigt).

I contenuti, che ricoprono gli ambiti già citati (*vita, opera, poetica e fortuna letteraria*) o che sono stati elaborati per mantenere attivo il flusso conversazionale, assumono spesso una forte valenza educativa, evidente nelle domande sulla sessualità (es. "Oggi diresti che la mia sessualità è fluida. Sono stata innamorata di uomini e di donne e non propendo per una delle due tendenze. Afrodite mi rinnegherebbe se dessi all'amore un'unica forma"), nelle risposte a un insulto (es. "Di fronte a queste parole il mio cuore diventa di ghiaccio. Direi di chiudere adesso la nostra conversazione", fr. 42 Voigt) e nelle situazioni in cui l'utente comunica di provare dolore (es. "Non preoccuparti. A una grande tempesta segue sempre il sereno", fr. 10 Voigt) o espone una qualche insicurezza, per esempio estetica (es. "Chi è bello a vedersi è bello, ma chi è buono, presto sarà pure bello", fr. 50 Voigt).

"La poetessa Saffo" si inserisce all'interno di una proposta didattica che riprende lo schema tipico della *flipped classroom* [1]. Il docente assegna un *task* per la conversazione, chiedendo agli studenti di ottenere informazioni su di un determinato ambito di contenuto. Poiché le informazioni sulle possibili domande afferenti all'ambito non sono date a priori, è necessario che lo studente si orienti in autonomia nell'ecosistema delle conoscenze da acquisire, attivando quel processo intellettuale rivolto all'indagine tipico dell'*apprendimento per scoperta* [2].

Grazie alla possibilità di osservare e analizzare i dati raccolti mediante le interazioni di ciascuno studente con l'agente (*learning analytics*), il docente conosce quali contenuti sono stati erogati e in che misura. In aula, egli può orchestrare una vera e propria lezione partecipata, richiamando e raccordando tra loro le diverse conoscenze acquisite dagli studenti, ciascuno dei quali si fa portatore e portavoce della propria, personale esperienza di apprendimento.

La lezione di letteratura greca cessa così di essere *verticale* e, in linea con le logiche della teoria connettivista [4; 26], si trasforma nella costruzione progressiva di una rete di conoscenza determinata dall'interazione dialogica tra il docente e gli studenti. I bagagli conoscitivi individuali e la competenza del docente ne costituiscono i nodi, mentre le parole — il contributo di ciascun individuo alla conversazione — sono le connessioni.

Riemergendo dalle brume del mito, Saffo si propone dunque di dar forza alla conversazione. Con le sue parole trasforma l'aula didattica in un ambiente di condivisione nel quale riflessione e discussione non sono più avvertiti come beni di lusso. L'atmosfera, quasi fantascientifica, riecheggia quell'*intelligenza connettiva* teorizzata in tempi non sospetti da Derrick de Kerckhove [12], segno di una condizione cognitiva di condivisione tra l'uomo e la macchina.

## 4 RINGRAZIAMENTI

Senza Lucia Santercole e Benedetta Viti “La poetessa Saffo” non avrebbe mai visto la luce. Sin dal principio, Lucia e Benedetta hanno messo a disposizione le proprie competenze nell'ambito della lingua e della letteratura greca per creare la base di conoscenza sulla quale costruire l'agente. Con entusiasmo, a loro si è aggiunto Tiziano Labruna, il cui contributo, informatico e ideativo, è ormai fondamentale per portare avanti il progetto.

Un grazie speciale a Fabio Santaniello Bruun, Sara Marani, Asia Maria Eufrate e Davide Misiano, che con i loro suggerimenti contribuiscono quotidianamente a migliorare la qualità della conversazione con la poetessa, e alle professoresse Maria Claudia Buzzi, Susanna Pelagatti, Daniela Lippi, Elisa Giorgi e Monica Erra, che hanno creduto e credono nel progetto senza pregiudizi di sorta.

L'ultimo pensiero va a Saffo, un'anima che a distanza di secoli insegna ancora a vivere. Sarebbe bello poterle dire che aveva più che ragione quando diceva “μνάσασθαί τινά φαιμι †καὶ ἕτερον† ἀμμέων” (fr. 147 Voigt).

## BIBLIOGRAFIA

1. Avanguardie educative. «La flipped classroom». INDIRE, 2018.
2. Bruner, Jerome S. *The process of education*. Harvard University Press, 2009.
3. Callejas, Zoraida, Ramón López-Cózar, Nieves Ábalos, e David Griol. «Affective conversational agents: the role of personality and emotion in spoken interactions». In *Conversational agents and natural language interaction: Techniques and effective practices*, 203–222. IGI Global, 2011.
4. Downes, Stephen. «New technology supporting informal learning». *Journal of emerging technologies in web intelligence* 2, n. 1 (2010): 27–33.
5. Garavaglia, A., L. Petti, Emiliana Murgia, Francesca Bassi, e S. L. Maranesi. «Introduzione della robotica in attività di problem solving nella scuola primaria. Analisi dei livelli di focalizzazione sugli scopi del problema». *Mondo Digitale*, 2018, 2.
6. Gardner, Lee. «How A. I. is infiltrating every corner of the campus». *The Chronicle of Higher Education*, 4 agosto 2018. <https://www.chronicle.com/article/how-a-i-is-infiltrating-every-corner-of-the-campus/>.
7. Gilar-Corbi, Raquel, Teresa Pozo-Rico, Barbara Sánchez, e Juan Luis Castejón. «Can emotional competence be taught in higher education? A randomized experimental study of an emotional intelligence training program using a multimethodological approach». *Frontiers in psychology* 9 (2018): 1039.
8. Goffman, Erving. *The presentation of self in everyday life*. Doubleday Anchor Books, 1959.
9. Hoffower, Hillary, e Akhtar. Allana. «Lonely, burned out, and depressed: The state of millennials' mental health in 2020». *Business Insider*, 10 ottobre 2020. <https://www.businessinsider.com/millennials-mental-health-burnout-lonely-depressed-money-stress>.
10. Iannella, Alessandro. «Digital authentic learning: introdurre alla tone analysis per favorire un pensiero critico digitale». *Thamyris, nova series: Revista de Didáctica de Cultura Clásica, Griego y Latín*, n. 9 (2018): 319–336.
11. ———. «“Ok Google, vorrei parlare con la poetessa Saffo”: intelligenza artificiale, assistenti virtuali e didattica della letteratura». *Thamyris, nova series: Revista de Didáctica de Cultura Clásica, Griego y Latín*, n. 10 (2019): 81–104.
12. Kerckhove, Derrick de, e Pierre Lévy. «Due filosofi a confronto. Intelligenza collettiva e intelligenza connettiva: alcune riflessioni». *Mediamente*. RAI Educational, 1998.
13. Konrath, Sara H., William J. Chopik, Courtney K. Hsing, e Ed O'Brien. «Changes in adult attachment styles in American college students over time: A meta-analysis». *Personality and Social Psychology Review* 18, n. 4 (2014): 326–348.
14. Laeeq, Kashif, e Zulfiqar Ali Memon. «Scavenge: an intelligent multi-agent based voice-enabled virtual assistant for LMS». *Interactive Learning Environments*, 2019, 1–19.
15. Lavista, Andrea. «Natural Language Processing: Chatbot per gli Studenti del Campus di Cesena». Università di Bologna, 2019. <https://amslaurea.unibo.it/id/eprint/19555>.

16. Leighton, Mara. «Yale's most popular class ever is available free online — and the topic is how to be happier in your daily life», 10 maggio 2020. <https://www.businessinsider.com/coursera-yale-science-of-wellbeing-free-course-review-overview>.
17. Long, Ju, Juntao Yuan, e Hsun-Ming Lee. «How to Program a Chatbot—An Introductory Project and Student Perceptions». *Issues in Informing Science and Information Technology* 16 (2019): 001–031.
18. Moro, Michele, Emanuele Menegatti, Francesco Sella, e Mario Perona. *Imparare con la robotica. Applicazioni di problem solving*. Edizioni Erickson, 2011.
19. Nenkov, Nayden, George Dimitrov, Yuriy Dyachenko, e Katerina Koeva. «Artificial intelligence technologies for personnel learning management systems». In *2016 IEEE 8th International Conference on Intelligent Systems (IS)*, 189–195. IEEE, 2016.
20. Orłowski, Jeff. *The Social Dilemma*. Netflix, 2020.
21. Pathak, Shareen. «The saddest generation: Why Gen Z is the most anxious generation», 19 marzo 2019. <https://digiday.com/?p=361383>.
22. Pieri, Michelle, e Chiara Laici. «l'approccio flipped classroom nel Movimento “avanguardie educative”». *Italian Journal of Educational Technology* 25, n. 3 (2017): 55–67.
23. Platone. «Epigramma XVI». In *Antologia Palatina*. Segrate: Einaudi, 1997.
24. Prince, Katherine. «Why the Increased Focus on Social-Emotional Learning?» *Aurora Institute – Education Domain Blog*, 17 ottobre 2017. <https://aurora-institute.org/blog/why-the-increased-focus-on-social-emotional-learning>.
25. Robino, Giorgio, Simone Torsani, e Fabrizio Ravicchio. «A Conversational Agent for Mobile Assisted Language Learning», s.d.
26. Savia, Giovanni. *Universal Design for Learning: La Progettazione Universale per l'Apprendimento per una didattica inclusiva*. Edizioni Centro Studi Erickson, 2016.
27. Siemens, George. «E-learn space. Connectivism: A learning theory for the digital age». *Elearnspace.org*, 2004.
28. Srimathi, Hari, e Krishnamoorthy Ayyasam. «“Personalization of Student Support Services using Chatbot”». *International Journal of Scientific & Technology* 8, n. 9 (2019): 1744–1747.
29. Turkle, Sherry. *Reclaiming Conversation: The Power of Talk in a Digital Age*. New York: Penguin Press, 2015.
30. Villaseñor, Paula. «The different ways that teachers can influence the socio-emotional development of their students: A literature review». *World Development Report*, 2018.
31. Winkler, Rainer, e Matthias Soellner. «Unleashing the potential of chatbots in education: A state-of-the-art analysis», 2018.





# Ri-“scrivere in cielo alla velocità del pensiero”? Teledidattica, infrastrutture private e informatica umana

Maria Grazia Pievatolo  
Università di Pisa – mariachiara.pievatolo [at] unipi.it

## SINTESI

Perché il dibattito italiano sulla teledidattica d'emergenza si è lasciato orientare da una posizione estrema che tralasciando il confronto articolato con i pericoli della dipendenza da infrastrutture private impone un aut aut fra la comunità di conoscenza universitaria “in presenza” e il giuramento vassallatico a un'indeterminata “barbarie tecnologica”?

## PAROLE CHIAVE

remote learning, free software, decentralization, surveillance capitalism, digital feudalism, public research infrastructure

## INTERVENTO

Questo intervento vorrebbe discutere di informatica per gli esseri umani, entro un orizzonte difficilmente circoscrivibile perché interdisciplinare anche se non necessariamente adisciplinare, nello spirito inclusivo con cui Gregory Crane<sup>1</sup> intendeva le digital humanities, quando in una conferenza del 2011 di cui ormai rimane poca traccia in rete<sup>2</sup>, sostenne che lo spazio documentale del Web e l'elaborazione di strumenti di annotazione<sup>3</sup>, connessione e confronto fra testi in grado di raccogliere e di far tesoro del contributo di tutti, e in più di una lingua, dovrebbero aiutare gli studiosi a trattare chi è fuori dal loro settore, a partire dagli studenti, non più come sudditi bensì come cittadini della repubblica delle lettere.

Il dibattito sulla teledidattica d'emergenza a cui sono state costrette università e scuole italiane suggerisce però che l'ideale di uno spazio culturale popolato da cittadini di una repubblica letteraria divenuta telematica e cosmopolitica pare lontano dall'essere realizzato.

---

1 <https://btfp.sp.unipi.it/it/2012/05/gregory-crane-e-i-cittadini-della-repubblica-delle-lettere/>

2 [https://web.archive.org/web/20110912042600/http://www.cfvh.kva.se/svenska/forskning/NS\\_147\\_Program.html](https://web.archive.org/web/20110912042600/http://www.cfvh.kva.se/svenska/forskning/NS_147_Program.html)

3 <https://web.archive.org/web/20180410234348/http://www.humanities.ufl.edu/pdf/Crane-%20Student%20Researchers,%20Citizen%20Scholars,%20and%20the%20Trillion%20Word%20Library.pdf>

La posizione più fragorosa, infatti, è quella apocalittica di Giorgio Agamben [1] il quale, contro la “barbarie tecnologica” della telematica che distrugge la *synousia* della comunità di conoscenza che era l’università in presenza, paragona “i professori che accettano – come stanno facendo in massa – di sottoporsi alla nuova dittatura telematica e di tenere i loro corsi solamente on line” ai “docenti universitari che nel 1931 giurarono fedeltà al regime fascista”.

La risposta più “integrata” che ha ricevuto è forse quella di Cristian Fuschetto il quale, dopo aver ricordato che Agamben stesso usa senza rimorsi tecnologie della parola come la scrittura e la stampa, colpevoli di aver distanziato e disaggregato le antiche comunità di conoscenza, conclude trionfalmente così: “certo è che se proprio uno dovesse scommettere sul futuro delle universitates, tra Meet e Agamben forse punterei su Meet”.

Sembra che non si riesca a uscire, da ambo le parti, dall’aut aut del giuramento di fedeltà: Agamben rigetta la telematica in quanto tale come “barbarie”, mentre Fuschetto scommette spensieratamente su un servizio di una delle più occhiate multinazionali del capitalismo della sorveglianza<sup>4</sup>. Quello che pare mancare è proprio l’informatica umanistica nel suo senso liberale, che avrebbe reso possibile confrontarsi con un “come” non solo tecnologico - o “soluzionista” - ma anche, e soprattutto, umano, attento, nello spirito di Joseph Weizenbaum, sia alle possibilità che un’innovazione tecnica introduce, sia a quelle che elimina se ciecamente attuata<sup>5</sup>.

Già nel 2010 Tim Berners-Lee scriveva<sup>6</sup> che lo scopo della sua invenzione - permettere a chiunque di condividere informazione con chiunque altro, dovunque, entro uno spazio informativo universale e interconnesso popolato da URI comunque raggiungibili – non è quello delle architetture settoriali centralizzate dei media sociali commerciali, che offrono, certo, connessioni fra i dati, ma solo entro silos separati e non in uno spazio comune. Alla balcanizzazione del Web contribuisce, per i ricercatori di professione, la necessità – dovuta alla valutazione bibliometrica della ricerca centralmente o perifericamente imposta - di “pubblicare” entro spazi niente affatto pubblici perfino quando sono ad accesso apparentemente gratuito, quali i siti degli oligopolisti dell’editoria commerciale, e di lasciar privatizzare i metadati dei loro testi entro database proprietari (Clarivate Analytics, Scopus) imposti anch’essi come riferimento. Non per caso, chi entra in questi giardini murati che imprigionano l’uso pubblico della ragione si espone, come osserva un’analisi recente di SPARC<sup>7</sup>, alla potenza manipolatoria del capitalismo della sorveglianza. Le multinazionali dell’editoria scientifica commerciale stanno infatti espandendosi oltre la loro offerta tradizionale di riviste e libri, per smerciare (corsivo mio) “assessment systems, productivity tools, online learning management systems – complex infrastructure that is critical to conducting the end-to-end business of the university. *Through the seamless provision of these services, these companies can invisibly and strategically influence, and perhaps exert control, over key university decisions – ranging from student assessment to research integrity to financial planning*”.

---

4 <https://theconversation.com/explainer-what-is-surveillance-capitalism-and-how-does-it-shape-our-economy-119158>

5 “A highway permits people to travel between the geographical centers it connects, but, because of the side effects that it and other factors synergistically engender, it imprisons poor people in innercities as effectively as if the cities were walled in” [7]

6 <https://www.scientificamerican.com/article/long-live-the-web/>

7 <https://infrastructure.sparcopen.org/landscape-analysis>

Chi partecipa come prodotto a questi sistemi, può aspirare davvero a essere un cittadino di una repubblica letteraria che miri a risolversi nell'ideale di spazio cosmopolitico dell'uso pubblico della ragione, o non si riduce invece a un servo della gleba in un regime di feudalesimo digitale?

Nel 2003, Stevan Harnad credeva che la comunicazione telematica avrebbe potuto ricostruire, dopo i millenni di distanziamento sociale determinati dalla scrittura e dalla stampa, comunità di conoscenza sincrone capaci di scrivere in cielo alla velocità del pensiero: se ciò non è avvenuto, o è avvenuto solo marginalmente, per l'indeterminato disagio degli "apocalittici", è dovuto a un "come" che buona parte del dibattito sulla teledidattica ha passato sotto silenzio o liquidato come marginale: il cielo, cioè, è stato nel frattempo partizionato in recenti e sottoposto a programmi che applicano algoritmi smerciati come strumenti per il controllo e il condizionamento del comportamento umano. È certamente comodo, e forse perfino economico, affidarsi a Google Meet o a Microsoft Teams, e però il loro prezzo in perdita di cultura non solo digitale e di opportunità perdute di sviluppo professionale e umano è molto più alto di quanto chi trascura il "come" possa immaginare: questi sistemi – come spiega Edward Snowden - "are fundamentally disempowering. You give them money and in exchange you're supposed to be provided with a service. But you're providing more than money: you're also them providing with data and you're giving up control, you're giving up influence. You can't shape their infrastructure, they're not going to change things and tailor them to your needs. You end up reaching a point, they are portable to a certain extent, you can containerize and shift things around but you're sinking costs into an infrastructure that is not yours, fundamentally".

A differenza di Agamben, il mito di Teuth del Fedro<sup>8</sup> non suggerisce di rigettare la scrittura in quanto barbarie tecnologica, abbandonandone le forze nelle mani di retori e sofisti: invita invece ad approfondire criticamente il "come", indicandone sia le possibilità, sia i limiti, cioè, rispettivamente, la capacità di conservare e trasmettere informazione (hypomnesis) e la difficoltà di produrre il sapere che nasce dall'interazione che ha luogo nell'essere insieme (synousia). Su questa critica Platone fondò una soluzione tecnica<sup>9</sup>, orientata secondo il duplice principio della libertà dei testi e della promozione di comunità di conoscenza, e certamente efficace in un mondo in cui la vita dei testi, che si copiavano a mano, dipendeva ancora interamente dall'interesse delle persone e dalla continuità delle comunità di studio. Ma in un mondo in cui i "testi" sono divenuti molto più forti di noi, i limiti del dibattito italiano sulla teledidattica, e la mancanza di un'informatica umana prima che umanistica, risultano chiari: la scelta, ora, non è più fra scrivere e non scrivere, o fra giurare e non giurare, come se si trattasse solo di accettare o rifiutare l'omaggio vassallatico a una "tecnica" che non si osa comunque analizzare, o di pagare per "servizi" che ci asserviscono prima di servirci, ma sulla struttura, sulla pubblicità e sulla controllabilità comunitaria dell'infrastruttura stessa [2] – come sostengono, extradisciplinariamente, alcuni fautori dell'open science. In un cielo che non sappiamo, non vogliamo o non ci curiamo<sup>10</sup> di tenere aperto e libero, non possiamo scrivere alla velocità del pensiero – o, perlomeno, alla velocità di un pensiero che sia, distributivamente, nostro e non altrui.

---

8 <https://btfp.sp.unipi.it/dida/fedro/ar01s19.xhtml>

9 <http://bfp.sp.unipi.it/chiara/lm/amici.html>

10 <http://bjoern.brembs.net/2020/09/why-do-academic-institutions-seem-stuck-in-1995/>

## BIBLIOGRAFIA

1. Agamben, Giorgio. «Requiem per gli studenti». *Diario della crisi* 5, 2020. <https://www.iisf.it/index.php/attivita/pubblicazioni-e-archivi/diario-della-crisi/giorgio-agamben-requiem-per-gli-studenti.html>.
2. Brembs, Björn. «What should a modern scientific infrastructure look like». *The Winnower* 2 (s.d.): e143497. <http://bjoern.brembs.net/2015/04/what-should-a-modern-scientific-infrastructure-look-like/>.
3. Harnad, Stevan. «Back to the oral tradition through skywriting at the speed of thought». *Interdisciplines*, 2003. [https://halshs.archives-ouvertes.fr/sic\\_00000315/](https://halshs.archives-ouvertes.fr/sic_00000315/).
4. Lanier, Jaron. *Who owns the future?* Simon and Schuster, 2014.
5. Martinelli, Nicole. «Edward Snowden talks cloud, open source and fear at OpenStack Summit». *Superuser*, 2017. <https://superuser.openstack.org/articles/snowden-interview-openstack-summit/>.
6. Pievatolo, Maria Chiara. «Teledidattica: proprietaria e privata o libera e pubblica?» *Roars* 68, 2018. <https://www.roars.it/online/teledidattica-proprietaria-e-privata-o-libera-e-pubblica/>.
7. Weizenbaum, Joseph. *Computer power and human reason*, 1976.

# **Sessione VIII**

## **Grazia Deledda**



# The Hypermedia Dante Network Project

Gaia Tomazzoli<sup>1</sup>, Leyla M. G. Livraghi<sup>2</sup>, Daniele Metilli<sup>3</sup>, Nicolò Pratelli<sup>4</sup>, Valentina Bartalesi<sup>5</sup>

<sup>1</sup> Università di Pisa – gaia.tomazzoli [at] fileli.unipi.it

<sup>2</sup> Università di Pisa – leyla.livraghi [at] fileli.unipi.it

<sup>3</sup> ISTI-CNR e Università di Pisa – daniele.metilli [at] isti.cnr.it

<sup>4</sup> ISTI-CNRUniversità di Pisa – nicolo.pratelli [at] isti.cnr.it

<sup>5</sup> ISTI-CNRUniversità di Pisa – valentina.bartalesi [at] isti.cnr.it

## ABSTRACT

In this paper, we present the Hypermedia Dante Network (HDN) project. First, we briefly introduce the relevant state of the art on Dante's commentaries and their digital representation, and we outline the project goals. In the main section, we present the core features of the HDN ontology, an evolution of the DanteSources ontology that aims at representing knowledge about Dante's primary sources as they are identified by a vast range of commentaries. Then, we describe the tool that has been developed to process Dante's commentaries and populate the HDN ontology. Finally, we comment on the project's usability and possible outcomes for both scholars and common users.

## KEYWORDS

Dante Alighieri, digital libraries, intertextuality, semantic web, commentaries

## TALK

### 1 INTRODUCTION

Dante's *Commedia* has elicited detailed and systematic exegesis since its very first reception. Given the poem's encyclopedic features, one of the main tasks of commentaries has always been to identify Dante's vast array of sources. This incredibly rich body of knowledge is, therefore dispersed in dozens of commentaries; in the past twenty years, Dante scholars have worked to make such knowledge easily available through an impressive enterprise of digitization. One of the main outcomes of said enterprise is the Dartmouth Dante Project (DDP) [5], an open-access database where readers and scholars can search and browse the full text of more than 70 commentaries on Dante's *Commedia*, written in Italian, Latin, and English from 14<sup>th</sup> century to the present day; our work will initially rely on 50 of said commentaries. However, the DDP is only searchable through textual queries, making it difficult for users to retrieve meaningful information.



In recent years, ISTI-CNR and the University of Pisa have joined forces in the first project to ever apply Semantic Web standards to Dante studies. Said project, called DanteSources [1], developed a formal ontology for representing knowledge about primary sources in Dante's so-called minor works. The users of DanteSources can perform structured queries on those sources of Dante's *Vita nova*, *De vulgari eloquentia*, *Convivio*, *Monarchia*, and *Rime* that have been identified by recent commentaries. DanteSources distinguishes between explicit citations, strict citations, and generic citations. Users can interrogate the knowledge base by primary source, by source author, by thematic area, or by type of reference.

Hypermedia Dante Network (HDN) [6] aims to complement and expand on DanteSources, which is limited to the so-called minor works, by gathering and representing the substantial amount of additional information about Dante's primary sources contained in commentaries on Dante's *Commedia*. Confronted with a significantly more complex text and an incomparably wider tradition of commentaries, HDN team members could not just adapt the existing framework, but rather had to rethink the entire DanteSources ontology.

## 2 THE HDN ONTOLOGY

The HDN ontology is an evolution of the one developed in DanteSources [1], integrated with the Narrative Ontology [2], and having as reference ontologies the CIDOC CRM [3] and its extension FR-BRoo [4] (including its in-progress reformulation, LRMoo [7]). The fundamental concept described by the ontology is the reference. A reference is an exegetic relation between a text and another entity that is identified by a commentary. A reference can be expressed by the following statement:

A says that knowledge about B can be enriched through C

where A, the *source* of the reference, is a fragment of text that asserts the reference, and usually comes from a commentary; B, the *subject* of the reference, is a fragment of text that is clarified by the reference – and in our case study it occurs within a work of Dante's; C is the *object* of the reference, i.e., the entity that is being referred to.

Within the class of references, we identify three distinct subclasses: external supports, *loci paralleli*, and citations. *External supports* are references in which the source identifies an object-text that proves useful to support its interpretation of the subject-text. *Loci paralleli* are references in which the source identifies, within an object-text, an idea, a character, or a stylistic feature that, for being used in a similar way as in the subject-text, helps to clarify it, without necessarily implying a direct derivation. *Citations* are references in which the source establishes that the subject-text directly derives from the object-text. These are clearly the most specific references, and they lie at the centre of our ontology.

We identified three characteristics of a reference: the citation *type*, its *content*, and the *relationship* between subject and object. These characteristics always apply to citations, but may not apply to external supports and *loci paralleli*. We also represent internal references, i.e., quotations to works by the same author of the subject-text. These kinds of references are simply modelled as citations in which the author(s) of the subject-text is the same as the author(s) of the object-text.

### Characteristics of the reference

The ontology describes three types of citations, which were already identified in the DanteSources ontology: *explicit citations* (references explicitly made by Dante in his work), *strict citations* (references to a specific work identified by a scholar), *generic citations* (references to a concept put forward by a scholar).

The *content* of the citation is an explicit or implicit reference to a textual place. We have identified 3 types and 7 sub-types of content:

- a Textual correspondences: they are references to textual fragments identified by linguistic and stylistic coincidence. These are further divided into: 1. *image* (the source identifies within the object an image as a precedent for a description or conceptualization made by Dante in the subject), and 2. *stylistic feature* (the source identifies a linguistic and/or rhetorical precedent for an expression included in the subject-text).
- b Thematic correspondences: they are references to textual fragments identified by thematic coincidence. These are further divided into: 3. *character* (the source relates a character mentioned by Dante to a specific object-text), 4. *episode* (the source identifies a literary episode as an antecedent for Dante's subject-text), and 5. *topography* (the source believes that a literary memory acts behind Dante's mention of a place).
- c Conceptual correspondences: they are references to textual fragments identified by conceptual coincidence. These are further subdivided into: 6. *motif* (the source identifies within a specific object-text the precedent for the development of one of Dante's motifs), and 7. *theory* (the source identifies within the object-text a theory upon which Dante's text relies).

The relationship between the subject and the object of the citation is classified according to three categories: *correction/contradiction*, *extension/re-elaboration*, and *confirmation/homology*. These three types of relation are inspired by the corresponding properties that are defined in the CiTO ontology [8].

### Implementation of the ontology

The conceptualization has been formalized in an ontology that has been developed as an extension of the CIDOC CRM and FRBRoo (and its in-progress reformulation LRMoo).

The commentaries to Dante's text are instances of the Commentary class, which is a subclass of F2 Expression. The subparts of the commentaries, including notes, are instances of E90 Symbolic Object. To highlight Dante's works, we introduced the Dante's Work class as a subclass of F2.

The ontology defines a class for each reference category, that is: Reference, External Support, Loci Paralleli, and Citation. Loci Paralleli and External Support are subclasses of Reference. Citation is a subclass of Loci Paralleli. Reference is defined as a subclass of E73 Information Object.

The taxonomy of the classes seen so far is shown in Fig. 1 (the ontology classes in light blue and the subclass relations in blue).

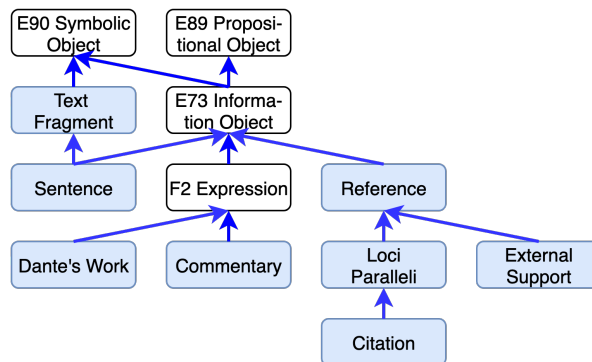


Figure 1: the main classes of the HDN ontology

The citation types and subject-to-object relationships are individuals, that is, instances of the E55 Type class. To distinguish them, the ontology includes two subclasses of E55: Citation Type, having as instances: genericCitation, strictCitation and explicitCitation; Subject-Object Relationship, having as instances correction, extension, and confirmation.

The subject of the Citation can vary significantly, therefore we consider it as an instance of E1 CRM Entity, the most general class of CRM.

An instance of Text Fragment is linked to the part of the work to which it belongs by the property R15i is Fragment Of. A part of the work (e.g., a *canto*) is linked to the part that includes it (e.g., a *cantica*) by property P148 is component of.

The references have as elements the source, the subject, and the object. The links between these elements and the references they belong to are represented by the following properties: hasRefSource links a reference to its source, and it is a subproperty of P129i is subject of; hasRefSubject links a reference to its subject, and it is a subproperty of P106 is composed of; hasRefObject links a reference to its object, and it is a subproperty of P129 is about.

### Ontology Population Tool

In order to extract the knowledge contained in the commentaries on Dante's *Commedia* and represent it according to our ontological model, we are developing a tool that will be distributed to scholars to populate the ontology.

The commentaries that will be processed in the HDN project are a selection of the corpus digitized by the Dartmouth Dante Project [5]. The text of the commentaries was pre-processed to identify the notes and, for each note, the following elements: the body of the note, the fragment of Dante's text, the verse, the *canto* and the *cantica*.

Once the pre-processing has been performed, scholars will use the tool to populate the ontology with the information that cannot be extracted automatically, that is, the references found in each note and all their characteristics (type, function, content, etc.).

### 3 CONCLUSIONS AND OUTLOOK

The outcomes of Hypermedia Dante Network will be manifold.

First of all, our project will make meaningful information quickly and easily available to users despite the vastness and complexity of both Dante's *Commedia* and its commentaries. The knowledge base will be made available through a user-friendly web application, where different queries can be performed. For example, a scholar could be interested in retrieving information about the main sources for Dante's topography, or, on the other hand, about the most significant uses of a particular source. Not only will HDN provide the user with information about which authors and works are quoted by Dante and in what forms, but it will also distinguish the different ways in which commentaries have put forward sources to explain Dante's poem, thereby giving a more detailed account of what features of a given author or source were more important to the poet at a given phase of his reception. By doing so, it will help us deepen our understanding of both Dante's poem and the almost self-sufficient tradition of its commentaries.

Although such web application is mainly intended for scholarly use, it will also prove useful in familiarizing everyday readers with research on Dante's primary sources, and it will prompt new and unexpected explorations in the masterpiece of Italian literature. We intend to design a web application that is flexible, accessible, and usable by a wide public.

In the future, we plan to progressively enrich our knowledge base with new commentaries and new languages, in order to make knowledge on Dante's primary sources available to everyone. Moreover, we hope that our endeavour will inspire the development of intertextuality-based ontologies for other literary works, thereby extending the scope of our multifaceted approach to stylistic, linguistic, and thematic features of literary texts.

## REFERENCES

1. Bartalesi, Valentina, Carlo Meghini, and Daniele Metilli. 'A Conceptualisation of Narratives and Its Expression in the CRM'. *International Journal of Metadata, Semantics and Ontologies* 12, no. 1 (2017): 35–46.
2. Bartalesi, Valentina, Carlo Meghini, Daniele Metilli, Mirko Tavoni, and Paola Andriani. 'A Web Application for Exploring Primary Sources: The DanteSources Case Study'. *Digital Scholarship in the Humanities* 33, no. 4 (2018): 705–723.
3. Doerr, Martin. 'The CIDOC Conceptual Reference Module: An Ontological Approach to Semantic Interoperability of Metadata'. *AI Magazine* 24, no. 3 (2003): 75–75.
4. Doerr, Martin, Chryssoula Bekiari, Patrick LeBoeuf, and Bibliothèque nationale de France. 'FRBRoo, a Conceptual Model for Performing Arts'. In *2008 Annual Conference of CIDOC, Athens*, 15–18, 2008.
5. Hollander, Robert. 'The Princeton Dante Project'. *Humanist Studies & the Digital Age* 3, no. 1 (2013): 53–59.
6. Meghini, Carlo, Mirko Tavoni, and Michelangelo Zaccarello. 'Mapping the Knowledge of Dante Commentaries in the Digital Context: A Web Ontology Approach'. *Forthcoming Romanic Review* 111, no. 2 (2020).
7. Riva, Pat, and Maja Žumer. 'FRBRoo, the IFLA Library Reference Model, and Now LRMoo: A Circle of Development', 2017.
8. Shotton, David. 'CiTO, the Citation Typing Ontology'. *Journal of Biomedical Semantics* 1, no. S6 (2010).

# **Futuro antico. Applicazioni in AR per la creazione di never ending books**

Nicola Barbuti  
Università di Bari Aldo Moro – nicola.barbuti [at] uniba.it

## **SINTESI**

Il paper presenta i risultati della ricerca finalizzata alla progettazione e realizzazione di un prototipo innovativo di ARBook cartaceo, che integra ristampe anastatiche di qualità di libri antichi e di pregio con espansioni interattive in Realtà Aumentata (AR) fruibili tramite dispositivi mobili. La sperimentazione ha previsto l'acquisizione ottica ad alta definizione degli originali, la stampa del volume e l'applicazione sul contenuto testuale cartaceo di marker finalizzati alla fruizione interattiva di espansioni digitali in AR. Il prototipo evolve le tradizionali funzioni del libro a stampa, in quanto per la prima volta l'applicazione di marker per la fruizione di contenuti in realtà aumentata è stata sperimentata su testi anziché su immagini o su grafica, generando un potenziale never ending book in grado di intercettare l'interesse anche di ampie fasce di utenti nativi digitali.

## **PAROLE CHIAVE**

augmented reality, Arbook, never ending book

## **INTERVENTO**

## **INTERVENTO**

### **1 STATO DELL'ARTE**

Recenti analisi di settore hanno confermato che, nelle nuove generazioni, la lettura tende sempre più a ridursi, in parte anche a causa della Digital Trasformation [1].

Questo, nonostante a partire dai primi anni 2000 siano state avviate sperimentazioni di libri a stampa integrati con soluzioni digitali. Si è ipotizzato, infatti, che l'espansione con grafica, animazioni e audio in Augmented Reality (AR), trasformando la lettura in un'esperienza dinamica e coinvolgente, potesse favorire un ritorno di interesse per l'editoria cartacea.

La ricerca su applicazioni di AR ai contenuti librari è stata inizialmente condotta prevalentemente nell'ottica di progettare risorse che favorissero l'apprendimento di determinate categorie di fruitori, facilitando la comprensione dei contenuti bidimensionali stampati tramite la visualizzazione di oggetti tridimensionali che li espandessero.

Le prime sperimentazioni di testi integrati con contenuti in AR sono state implementate nel primo decennio del nuovo secolo nell'editoria di ambito matematico [3], biochimico [4] e dell'education [2], avviando di fatto la ricerca che ha poi portato alla definizione del modello di ARBook.

Una prima soluzione abbastanza evoluta di questo modello è stata studiata nel 2010 nell'ambito della ricerca su testi utilizzabili da soggetti affetti da sordità [6].

Tuttavia, tutte queste sperimentazioni erano basate su configurazioni di AR per PC desktop, risultando fortemente limitate nelle ricadute produttive da diverse criticità: costi elevati, scarso appeal dei prodotti, limitate capacità degli utenti. La svolta si è avuta con il passaggio a soluzioni di AR per dispositivi mobili, ben noti e ampiamente utilizzati dalle diverse fasce d'età delle nuove generazioni di nativi digitali.

Il modello di ARBook che ha avviato questa nuova fase della ricerca è stato il MagicBook, realizzato con funzioni di education ed edutainment per bambini e giovani in età scolare, che consentiva agli utenti di leggere il libro fisico come un normale testo e visualizzare le espansioni in AR delle immagini tramite palmare [5].

L'analisi di questo modello e delle sperimentazioni precedenti hanno rappresentato i presupposti metodologici dai quali ha tratto origine la ricerca che si descrive, finalizzata a sviluppare un pilot di AR-Book nel quale, per la prima volta, marker per l'AR sono stati applicati per espandere testo stampato anziché immagini o grafica.

## 2 METODOLOGIA

La ricerca, svolta in collaborazione con la D.A.BI.MUS. S.r.l., spin off dell'Università di Bari specializzata in progettazione e sviluppo di servizi di digitalizzazione e prodotti digitali per il patrimonio culturale, e con la casa editrice Quorum Italia S.r.l., si è focalizzata su edizioni anastatiche di testi antichi. L'interesse per quest'ambito così peculiare è stato generato dalla constatazione della scarsa conoscenza sulla produzione libraria di intellettuali e scienziati pugliesi dei secoli passati, a fronte di una loro partecipazione da protagonisti nei dibattiti culturali e scientifici delle diverse epoche.

Il pilot è stato progettato con la massima cura sia del design tipografico, sia dell'espansione dei contenuti testuali tramite applicazioni in AR, con l'obiettivo di aumentare l'esperienza di fruizione degli utenti e intercettare l'interesse di un pubblico anche giovane.

La metodologia di sperimentazione si è articolata in tre attività. La prima si è focalizzata sulla scelta dell'opera e sull'acquisizione ottica. È stata scelta una raccolta di poesie della prima metà dell'Ottocento, che rappresenta la più antica testimonianza conosciuta dell'originale idioma vernacolare barese<sup>1</sup>. Il volume è stato acquisito otticamente in formato BITMAP ad alta definizione per la stampa. Le immagini sono state sottoposte a post elaborazione lavorando sul contrasto tra testo e sfondo e sulla definizione del tratto grafico, al fine di favorire la migliore applicazione dei marker.

---

1 F. S. Abbrescia, *Le rime italiane e baresi*, Bari, F. Petruzzelli, 1848.

La seconda attività ha previsto il design tipografico (Figura 1) e la stampa del prototipo cartaceo sul quale applicare i marker AR. La qualità estetica del libro gioca un ruolo importante nell'impatto con l'utente, in quanto fattore che contribuisce a suscitare l'interesse e a stimolare l'interazione fisica con l'oggetto. A riguardo, si è ritenuto importante associare alla manipolazione classica dell'oggetto fisico modalità di utilizzo del dispositivo mobile basate su Tangible User Interface (TUI), con il dito quale elemento di puntamento per l'interazione con i contenuti in AR.

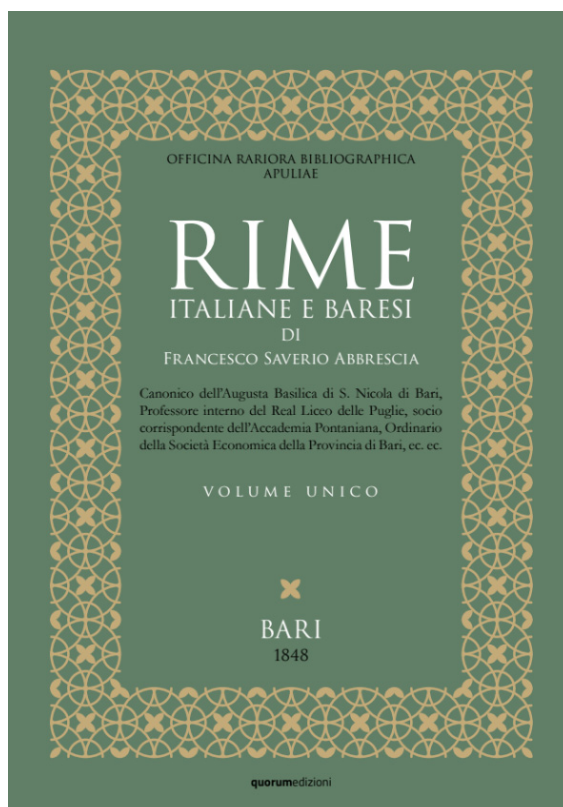


Figura 1 - Design della prima di copertina

La terza attività è stata incentrata sullo sviluppo dell'applicazione in AR. Sono state analizzate diverse webApp e per sviluppare la prima versione del pilot è stata scelta l'open source Blippar<sup>2</sup>, provvista dei tools necessari supportare il modello applicativo e utilizzabile con qualsiasi sistema per dispositivi mobili (iOS e Android). Attualmente si stanno valutando altre due soluzioni, delle quali si stanno studiando le funzionalità per implementare modalità di interazione con i contenuti in AR ancora più avanzate delle precedenti.

La configurazione dell'applicazione AR si è basata su quattro componenti principali: il dispositivo mobile per la fruizione dei contenuti in AR; la telecamera del dispositivo mobile per inquadrare e scansionare l'oggetto in tempo reale; il libro fisico; i marker per l'interazione tangibile e i target di tracciamento.

---

2 <https://www.blippar.com/>



Il design del modello ha integrato i concetti di pagina di storia e di pagina marker. La pagina di storia copre il contenuto delle pagine del volume. La pagina marker è composta dai marcatori che animano i contenuti in AR collegati a ciascuna pagina. Per visualizzarli e interagire, il prototipo utilizza il display dei dispositivi mobili che, puntato direttamente sulla pagina in lettura, la scansiona ed evidenzia l'interfaccia aumentata, consentendo all'utente di esplorarne i contenuti con tools dedicati e di fruirli mantenendo nel contempo intatta la normale lettura del libro (Figure 2 e 3).

Trattandosi di un testo ottocentesco comprendente poesie nell'antico vernacolo barese, le espansioni dei contenuti testuali consistono in un dizionario multimediale in formato PDF con i significati dei lemmi più arcaici, e nella lettura audio dei medesimi lemmi e di alcune poesie, al fine di far conoscere la corretta pronuncia dell'antico idioma barese, oggi quasi del tutto scomparso. I contenuti sono fruibili tramite lettura di testo e ascolto audio con lettura interattiva eseguita da un avatar digitale. Successive release espanderanno ulteriormente i contenuti, consentendo ai fruitori di aumentare la loro interazione durante la della lettura del testo.

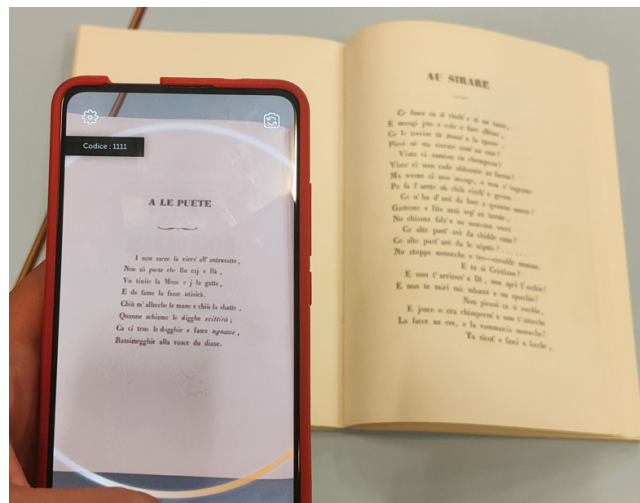


Figura 2 – Scansione della pagina

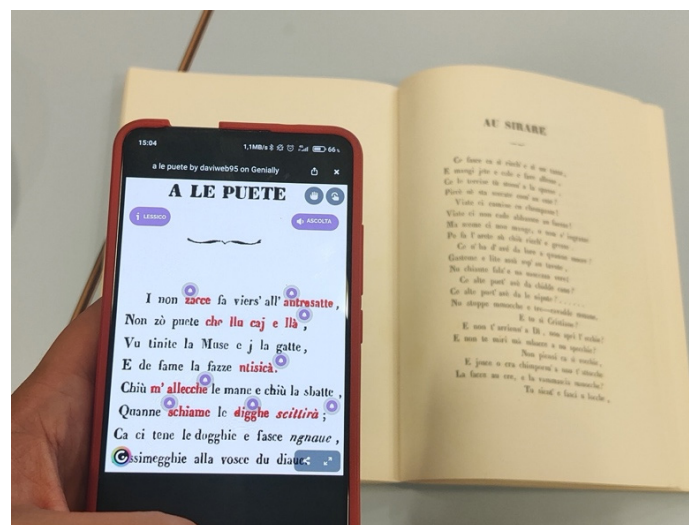


Figura 3 – Interazione con i contenuti testuali in AR

### 3 CONCLUSIONI

La sperimentazione ha mostrato come il libro possa assumere forme e funzioni del tutto nuove nell'interazione con gli utenti: entità non più statica e chiusa, ma dinamica, diacronica e fortemente interattiva, espandibile potenzialmente senza fine. I contenuti in AR, infatti, possono essere implementati con successive release, trasformando di fatto ciascuna edizione in un gate culturale costantemente in espansione.

Attualmente, la ricerca si è concentrata sulla biblioteca scolastica di un liceo pugliese, dove gli studenti partecipanti stanno sperimentando l'applicazione del modello descritto a volumi in essa conservati, al fine di renderli altrettanti never ending book, portali interattivi tramite cui accedere a espansioni in AR attinenti ai contenuti testuali.

### BIBLIOGRAFIA

1. Faggiolani, Chiara, e Alessandra Federici. «L'affordance della biblioteca. Modalità d'uso emergenti dalle indagini Istat». *Biblioteche Oggi* 3, n. 36 (2018).
2. Grasset, Raphaël, Andreas Dünser, e Mark Billinghurst. «Edutainment with a Mixed Reality book: a visual augmented illustrative children's book». *ACM Advance in computer entertainment technology*, 2008.
3. Kaufmann, Hannes, e Dieter Schmalstieg. «Mathematics and Geometry in Education with Collaborative Augmented Reality». In *ACM SIGGRAPH 2002 Conference*, 2002.
4. Medina, Eliana, Yu-Chien Chen, e Suzanne Weghorst. «Understanding biochemistry with Augmented Reality». In *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications*, 2007.
5. Tomia, Azfar Bin, Rohaya Dayang, e Awang Rambli. «An Interactive Mobile Augmented Reality Magical Playbook: Learning Number with the Thirsty Crow». *International Conference on Virtual and Augmented Reality in Education - Procedia Computer Science* 25 (2013).
6. Zainuddin, Norziha Megat Mohd, Halima Badioze Zaman, e Ahmad Azlina. «A participatory design in developing prototype an augmented reality book for deaf students». *IEEE Proceedings of the 2nd International Conference on Computer Research and Development*, 2010.



# Nuove voci digitali per incontrare Tolkien

Enrico Spadaro  
Eberhard Karls Universität Tübingen – henryspad7 [at] icloud.com

## PAROLE CHIAVE

Tolkien, Italia, traduzione, diffusione, social network, facebook, web-radio, condivisione culturale

## INTERVENTO

### 1 INTRODUZIONE

Lo scrittore inglese J.R.R. Tolkien è ormai uno degli autori più letti e influenti dal punto di vista sia accademico che commerciale. Negli ultimi anni, grazie alle tecnologie digitali, al web e alle reti sociali, sono nati nuovi strumenti di diffusione, discussione e studio che abbracciano le opere di Tolkien: nelle pagine seguenti, dopo una breve introduzione riguardo alla sua produzione letteraria, verranno presentati alcuni esempi attuali di diffusione digitale di Tolkien, in particolare quello riguardante il gruppo dei Tolkieniani Italiani.

### 2 AUTORE DEL SECOLO E OLTRE

Nel settembre del 2000, Thomas A. Shippey (n. 1943), accademico britannico e studioso di letteratura medievale, pubblicava *J.R.R. Tolkien: Author of the Century*, edito in traduzione italiana da Simonelli nel 2004 (*Tolkien Autore del Secolo*). Shippey, considerato uno dei più grandi esperti tolkieniani nel mondo, cerca di avvalorare la tesi per cui le storie fantastiche di Tolkien, a lungo considerate dalla critica “letteratura d’evasione”, meritino un posto di riguardo, al pari dei grandi autori del Novecento. L’autore trae spunto dal risultato di un sondaggio commissionato verso la fine degli anni novanta dalla catena inglese di librerie Waterstone che, secondo le preferenze di oltre 25.000 lettori, ha decretato *Il Signore degli Anelli* il miglior libro del XX secolo.

La popolarità delle opere di John Ronald Reuel Tolkien (1892-1973), in particolare de *Il Signore degli Anelli* (1954-55) e de *Lo Hobbit* (1937), ha sempre raggiunto livelli altissimi, fin dai tempi della loro pubblicazione: negli anni sessanta il professore di Oxford era stato eletto icona del movimento hippie e chiari riferimenti ai suoi testi compaiono nelle canzoni di gruppi come *Pink Floyd* o *Led Zeppelin*.

All’indomani dell’inizio del terzo millennio, gli adattamenti cinematografici di Peter Jackson – prima de *Il Signore degli Anelli* e poi de *Lo Hobbit* – hanno maggiormente contribuito ad assicurare a Tolkien un successo planetario. Nel 2019, Dome Karukoski ha diretto il biopic *Tolkien* che narra gli anni giovanili dello scrittore britannico tra Birmingham e Oxford, la partecipazione alla prima guerra mondiale e l’inizio della sua storia d’amore con Edith Bratt, che tanto segneranno la carriera accademica e letteraria di Tolkien.

Amazon sta inoltre girando una serie televisiva, che sarà disponibile online con il servizio streaming Prime Video, in cui saranno adattati gli eventi della Seconda Era del mondo, precedenti a quelli che costituiscono *Lo Hobbit* e *Il Signore degli Anelli*.

Questi due romanzi sono solo la punta di un grandissimo iceberg, che costituisce il cosiddetto *Legendarium* tolkieniano, vale a dire quello che per Tolkien doveva diventare una vera e propria mitologia per l'Inghilterra: era questo l'intento del giovane studioso di filologia e letteratura quando nelle trincee della prima guerra mondiale iniziò a scrivere "the book of Lost Tales". A partire dal 1917, Tolkien non smise mai di elaborare la propria mitologia, lasciando alla sua morte un corpus immenso di storie, racconti e leggende, di cui divenne erede letterario il terzo figlio Christopher. Questi si trovò di fronte ad una mole impressionante di scritti, appunti, elenchi, annali, bozze, opere incompiute e molteplici versioni delle stesse, e s'impegnò pertanto a dare una forma completa a questa mitologia: nel 1977, *Il Silmarillion* veniva dato alle stampe. Il lavoro di Christopher Tolkien non si arrestò e già nel 1980 era la volta di una nuova opera inedita, *Unfinished Tales (Racconti incompiuti)*, raccolta di testi che precisavano e ampliavano molte delle storie del *Legendarium*, cui seguiva, l'anno successivo, l'epistolario dell'autore (*La realtà in trasparenza. Lettere*, curato del suo biografo Humphrey Carpenter), che svelava la genesi e l'evoluzione delle idee e dei progetti di Tolkien, nonché i suoi rapporti a livello accademico e con editori, lettori e critici. Nel 1983 venne intrapresa la realizzazione di un'edizione critica degli scritti di Tolkien, includendo tutti gli appunti, le bozze e le varie versioni di ogni singolo racconto. Nacque così la *History of Middle-earth (HoME)*, costituita da dodici volumi (più uno di indici), pubblicati in serie dal 1983 al 1996. Oggi, dopo la morte di Christopher Tolkien nel gennaio 2020, la *Tolkien Estate* detiene i diritti sull'autore e numerose sono le società tolkieniane che si ritrovano in tutto il mondo: la più importante è la *Tolkien Society*, fondata nel 1969, in cui confluiscono studiosi e appassionati dai cinque continenti e che si occupa di diffondere la ricerca su Tolkien attraverso la pubblicazione – cartacea e digitale – di riviste quali *Mallorn* e *Amon Hen* e l'organizzazione annuale di eventi e congressi, l'ultimo dei quali, l'*Oxonmoot 2020*, si è tenuto in formato digitale su Zoom tra il 18 e il 20 settembre 2020.

### 3 TOLKIEN IN ITALIA

Il rapporto di Tolkien e delle sue opere con l'Italia è sempre stato e rimane ancora oggi travagliato, per motivi di ricezione nel nostro Paese, di resa traduttiva e d'interpretazione ideologica. Le ricerche effettuate da Oronzo Cilli nel suo *Tolkien e l'Italia* sottolineano l'evoluzione della prima diffusione del nostro autore nella penisola: all'indomani della pubblicazione di *The Fellowship of the Ring*, prima parte de *The Lord of the Rings*, Mondadori ebbe la possibilità di tradurre in italiano tale volume ma il progetto non andò avanti. Allo stesso modo nel 1962, Vittorini e Sereni, collaboratori dell'editore milanese, giudicarono inopportuna la pubblicazione, a causa dell'imperante neorealismo nel panorama letterario italiano.

Fu solo nel 1967 che l'editore Astrolabio-Ubaldini, affidando la traduzione del romanzo alla giovane Vittoria Alliata di Villafranca, propose sul mercato la prima edizione de *La Compagnia dell'Anello*. Le difficoltà economiche e l'iniziale insuccesso portarono la cessione dei diritti a Rusconi, che, grazie alle collaborazioni con Quirino Principe ed Elémire Zolla, pubblicò *Il Signore degli Anelli* in volume unico nel 1970.

Il marchio Rusconi tradusse progressivamente le varie opere di Tolkien, benché dei dodici volumi della HoME solo i primi due esistano ad oggi in versione italiana, fino al passaggio dei diritti a Bompiani nel 2000.

Anche nel nostro Paese l'opera di Tolkien ha acquisito lo status di best seller e numerosi sono gli studiosi che si occupano dello scrittore britannico, così come i fan che ne mantengono vivo il culto e le associazioni che curano le edizioni italiani ed organizzano eventi su tutto il territorio nazionali. La più antica e conosciuta rimane senza dubbio la Società Tolkieniana Italiana, ideata inizialmente nel 1992 e attiva dal 1994, il cui obiettivo primario è lo studio dell'Opera e del Pensiero di Tolkien. "A tale scopo la STI promuove in tutta Italia eventi e convegni che vedono la partecipazione di importanti studiosi [...] Molti sono i progetti condotti felicemente a termine, come la stesura del *Dizionario dell'universo tolkieniano* e del volume *I popoli di Tolkien*. [...] L'Associazione intende favorire anche il recupero e la diffusione dei testi tolkieniani che esulano dal corpus letterario della Terra di Mezzo"<sup>1</sup>.

Negli ultimi anni ha preso sempre più piede un'altra associazione che si occupa dello studio e della diffusione dell'opera di Tolkien in Italia, l'Associazione Italiana di Studi Tolkieniani (AIST). Nella pagina informativa del sito internet si legge: "l'AIST non è interessata a una "via italiana a Tolkien", bensì a un contributo italiano agli studi tolkieniani, collocandosi allo stesso livello delle società tolkieniane anglosassoni"<sup>2</sup>. Oltre ad aver curato alcune pubblicazioni riguardanti il mondo tolkieniano, tra cui *Tolkien e i classici* (Effatà Editrice, 2015), e un'ampia attività online sul sito<sup>3</sup> e sul corrispettivo gruppo Facebook, AIST si è prodigata affinché venisse intrapresa una nuova traduzione italiana de *Il Signore degli Anelli*. Bompiani ha affidato il lavoro a Ottavio Fatica e tra l'autunno del 2019 e agosto 2020 sono uscite in libreria le tre parti che costituiscono il romanzo, con il volume unico pubblicato il 28 ottobre 2020. AIST, così come dichiarato dal presidente Roberto Arduini in occasione dell'ultimo Salone del Libro di Torino svoltosi online, ha seguito la traduzione dal punto di vista filologico e scientifico.

La nuova traduzione è stata accolta con non poche critiche, anche frutto del paragone con la precedente, ma non è questa evidentemente la sede per discuterne: in ogni caso, il dibattito tolkieniano in Italia rimane ancora fervente e intenso e sorgono sempre più realtà che si interessano allo scrittore oxoniese.<sup>4</sup> Le nuove tecnologie digitali, soprattutto i social network, contribuiscono e svolgono un ruolo fondamentale in tal senso e nelle prossime sezioni sarà proprio una di queste realtà "tolkieniane" ad essere presa in esame.

---

1 <https://www.tolkien.it/la-societa-tolkieniana-italiana/attivita/>

2 <https://www.jrrtolkien.it/about/>

3 [www.jrrtolkien.it](http://www.jrrtolkien.it)

4 A tal proposito AIST e l'Università di Trento hanno organizzato un convegno online sulla piattaforma Zoom, "Fallire sempre meglio: tradurre Tolkien, Tolkien traduttore".

<https://webmagazine.unitn.it/evento/lettere/84413/fallire-sempre-meglio-tradurre-tolkien-tolkien-traduttore>

## 4 I TOLKENIANI ITALIANI

Il progetto dei Tolkieniani Italiani prende vita inizialmente il lunedì di Pasquetta del 2018 da un'idea di Giuseppe Scattolini, studioso di Tolkien e fondatore del gruppo dei “Cavalieri del Mark”, con l'intento di raggruppare – in maniera soprattutto digitale – le diverse realtà italiane che si occupano di Tolkien. Come si legge nella descrizione del gruppo Facebook «Tolkieniani Italiani – gruppo pubblico»<sup>5</sup>, questo “nasce grazie alla condivisione di risorse e piattaforme che abbiamo messo in atto come rete di associazioni, gruppi, semplici appassionati e studiosi”. Il “fare rete”, termine ampiamente in voga oggigiorno, è uno degli elementi essenziali dei Tolkieniani Italiani.

Come detto in precedenza, esistono oramai un gran numero di associazioni, gruppi di appassionati e singoli studiosi che s'interessano a Tolkien e alla sua opera e affinché ciascuno di questi non perda la propria specificità o indipendenza, nell'ottica di una cultura condivisa e aperta, i Tolkieniani Italiani non si configurano come un'associazione ben specifica che persegue un fine e necessita un'iscrizione o tesseramento, bensì come un'idea che abbraccia le realtà tolkieniane che vi aderiscono. In tal modo si elimina qualunque principio d'ideologia o d'intento politico, con il semplice obiettivo di discutere riguardo a Tolkien, un autore, peraltro, che sottolinea di prediligere “l'applicabilità” della propria opera, secondo la libertà del lettore cui non deve essere imposta l'allegoria prefissa dallo scrittore [8]. Si tratta pertanto di un aspetto unico nel panorama italiano tolkieniano, in cui negli anni le ideologie politiche hanno influito non poco: tutto questo è reso possibile grazie alla libertà d'espressione che la rete internet concede ai suoi utenti e che si attua appieno all'interno dei Tolkieniani Italiani, dove la condivisione del sapere tolkieniano la fa da padrona.

Il gruppo è animato da principi fondamentali, quali la centralità assoluta dell'opera letteraria di Tolkien, l'attenzione alla critica e all'interpretazione degli scritti dello stesso, pertanto gli esperti e gli accademici presenti nel gruppo si prefiggono il ruolo di custodi della realtà dei Tolkieniani Italiani. La “divulgazione” scientifica riveste dunque un ruolo di primo piano, ma è pensata diversamente, senza una netta separazione tra “alti studi” e “divulgazione” vera e propria: tutti possono esserne introdotti e devono far parte, in qualche modo, degli studi più avanzati. Non si tratta di una semplice partecipazione agli incontri pensati più “per il popolo” (Tolkien e il fantasy, Tolkien e il Trono di Spade, ed altri elementi molto pop), ma una condivisione vera e propria che permetta l'arricchimento collettivo, perché tutti hanno desiderio di una conoscenza più approfondita della realtà e possono arrivare a comprendere la verità, se presentata in modo non specialistico.

---

5 <https://www.facebook.com/groups/235908023657984/announcements>

Tutto questo avviene, su scala nazionale, all'interno del gruppo pubblico su Facebook, sfruttando quindi una delle funzioni più interessanti della rete sociale più famosa al mondo: all'inizio si è cercato di raccogliere sempre più persone e realtà, invitando a collaborare tutte le pagine di Facebook che trattavano di Tolkien. La pagina che ha risposto più entusiasticamente e con cui è nata un'importante collaborazione nonché una bella amicizia è quella di «Pedo Mellon a Minno»<sup>6</sup> e da lì se ne sono aggiunte tante altre tra cui le più importanti, che superano i diecimila follower, sono «Tolkien Italia»<sup>7</sup> e «Le Migliori Frasi del Silmarillion»<sup>8</sup>. Le pagine si collegano al gruppo pubblico portandovi i loro seguaci, poi condividendo il proprio materiale sul gruppo oltre che recuperare i propri, che magari sulla pagina sono quiescenti. Si entra così in contatto con il pubblico di altre pagine, incrementando il proprio, puntando alla creazione di una rete nazionale in cui il sapere è reso disponibile a tutti. Il gruppo pubblico dei «Cavalieri del Mark», che poi si è modificato nel gruppo dei «Tolkieniani Italiani», è diventato e rimane un punto di incontro per tutti, di condivisione dei follower e di comunione per tutti i Tolkieniani. La rete su Facebook si fonda specificatamente sul flusso e riflusso delle persone di passaggio sul gruppo pubblico e la pubblicazione coordinata di notizie, eventi e articoli sulle pagine, che mantengono la loro specificità ma che sono comunque tutte in linea a uno stesso obiettivo e tentano di non accavallarsi mai. I Tolkieniani Italiani nascono per coordinare e permettere la collaborazione di realtà già esistenti, quindi si può affermare che qualunque cosa venga organizzata da chi fa parte della rete, che siano i «Cavalieri del Mark», la Società Tolkieniana Italiana o altri gruppi ad essa afferenti, è come se fosse organizzata interamente dai Tolkieniani Italiani. Le visualizzazioni e gli accessi al gruppo sono passati dai 5866 del 31/12/2019 ai 9729 del 22/11/2020, con una media di circa duecento interazioni (“mi piace”, “commenti” e “condivisioni”) giornaliera. Inoltre, stando ai dati forniti da Greta Bertani, una delle responsabili dei Tolkieniani Italiani, il numero degli iscritti al gruppo, monitorato costantemente, era di 1.798 il 24/8/2018, mentre al 24/11/2020 essi corrispondono a 12.762, con un incremento del 709,79%, di cui 149,28% nel solo 2020. Per quanto riguarda le principali pagine afferenti al gruppo è possibile constatare un costante aumento nel numero dei seguaci: la pagina «I Cavalieri del Mark» è passata da 4.745 il 25/8/2017 a 17.913 il 20/11/2020, con un incremento del 377,51%<sup>9</sup>; «Tolkien Italia» dai 14.486 del 20/8/2017 ai 26.379 del 20/11/2020 (182,10%)<sup>10</sup>; «Le Migliori Frasi del Silmarillion» da 15.364 il 27/11/2018 a 25.107 il 16/11/2020 (163,41%)<sup>11</sup>.

---

6 <https://www.facebook.com/StoriediTolkien>

7 <https://www.facebook.com/TolkienItalia>

8 <https://www.facebook.com/Silmarillionmigliorifras>

9 <https://www.facebook.com/tolkiennellemarche>

10 <https://www.facebook.com/TolkienItalia>

11 <https://www.facebook.com/Silmarillionmigliorifras>



Al gruppo Facebook, che rappresenta un uso ottimale del social network, si aggiunge anche il sito web «Tolkien Italia»<sup>12</sup>, in cui si ritrovano le notizie sul mondo di Tolkien sia dall'Italia che dal resto del mondo, si discute sulle nuove uscite nell'ambito della ricerca tolkieniana, con ampio spazio dedicato alla nuova traduzione de *Il Signore degli Anelli* da parte di Ottavio Fatica; è inoltre presente un corso dettagliato di lingue elfiche, organizzato da Gainluca Comastri, che aderisce anch'egli al progetto dei Tolkieniani Italiani.

Si punta in futuro all'organizzazione di eventi culturali, meeting e congressi con pubblicazioni scientifiche che facciano seguito alle varie iniziative che caratterizzano tale progetto. Una di queste è stata senz'altro il convegno “Barlumi di cose più alte, più profonde o più oscure della sua superficie”, tenutosi presso l'Università di Macerata tra il 4 e il 6 dicembre 2019 con l'intento di ampliare la rete dei Tolkieniani Italiani non solo a docenti universitari ma anche a tutta la rete delle università. A Macerata si è realizzato solo il primo passo grazie all'intervento di accademici o di diversi esperti di Tolkien, quali Gianluca Comastri, Luisa Paglieri, Greta Bertani, Costanza Bonelli, Francesca Montemagno, o ancora Oronzo Cilli, il cui ultimo libro *Tolkien's Library* è stato premiato come miglior pubblicazione ai «Tolkien Society Awards 2020». L'evento di Macerata ha anche beneficiato di una diretta sul web grazie a quella che è forse l'iniziativa più interessante del progetto dei Tolkieniani Italiani, soprattutto a livello digitale, vale a dire la creazione e la gestione di una web-radio, «La Voce di Arda».

## 5 RADIO LA VOCE DI ARDA

Nelle opere di Tolkien, Arda è il mondo in cui si svolge ogni evento: pertanto, la radio «La Voce di Arda» intende comunicare tutto quello che riguarda l'universo tolkieniano. Si tratta di una stazione radiofonica digitale, i cui organizzatori, speaker e animatori sono Giuseppe Scattolini e Simone Claudiani, e che, anche grazie alla collaborazione con la Società Tolkieniana Italiana, è diventata a tutti gli effetti il canale preferenziale di comunicazione ad ampio raggio del gruppo dei Tolkieniani Italiani. Essendo una web-radio, è possibile ascoltare le puntate, in cui si trattano esclusivamente le opere e la poetica di Tolkien, sulla piattaforma «Spreaker», un servizio di distribuzione di podcasting attivo in Italia dal 2010. La radio è in onda in diretta i primi tre venerdì di ogni mese alle ore 21.00 salvo qualche eccezione, e ogni puntata, della durata di tre ore, è caratterizzata dalla lettura e da commenti di passi scelti attraverso sondaggi nel gruppo Facebook, dalla discussione riguardo a pubblicazioni e ricerche e coinvolge studiosi, appassionati ed esperti del mondo tolkieniano. Benché non si tratti del primo tentativo di “radio tolkieniana” – ricordiamo il progetto conclusosi di “Radio Brea di Sentieri Tolkieniani” –, è lo strumento digitale di discussione e condivisione tolkieniana che sta riscuotendo maggior successo nel nostro Paese. La piattaforma Spreaker consente d'intervenire in diretta durante le puntate attraverso la funzione dei commenti, in modo tale che gli ascoltatori possano fornire la propria opinione istantaneamente su quello che viene detto durante la puntata e diventare così parte attiva della radio. S'instaura in tal modo una prospettiva di dialogo continuo tra gli animatori radiofonici, gli ospiti in diretta – che sono spesso grandi esperti o studiosi affermati di Tolkien – e ogni singolo utente all'ascolto, realizzando quell'intento di condivisione del sapere che è tra i punti cardine del progetto dei Tolkieniani Italiani.

---

12 <https://tolkienitalia.net>

Una volta terminata la puntata, è possibile ascoltarla in differita sia su Spreaker sia su Spotify, nella sezione «Podcast», così da poter riprendere e apprezzare ancora il contenuto di ogni singolo episodio; si ritrovano ancora online<sup>13</sup> i titoli di tutti gli episodi andati in onda finora e, tra questi, si potrebbe ricordare il corso di lingue elfiche di cui è possibile recuperare una completa documentazione su «Tolkien Italia»<sup>14</sup>. Ampio spazio è dedicato alla lettura di passi de *Il Signore degli Anelli*, de *Lo Hobbit* o degli scritti che costituiscono la mitologia di Tolkien raccolta nelle diverse fasi della sua vita. Ad essi si aggiungono le discussioni sulle recenti pubblicazioni italiane concernenti l'opera del professore di Oxford, come *Colui che raccontò la grazia* di Mauro Toninelli, l'approfondimento di tematiche presenti in Tolkien ma che hanno risvolti anche sulla realtà contemporanea e una critica a trecentosessanta gradi sulla nuova traduzione de *Il Signore degli Anelli* da parte di Ottavio Fatica.

Il periodo di lockdown, in cui molti italiani sono stati costretti a restare in casa a causa della pandemia di Coronavirus, ha favorito la diffusione e l'aumento degli ascolti della radio: sempre più utenti hanno preferito passare i propri venerdì sera in compagnia di Tolkien e probabilmente di quel messaggio positivo e provvidenziale che traspare nei suoi romanzi; non a caso la parola “hope”, “speranza/sperare”, è una delle più ricorrenti tra le pagine de *Il Signore degli Anelli*, ben 74 volte solo nel libro I. Il ritorno alla speranza è forse qualcosa che molti hanno riscoperto durante gli ultimi difficili mesi, ed è importante sottolineare come questa speranza sia stata comunicata attraverso un canale digitale: la rete è spesso bistrattata in quanto carica di odio e messaggi negativi, per cui è fondamentale rimarcare quando brillano luci positive foriere di messaggi nuovi e diversi.

Grazie ad un progetto ben strutturato, ad una costanza nell'impegno e nella collaborazione, la popolarità della radio ha raggiunto livelli forse insperati inizialmente, dai soli 419 ascolti del 22 marzo 2019 fino al picco – circa 250 mila –, registratosi lo scorso 19 giugno in occasione della partecipazione in diretta di Vittoria Alliata di Villafranca, la prima traduttrice italiana di Tolkien, che, sul finire degli anni '60, da adolescente aveva intrapreso con la macchina da scrivere e semplici dizionari la resa in italiano de *Il Signore degli Anelli*. Le ambizioni della radio non si fermano di certo qui e sono in cantiere altri progetti, tra cui su tutti un corso di lingua inglese attraverso la lettura e l'analisi dell'opera di Tolkien<sup>15</sup>, l'esplorazione dei suoi scritti iniziali o il paragone tra il nostro autore e altri classici della letteratura occidentale.

Come detto, il fine ultimo dei Tolkieniani Italiani rimane l'organizzazione di eventi culturali e la pubblicazione scientifica riguardo a Tolkien, anche grazie alla collaborazione attiva con la Società Tolkieniana Italiana – riconosciuta dalla Tolkien Estate, che custodisce i diritti dell'opera di Tolkien. Nondimeno, la diffusione sui social network, primo fra tutti Facebook nella forma del gruppo pubblico, ma anche le rispettive pagine Instagram che stanno sempre più prendendo piede, permette un accesso libero e illimitato da parte di qualsiasi utente, il quale può divenire parte attiva del progetto.

Grande importanza viene data all'esperienza e a come si vive personalmente, individualmente e comunitariamente Tolkien, che può essere definito a tutti gli effetti un padre spirituale per tutti i Tolkieniani, o come lo definisce Tom Shippey “l'autore del XX secolo” [5].

---

13 <https://www.spreaker.com/show/lo-show-di-radio-la-voce-di-arda>

14 <https://tolkienitalia.net>

15 <https://www.facebook.com/CorsoLetturaTolkienInglese>

Le sue idee, i suoi messaggi, gli studi, le interpretazioni diventano il fulcro di una nuova comunicazione digitale, che attraverso molteplici forme diffonde vecchie e nuove idee, permette un dialogo più ampio e una fruizione totalmente condivisa e partecipata del patrimonio culturale.

## BIBLIOGRAFIA

1. M. Toninelli. *Colui Che Raccontò La Grazia*. Assisi: Cittadella editrice, 2019.
2. O. Cilli. *Tolkien e l'Italia – Il Mio Viaggio in Italia*. Rimini: Il Cerchio, 2017.
3. ———. *Tolkien's Library. An Annotated Checklist*. Edinburgh: Luna Press, 2019.
4. S. Giuliano. *J.R.R. Tolkien. Tradizione e Modernità Nel Signore Degli Anelli*. Milano: Edizioni Bietti, 2013.
5. Shippey, Tom. *J.R.R. Tolkien: Author of the Century*. London: Mariner Books, 2000.
6. T. Shippey. *J. R. R. Tolkien: Author of the Century*. Londra: Mariner Books, 2000.
7. Tolkien. *Lo Hobbit Annotato*. Milano: Bompiano, 2012.
8. Tolkien, J.R.R. *Il Signore Degli Anelli*. Milano: Bompiani, 2000.
9. ———. *Il Silmarillion*. Milano: Bompiani, 2000.
10. ———. *The History of Middle-Earth (12 Vol. Set)*. Vol. 1-12. Londra: Harper Collins, 2014.
11. ———. *The Lord of the Rings*. London: Harper Collins, 2016.
12. Wu Ming 4. *Difendere La Terra Di Mezzo*. Bologna: Odoya, 2013.

## ULTERIORI RIFERIMENTI

[www.facebook.com/groups/235908023657984/announcements](http://www.facebook.com/groups/235908023657984/announcements)

[www.facebook.com/Silmarillionmigliorifrase](http://www.facebook.com/Silmarillionmigliorifrase)

[www.facebook.com/StoriediTolkien](http://www.facebook.com/StoriediTolkien)

[www.facebook.com/TolkienItalia](http://www.facebook.com/TolkienItalia)

[www.facebook.com/tolkiennellemarche](http://www.facebook.com/tolkiennellemarche)

[www.jrrtolkien.it/](http://www.jrrtolkien.it/)

[www.spreaker.com/show/lo-show-di-radio-la-voce-di-arda](http://www.spreaker.com/show/lo-show-di-radio-la-voce-di-arda)

<https://tolkieniano.blogspot.com>

<https://tolkienitalia.net>

[www.tolkien.it/la-societa-tolkieniana-italiana](http://www.tolkien.it/la-societa-tolkieniana-italiana)

[www.tolkiensociety.org](http://www.tolkiensociety.org)

<https://webmagazine.unitn.it/evento/lettere/84413/fallire-sempre-meglio-tradurre-tolkien-tolkien-traduttore>

# A Literary GIS of Trentino: Opportunities for Territorial Enhancement from Geographic Research and Literature

Elena Dai Prà<sup>1</sup>, Nicola Gabellieri<sup>2</sup>, Giannantonio Scaglione<sup>3</sup>

<sup>1</sup> University of Trento - elena.daipra [at] unitn.it

<sup>2</sup> University of Trento - nicola.gabellieri [at] unitn.it

<sup>3</sup> University of Trento - g.scaglione [at] unitn.it

## ABSTRACT

This contribution deals with a project that is currently being developed, which aims at creating a Literary GIS that will collect excerpts from literary works describing travelling experiences in the territory of the Province of Trento. The application of Geographic Information Systems (GIS) technologies to literary and geo-historical research has established itself as one of the most dynamic fields of development of Digital Humanities. However, in Italy, the use of GIS is still limited. The cataloguing, transcription and mapping of extracts from texts written by travellers who have crossed the Adige Valley and described it between the 17<sup>th</sup> and the 20<sup>th</sup> century will enable the development of a specific database and the integration of geolocation data using GIS software. The selected passages, ranging from travel reports to the first *ante litteram* tourist guides, can offer a series of clues on the social, environmental, landscape-related and territorial context of the past centuries; they also allow us to reflect on the perception and subsequent narration of the visited areas by foreign travellers. At the same time, the database can become the foundation of several activities of public dissemination, both in the field of education and in that of tourism enhancement. A web-GIS database could support the creation of literary and cultural itineraries, following the footsteps of famous writers and rediscovering their narration of past landscapes in Trentino.

## KEYWORDS

literary GIS, travel literature, historical geography, literary geography, cultural tourism

# TALK

## 1 TITLE

The use of literary sources in geography and historical geography dates to the dawn of the discipline ([19]). As early as in 1844, in his work *Cosmos: essai d'une description physique du monde*, Alexandre von Humboldt acknowledged literature as a privileged *medium* to discover the traces and the spirit of the relationship between men and space ([17]). Such field of investigation, generically defined as “literary geographies”, has progressively established itself since the 1960s, especially thanks to geographic studies addressing the creation and the nature of the concept of region. Literary descriptions were seen as a crucial instrument in the representation of geographic uniqueness and of the original historical traits of given territories (the so-called *genius loci*).

In the last decades, mostly because of the opposite and converging “cultural turn” in geographical studies and “spatial turn” in literature, the field of literary geographies has grown, investigating a series of different yet intertwined objects. *Inter alia*, we recall the use of texts as geohistorical sources, which, combined with other forms of documentation, support research of past landscapes and territories; literary representations as indicators of their authors’ perception of geographical objects; the construction of images and territorial stereotypes that serve as literary *topoi*; the role of landscape, territory and place in many literary plots ([1] [14] [3] [21] [13]).

In this context, travel literature, which includes travel reports, travellers’ private diaries, current or *ante litteram* tourist guides, travel novels, and even some ethnographic and demological pieces, qualifies as a porous yet well-defined line of research, which deserves to be considered for its heuristic potential in the study of past geographies ([28]).

Since the 1990s, said studies have found new areas of application in the fields of “public” and “applied” geography, in several initiatives that were aimed at spreading the knowledge of territories and landscapes and at enhancing them. Literary parks, cultural itineraries and landscape observatories have used literary narrations to increase the territorial awareness of local inhabitants, to encourage the rediscovery of areas that were only marginally affected by tourist flows, or to add value to already well-established tourist destinations ([26] [21]).

## 2 TITLE

Since the 1990s, Geographic Information Systems (GIS) have progressively asserted themselves as one of the most effective instruments for data management and analysis in all disciplines interested in studying spatial and territorial phenomena ([31] [9]). A GIS is a set of tools that are used to collect, store, analyse and visualise spatial data for scientific, management and information purposes ([4]). Starting from a software environment that enables the elaboration of data that are organised into Data Management Systems (DBMS) information structures, a geodatabase can be created, i.e. a database in which spatial and geographic data are stored, managed and interrogated.

It can host different types of spatial data, such as vectors and rasters with alphanumeric attributes and geographic coordinates ([5]). The set of instruments, structures, human resources and procedures that enable the acquisition, analysis and consultation of data has been defined as Territorial Information System (TIS) ([22]).

TIS has been applied to all fields concerned with the study, management and control of a territory. However, since the arrival of 2.0 web forms, researchers have started to also consider their communicative potential. Indeed, they put their efforts into identifying “forms of communication increasing the dissemination of research contents and enabling an opening towards new scientific horizons” ([12]). Joe Smith has highlighted the usefulness of Digital Geography, conceived as a set of practices «that might dramatically improve the conditions for the conduct, reception and debate of our geographical work» ([29]). Geographic Information Systems can represent an effective tool to inventory digital bibliographic archives and make them accessible; researchers and the general public would be able to use location and a cartographic interface to identify all sources linked to a given portion of space.

Especially in geohistorical, historical and archaeological research, the application of the GIS software has long been destined mainly to the digitalisation and analysis of cartographic and aero-photogrammetric sources, or as a tool for managing land surveys ([65]). The focus on textual sources has been far more limited, mostly because many disciplines, such as literature and cultural geography, do not often rely on quantitative and topographically located data as are those processed by GIS. Only since the 2000s, and particularly thanks to the epistemological proposals by Moretti and to the experiments carried out in the Anglo-Saxon world, have literary studies started to consider GIS as a heuristic instrument ([23] [24]). These preliminary ideas have recently been extended to interdisciplinary pioneer projects that envisage the application of territorial information systems to literature, within a heterogeneous field that has been defined as Literary GIS ([8]). The project “Mapping the Lakes: A Literary GIS” is an example; it has led to structuring a geodatabase that collects numerous historical travel reports describing the Lake District, a mountainous region in north-western England, traditionally associated with English literature and art. Such database has allowed researchers to explore and experience the imaginary and affective geographies of English writers and readers, thanks to the digitalisation and analysis of *corpora* of several pieces of work ([7] [18]). Similar experiences involve the “The digital Periegesis” project, that aims to map and analyse ancient Greek sources as the works of Pausania of Magnesia in order to identify itineraries and networks, as well as the works of Rachele Sprugnoli on computational methods and technologies applied to the treatment of historical travel texts in Italy.

Despite the rapid spread of Literary GIS and the increase of similar initiatives, this line of research is still largely unexplored in Italy.

### 3 TITLE

The paper presents a project for the implementation of a literary GIS that will collect reports by travelers who have visited the Trentino region between the 17th century and the beginning of the 20th century. Such time frame has been selected because it corresponds with the spread of *ante litteram* forms of travel and tourism (mostly linked to the Grand Tour) which saw the production of both published and unpublished memoirs. These are strongly linked to the representation of society, landscape and territory. The choice of the territorial case study also responds to precise motivations: the Adige Valley has long been one of the main gateways to the Italian Peninsula for numerous north-European travel-

lers who wanted to visit Italy's major cities for religious or cultural purposes ([1] [2] [27]). Such flow has led to the production of a large literary *corpus*, which includes diaries, mail correspondence, reports, travel guides. These texts highlight new elements of the alpine geography of the past, while reflecting the interpreting categories, the expectations and perception of the travellers themselves ([28]). In order to implement a few of the methodological proposals put forward in the Anglo-Saxon environment, some texts were chosen, digitalised and transcribed.

In this first stage of the project, a number of 12 texts (produced between the 16th and the 19th century) has been selected. In order to develop the methodology, the research involved edited books available in the local libraries, addressing writers of different language (English, French and German) and texts produced for different purposes (travel accounts for publishing, private travel accounts that have been published later, tourist guides).

The selection of texts with descriptions of landscapes and territories that could be located led to the creation of a digitised cataloguing system.

Useful information for the bibliographical, historical and geographic identification of said texts were included: the author's name, the title, the date of publication, the time of travel, the author's origin, the type of text (report, diary, letter, guide), a transcription of the text in its original language and – where necessary – its Italian translation, toponyms. Subsequently, each item in the dataset was geolocated in a GIS environment and linked to the centroid of the area described in the text. Geolocalization has been made manually, on the basis of mentioned place names and described areas. Such operation inevitably entails a certain degree of approximation, given the different nature of the descriptions, whose objects can be both well-defined territorial elements, like a building, and larger portions of space, like a valley.

The result is a geodatabase through which it is possible to: visualise the most described areas and the routes followed by travellers; identify different narrations of the same sites to compare them; carry out queries using given keywords such as adjectives related to perception (e.g. "sublime or divine", which are reading categories that have captured a lot of attention in literary sciences) and specific material elements of territory and landscape (urban elements, cultivation practices, viability and hospitality structures).

For instance, Figure 1 represents a density map of the different punctual elements recorded in the dataset. The image clearly represents the routes narrated by travellers; these moved from the Brenner Pass to the Padan Plain (*Pianura Padana*), mostly crossing the bottom of the valleys. Many descriptions are located in Trento, or in Torbole or Riva, on Lake Garda. This testifies the success that this area had and the interest it sparked in German travellers such as Goethe.

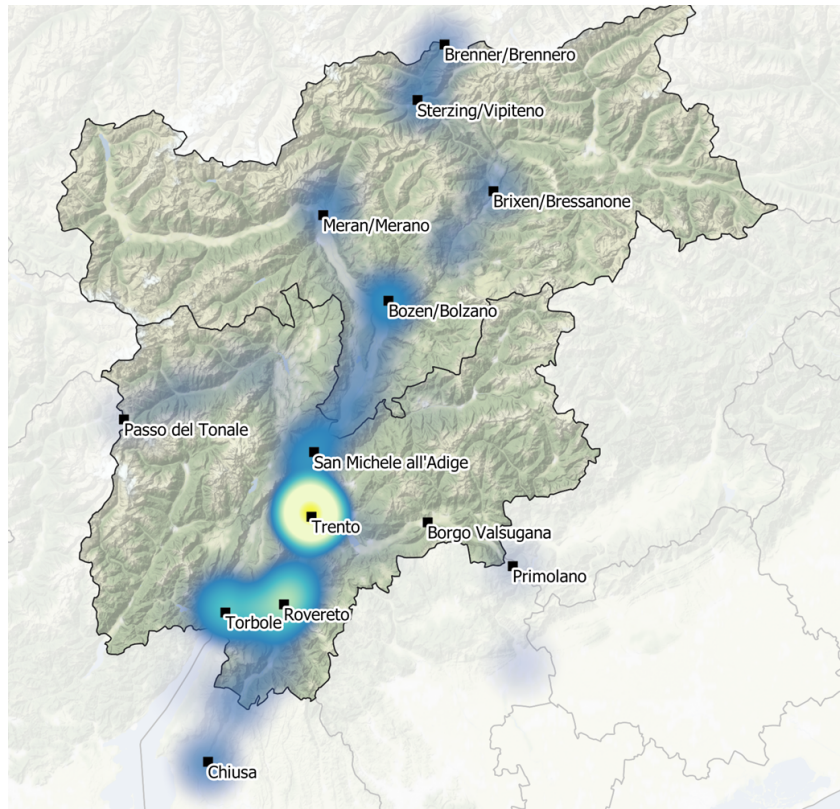


Figure 1. Density map made with the QGIS software representing the descriptions of landscapes and territories recorded in the geodatabase on travel literature in Trentino

## 4 FROM LITERARY GIS TO TOURISM ENHANCEMENT: POSSIBLE IMPLEMENTATIONS

Several decades after the creation of the first “Literary parks”, stating that literary sources can be a crucial instrument for the enhancement of certain territories in terms of tourism seems tautological. Literary parks, just like cultural routes, have been created with the aim of encouraging the discovery of more or less well-known landscapes. This can be achieved by supporting the fruition and interpretation of such landscapes through the descriptions by famous or lesser-known authors who have written about them in the past ([26] [13]). In this respect, literary works become a narrative plot that should attract and guide visitors while also promoting “green” and sustainable cultural tourism. However, as mere storytelling does not suffice, focusing on critical and conscious interpretations of local space, its transformation and its cultural values are fundamental ([21]).

In this context, the project aims to use travel literature of the past as a key to access and read the territory of Trentino. The works by authors who wrote about this area will become a *medium* to highlight relevant aspects of local culture and history, following two directions. On the one hand, the aim is to safeguard the works and ideas of poets and writers who are physically linked to a place; on the other hand, the work wants to promote and enhance local material and immaterial cultural heritage, thanks to the creation of cultural theme-based routes.



Accordingly, a location-based database for the production of cartography, to be accessed through web-GIS ([15]), can be a powerful tool for highlighting cultural and environmental resources and improving the fruition of a given territory ([25]). By locating the descriptions of landscapes, cultural tourism routes can be established, in order to facilitate the recognition and interpretation of territorial heritage and perpetuate the value intellectuals of the past attributed to certain places. This is the case of a cultural itinerary dedicated to Goethe, who, as Dai Prà writes, “leads us [...] in his journey towards the concept of landscape as an image and representation that is capable of reconciling the mind’s different skills through a sort of superior contemplation and pleasure, which will make him write at the beginning of his work: ‘Et in Arcadia ego’ (I have been to the land of happiness and beauty, too)” ([10]).

## REFERENCES

1. Black, Jeremy. ‘Italy and the Grand Tour: The British Experience in the Eighteenth Century’. *Annali d’Italianistica* 14 (1996): 532–541.
2. Brillì, Attilio. *Dalla Satira Alla Caricatura: Storia, Tecniche e Ideologie Della Rappresentazione*. Vol. 27. Edizioni Dedalo, 1985.
3. Brosseau, Marc. ‘In, of, out, with, and through: New Perspectives in Literary Geography’. In *The Routledge Handbook of Literature and Space*, 9–27. Routledge, 2017.
4. Burrough, Peter A. ‘Principles of Geographical’. *Information Systems for Land Resource Assessment*. Clarendon Press, Oxford, 1986.
5. Burrough, Peter A., Rachael McDonnell, Rachael A. McDonnell, and Christopher D. Lloyd. *Principles of Geographical Information Systems*. Oxford university press, 2015.
6. Chevalier, Michel. ‘Géographie et Littérature’. *Société de Géographie* 1 (2001): 260.
7. Cooper, David, Christopher Donaldson, and Patricia Murrieta-Flores. *Literary Mapping in the Digital Age*. Routledge, 2016.
8. Cooper, David, and Ian N. Gregory. ‘Mapping the English Lake District: A Literary GIS’. *Transactions of the Institute of British Geographers* 36, no. 1 (2011): 89–108.
9. Cope, Meghan, and Sarah Elwood. *Qualitative GIS: A Mixed Methods Approach*. Sage, 2009.
10. Dai Prà, Elena, and Franco Salvadori. ‘Il Viaggio in Italia Di Goethe: Ontologia Del Paesaggio Nel Solco Della Tradizione Speculativa Geografica (e Non Solo) Tedesca?’ *L’apporto Della Geografia Tra Rivoluzioni e Riforme*, 2019, 617–20.
11. Foka, Anna, Elton Baker, Konstantinidou, Mostofian Kyriaki, Demiroglu Nasrin, O. Cenk, Brady Kiesling, and Linda Talatas. ‘Semantically Geo-Annotating an Ancient Greek “Travel Guide” Itineraries, Chronotopes, Networks, and Linked Data’. In *Proceedings of the 4th ACM SIGSPATIAL Workshop on Geospatial Humanities*, 1–9. New York: Association for Computing Machinery (ACM), New York, NY, USA, 2020.
12. Forino, Giuseppe, Luca Salvati, Claudio Sesto, and Rossella Belluso. ‘Le Riviste Scientifiche Di Geografia (Roma, 8 Luglio 2009)’. *Semestrale Di Studi e Ricerche Di Geografia* 21, no. 2 (2009): 191–94.
13. Gabellieri, Nicola. *Geografia Letteraria Dei Paesaggi Marginali La Toscana Rurale in Carlo Cassola*. Vol. 9. All’Insegna del Giglio, 2019.

14. Gavinelli, Dino. 'Geografia e Letteratura'. In *La Letteratura Contemporanea Nella Didattica Della Geografia e Della Storia*. Milano: CUEM, 2006.
15. Grava, Massimiliano. 'Imágenes Estúpidas versus Imágenes Inteligentes. Empleo de Webgis y Cloud Services Para La Publicación de Geo-Datos'. *Revista Uruguay de Historia Económica* 6, no. 9 (2016): 73–80.
16. Grava, Massimiliano, Camillo Berti, Nicola Gabellieri, and Arturo Gallia. *Historical GIS. Strumenti Digitali per La Geografia Storica in Italia*. EUT Edizioni Università di Trieste, 2020.
17. Gregory, Ian N., and David Cooper. 'Thomas Gray, Samuel Taylor Coleridge and Geographical Information Systems: A Literary GIS of Two Lake District Tours'. *International Journal of Humanities and Arts Computing* 3, no. 1–2 (2009): 61–84.
18. Grendi, Edoardo. 'Dal Grand Tour a "La Passione Mediterranea"'. *Quaderni Storici* 34, no. 1 (1999): 121–134.
19. Humboldt, Alexander von. *Cosmos: essai d'une description physique du monde*. Paris: Gide et J. Baudry Editeur, 1844.
20. Lévy, Bertrand. 'Géographie et Littérature. Une Synthèse Historique'. *Le Globe* 146 (2006): 25–52.
21. Marengo, Marina. *Geografia e Letteratura: Piccolo Manuale d'uso*. Geografia e Organizzazione Dello Sviluppo Territoriale. Studi Regionali e Monografici 76. Bologna: Pàtron editore, 2016. PN56.G48 M37 2016.
22. Mogorovich, Paolo, and Piero Mussio. 'Automazione Del Sistema Informativo Territoriale'. In *Elaborazione Automatica Dei Dati Geografici*. Masson, 1988.
23. Moretti, Franco. *Graphs, Maps, Trees: Abstract Models for a Literary History*. Verso, 2005.
24. Murrieta-Flores, Patricia, Christopher Donaldson, and Ian Gregory. 'GIS and Literary History: Advancing Digital Humanities Research through the Spatial Analysis of Historical Travel Writing and Topographical Literature', 2017.
25. Pasqualin, Massimo. 'Beni Culturali, Informazione Cartografica e Potenzialità Ricreative Del Territorio: Due Contesti Diversi Nel Veneto', 2016.
26. Persi, Peris, and Elena Dai Prà. "*L'aiuola Che Ci Fa...*" *Una Geografia per i Parchi Letterari*. Urbino: Urbino University Press, 2001.
27. Ronchini, Monica, and A. De Bertolini. 'Il Paesaggio Rurale Dei Viaggiatori in Trentino'. In *Storia Dei Paesaggi Agrari Del Trentino, Terre coltivate.*, 84–88. Trento: Fondazione Museo Storico del Trentino, 2014.
28. Scaramellini, Guglielmo. *Paesaggi Di Carta, Paesaggi Di Parole: Luoghi e Ambienti Geografici Nei Resoconti Di Viaggio (Secoli 18.-19.)*. Giappichelli, 2008.
29. Smith, Joe. *Geography in Public and Public Geography: Past, Present and Future*. JSTOR, 2013.
30. Sprugnoli, Rachele. "'Two Days We Have Passed with the Ancients...': A Digital Resource of Historical Travel Writings on Italy", 2018.
31. Zorzi, Andrea. 'Documenti, Archivi Digitali, Metafonti'. *Archivi & Computer Automazione e beni culturali*, no. 10 (2000): 274–91.



# **Fuoco dal cielo. Luoghi di penitenza e di purificazione nella preistoria della scuola onlife**

Paolo Masini  
Istituto Polo Fermi-Giorgi – paolo.masini [at] gmail.com

## **PAROLE CHIAVE**

DH for the future, scuola onlife, formazione docenti, consumismo digitale, ambienti di apprendimento

## **INTERVENTO**

### **1 INTRODUZIONE**

Il presente contributo intende dare una evidenza critica sull'effettiva consistenza del legame tra le DH e i bisogni reali della società prendendo in considerazione alcune difficoltà che questo rapporto assume all'interno del "Sistema Scuola". Nei diversi ordini e gradi della Scuola italiana che cosa di fatto rappresentano le DH, quali implicazioni profonde hanno nel design didattico-pedagogico delle quotidiane attività in classe costruite dai docenti, quali sensibilità culturali orientano la consapevolezza professionale utile ad attingere al loro campionario per veicolare contenuti e valori adeguati, ad esempio, alla formazione del cittadino e ai principi della sostenibilità richiamati dall'Agenda 2030?

### **2 STRATEGIE DI COINVOLGIMENTO**

Trovandosi con un piede nel cuore dell'innovazione tecnologica, le DH subiscono inevitabilmente le pressioni potenti dell'industria e forse fanno talvolta fatica a perseguire l'impegno civile critico verso il capitalismo digitale [1]. Questo lato più o meno oscuro delle DH sta comunque dentro al piatto delle opportunità tecnologiche che il mercato offre ai docenti per realizzare le scelte didattiche coerenti con la loro progettazione, ed è inevitabile il rischio che le strategie di marketing li esponano a valutazioni superficiali in cui si perde di vista da un lato l'ibridazione di metodi, tecniche e linguaggi, dall'altro il valore culturale "rivoluzionario" legato alla pratica consapevole delle DH.

Ad esempio, per alcuni anni il tema del "fare" ha ispirato il dibattito intorno all'idea collettiva di una trasformazione radicale del capitalismo contemporaneo. Questo approccio presupponeva l'impegno che "l'uomo artigiano" poteva realizzare attraverso l'innovazione digitale e l'integrazione STEM. Ebbene, anche con varianti STEAM/STEAM che il mercato ha ampiamente articolato in prodotti didattici per tutti i gusti, in genere la scuola non è forse riuscita a mantenere nei contenuti il nesso profondo del "fare" con l' "humanitas" [5].

Ovviamente non dovunque questo accade e fortunatamente esistono, a macchia di leopardo, luoghi di approfondimento didattico e pedagogico rivolto alle DH. Mancando però una visione politica di “Sistema Scuola”, queste isole di ricerca e consapevolezza sono legate alle personalità periture di singoli Dirigenti, o ancora all’impegno di alcune Fondazioni e Istituzioni che investono nel loro territorio per radicare buone pratiche, oppure infine a network di Scuole finalizzate al potenziamento dell’innovazione didattica.

Quest’ultimo aspetto, le Reti formalizzate di scuole che esprimono cultura e valori attraverso l’integrazione delle DH nei curricula, merita una piccola nota indipendentemente dalle tipologie e senza lasciarci ingannare dalle dimensioni. Ci sono infatti casi di Reti di grande visibilità e quantitativamente importanti, la cui incidenza nell’aumento della cultura dell’innovazione didattica è oggettivamente difficile da valutare<sup>1</sup>. Altre invece hanno dimostrato che la chiave del successo sta nella creazione, ab origine, di uno forte spirito comunitario che sostiene la mission e rende possibile la condivisione di regole e codici culturali nei quali l’innovazione si esprime non attraverso la singola attività o metodologia didattica ma l’intero ambiente di apprendimento<sup>2</sup>.

Questo dimostra che le DH producono migliori frutti in ecosistemi scolastici nei quali la loro applicazione: è espressione coerente al flusso dei valori sociali, è il risultato di una progettazione condivisa, è integrata nella costruzione di una prossemica digitale nel contesto di “ambienti di comportamento” in presenza, a distanza e OnLife.

Diversamente è poco probabile che le DH riescano ad incidere veramente nei curricula della Scuola e non è infrequente imbattersi nel paradosso che l’uso di strumenti ideati per la partecipazione attiva a livello culturale, riescano a modulare tra gli addetti ai lavori solo alte frequenze di “Look at me!”, connotate da forte investimento emotivo (visibilità, popolarità, consolidamento o promozione di ruoli e posizioni, benefici ecc.) perseguendo sovente l’infida onda social di narcisismo digitale. È quindi importante, se intendiamo fare autentica memoria dell’Humanitas, che le DH debbano incrociarsi con le istanze radicali dell’innovazione didattica che passano dalla fatica della coscientizzazione socio-culturale [2], quindi probabilmente fuori dai coni di luce della Società dello Spettacolo.

### **3 THE BIG QUESTION(S): ETICA, SOCIALE, CULTURALE**

Il vuoto di curricula formativi nell’apprendistato scolastico formale, che mostrino la portata trasformativa delle DH applicate ai fatti e ai dati delle esperienze umane anche onlife, ha già conseguenze a tratti devastanti per la nostra vita attuale e futura. Richiamo qui almeno due recenti documenti sul tema: sta davvero accadendo qualcosa di incisivo nelle nostre aule? Perché le sole tecnologie, o la loro assenza, sono drammaticamente inadeguate.

---

1 Ad esempio la Rete Avanguardie Educative, un progetto di ricerca-azione di INDIRE.

- <http://innovazione.indire.it/avanguardieeducative/>

2 A titolo di esempio le Reti: Scuola Senza Zaino, Dada, Piccole Scuole. Sugli “ambienti di apprendimento” esiste uno specifico indirizzo di ricerca dell’INDIRE come riferimento nazionale.

- <https://www.scuolasenzazaino.org>

- <https://www.scuoledada.it>

- <https://www.indire.it/progetto/piccole-scuole/>

- <https://www.indire.it/progetto/architetture-scolastiche/>

Il primo documento ha per autore Andreas Schleicher, che è il Coordinatore del OECD Programme for International Student Assessment (PISA), un personaggio molto esperto ed autorevole [4]. Eli afferma: le indagini Pisa dimostrano che finora non ci sono apprezzabili miglioramenti negli apprendimenti degli studenti nei paesi che hanno investito pesantemente nelle tecnologie digitali per l'educazione. Forse non siamo ancora abbastanza bravi nel genere di pedagogia che sarebbe necessaria per trarne profitto: mettere insieme le tecnologie del XXI secolo con le pratiche pedagogiche del XX e con l'organizzazione scolastica del XIX secolo ha piuttosto il risultato di diminuire l'efficacia dell'insegnamento anziché di svilupparlo.

Sul piano del cambiamento sociale, Schleicher fa un esplicito riferimento alla escalation terroristica in Europa a partire dal 2014: gli attacchi terroristici in Europa, in particolare, hanno dimostrato che è fin troppo semplicistico ritrarre estremisti e terroristi come vittime della povertà o della scarsa istruzione. Le necessarie e numerose ricerche sul background e le biografie di estremisti e terroristi mettono in luce che queste persone spesso non provengono dalle parti più povere delle società. I radicali si trovano anche tra i giovani delle famiglie della classe media che hanno completato la loro istruzione formale. Ironia della sorte, quei terroristi sembrano ben attrezzati con le capacità imprenditoriali, creative e collaborative che sono diventate il fondamento di un'educazione del 21° secolo. [...] Quindi, come può l'istruzione combattere l'estremismo? Si arriva al cuore dell'educazione: insegnare i valori che possono dare agli studenti una bussola affidabile e gli strumenti per navigare con sicurezza attraverso un mondo sempre più complesso, volatile e incerto. [...] Per quanto sia difficile ottenere quel giusto equilibrio, gli educatori devono preparare gli studenti per le comunità culturalmente diverse e digitalmente collegate in cui lavoreranno e socializzeranno. È importante iniziare a riflettere su come i sistemi educativi si concretizzano in quella più ampia nozione di cittadinanza nel 21° secolo [4].

Il secondo è sempre un Documento dell'OECD dal titolo "The Future of Education and Skills Education 2030".

Nel 21° secolo, questo scopo è stato sempre più definito in termini di benessere. Ma il benessere implica più dell'accesso a risorse materiali, come reddito e ricchezza, posti di lavoro e guadagni, e alloggi. È anche legato alla qualità della vita, compresa la salute, l'impegno civico, le connessioni sociali, l'istruzione, la sicurezza, la soddisfazione della vita e l'ambiente. Un accesso equo a tutti questi elementi è alla base del concetto di crescita inclusiva.

L'istruzione ha un ruolo vitale da svolgere nello sviluppo delle conoscenze, abilità, attitudini e valori che consentono alle persone di contribuire e trarre beneficio da un futuro inclusivo e sostenibile [...]. L'istruzione deve mirare a fare di più che preparare i giovani al mondo del lavoro; ha bisogno di fornire agli studenti le competenze di cui hanno bisogno per diventare cittadini attivi, responsabili e impegnati [3].

Ed è in questo dominio che le DH possono rivestire un nuovo ruolo fondamentale, consentendo di estendere il potere generativo del loro essere nativamente OnLife a tutte le dinamiche dell'apprendimento.

## BIBLIOGRAFIA

1. Grimshaw, Mike. 'Towards a Manifesto for a Critical Digital Humanities: Critiquing the Extractive Capitalism of Digital Society'. *Palgrave Communications* 4, no. 1 (27 February 2018): 1–8. <https://doi.org/10.1057/s41599-018-0075-y>.
2. Manfredi, Silvia Maria, and Piergiorgio Reggio. 'Educazione e Coscienza Critica. Note Sul Concetto Di "Coscientizzazione" in Paulo Freire'. *Animazione Sociale* 5 (2007).
3. OECD. 'The Future of Education and Skills. Education 2030'. OECD, 2018.
4. Schleicher, Andreas. *How to Build a 21st-Century School System*. Paris: OECD, 2018.
5. Wise, Tim. 'Forget STEM, We Need MESH'. *Our Human Family* (blog), 17 September 2019. <https://medium.com/our-human-family/forget-stem-we-need-mesh-43ab6f6273cd>.

# **Sessione IX**

## Karl Popper





# Storia e rivoluzione digitale. Una riflessione tecnica e teoretica

Manfredi Scanagatta  
Università di Modena e Reggio Emilia – manfredi.scanagatta [at] gmail.com

## SINTESI

I processi di digitalizzazione che coinvolgono anche la disciplina della storia, per essere compresi richiedono di essere analizzati sia da un punto di vista tecnico informatico sia da un punto di vista epistemologico. Perché lo storico possa essere in grado di utilizzare al meglio i nuovi strumenti che ha a disposizione, è necessario che si interroghi su le procedure informatiche che sottendono ai processi di digitalizzazione delle fonti materiche, alla realizzazione di archivi digitali, alla metadattazione delle fonti e alle possibilità di utilizzo del web semantico per la costruzione di ontologie all'interno degli archivi. Dall'altra parte i tempi ci chiedono di riprendere in mano la discussione sul valore della storia in sé, quei percorsi cioè di analisi metastorica che oggi si devono confrontare con nuovi strumenti, così impattanti sulla disciplina da richiedere un'analisi critica dei significati epistemologici sui quali la storia fonda il proprio metodo. Oltre al complesso tema della metadattazione e dell'interoperabilità delle fonti digitalizzate in un'ottica semantica, dobbiamo riflettere sull'incredibile vastità di fonti digital born che vengono oggi prodotte, per iniziare a chiedersi come potersi relazionare con queste nuove fonti non solo in termini quantitativi, ma qualitativi.

## PAROLE CHIAVE

digitalizzazione, fonti, archivi digitali, digital born, metastoria, ontologie digitali, spiegazione, narrazione

## INTERVENTO

### 1 INTRODUZIONE

Cosa succede quando una fonte materica viene digitalizzata? In che modo muta il rapporto tra storico e archivio quando l'archivio è digitale? Quali fonti *digital born* saranno utili allo storico del futuro e in che modo potrà usarle? Queste sono solo alcune delle domande alle quali non pretendo di dare risposte, ma che spero possano contribuire alla costruzione di un ragionamento teso a riannodare tra loro fili che apparentemente provengono da differenti matasse.

I lavori di Pigliapoco, Vitali, Zanni Rosiello, Possimato, Dinoia, Meschini, Wesson, solo per citare alcuni tra ricercatori e docenti che lavorano ormai da anni sui temi della digitalizzazione degli archivi e le prospettive del web semantico, aiutano lo storico a provare a comprendere come stiano cambiando gli archivi che da fisici e materici si trasformano in digitali.

Seguendo l'inevitabile avanzare dei processi tecnologici lo storico costruisce un rapporto più profondo e complesso con il web, diventando *de facto* un abitante anche se non sempre consapevole, di ciò che Floridi chiama *infosfera*.

In questa trasformazione lo storico deve come prima cosa comprendere ciò che già esiste e che sta già modificando il lavoro di ricerca e di diffusione del contenuto storico. Basti pensare agli archivi digitali che utilizzano i linked open data e gli RDF, capaci dunque di generare ontologie digitali dove il simulacro della fonte materica si mostra al ricercatore all'interno di una rete di collegamenti ipertestuali di natura semantica. Possiamo azzardare che questi archivi siano o possano essere autoesplicativi e che mostrando la fonte diano forma alle parole di Rossana Lita, quando afferma che "solo nel suo medium la realtà storica diventa accessibile". E lo storico?

La realizzazione di ontologie digitali interne agli archivi deve riportare l'attenzione su un tema che dalla metà dell'ottocento contraddistingue il dibattito metastorico. Dobbiamo chiederci come si riorganizza il rapporto tra ricerca e spiegazione e quale sia il valore della narrazione.

E' necessario dunque riprendere il pensiero di autori quali Collingwood, Gallie, Dray, Hempel, White, Ricoeur, Ginsburg, Danto, per osservarlo in modo critico con la consapevolezza che la digitalizzazione, così come è nelle aspettative della Public History, consente una nuova fruizione della fonte primaria che con sempre maggior semplicità potrà divenire elemento intellegibile all'interno del discorso storico.

In un mondo digitalizzato e iperconnesso l'individuo singolare produce una mole di fonti prima neanche pensabile. Per dirla con Ferraris ogni azione compiuta su un piano digitale porta come conseguenza una registrazione e un nuovo livello di scrittura, di documentalità. Questo processo ci consente di riprendere in mano il dibattito sul valore del termine *Geschichte* che alla fine del XVIII secolo ha dato forma all'idea del singolare collettivo, dove le storie vanno a comporre la storia.

## 2 PROPOSTA

La rivoluzione digitale che sta coinvolgendo il nostro tempo ci chiama a dover rispondere a specifiche domande che investono tutti i campi della nostra esistenza. Le tecnologie digitali non sono soltanto strumenti che si limitano a modificare il modo in cui interagiamo con il mondo, sono soprattutto sistemi che danno forma, formattano e influenzano sempre di più il modo in cui comprendiamo il mondo e ci rapportiamo ad esso, così come il modo in cui concepiamo noi stessi e interagiamo tra noi.

Come scrive Floridi le tecnologie digitali «in altre parole son re-ontologizzanti, cioè modificano la natura intrinseca, l'ontologia di quello che toccano» [11].

Per Ferraris «il web è terribilmente materiale, ed è tutt'altro che irrealista, non è il mondo del sogno contrapposto a quello della realtà, bensì un mondo ancora più reale, perché più presente, concentrato, documentato» [9].

Il web è lo strumento che usiamo per osservare, comunicare e commentare la nostra esistenza passata, presente e futura e in questa circostanza il web è per noi da osservare come strumento di gestione e fruizione delle *tracce* [2] del nostro passato.

La digitalizzazione, la dematerializzazione e rimediazione [3] delle fonti primarie e la conseguente costruzione di archivi digitali, si presenta come fenomeno capace di mutare l'approccio alla ricerca storica, di favorire la narrativizzazione della storia attraverso la messa in scena delle fonti in un'ottica di Public History [18], e di aprire il mondo degli archivi e delle informazioni in essi contenute a chiunque abbia una connessione internet.

Quando si naviga un archivio digitale e si procede con le nostre ricerche interagendo con lo spazio virtuale e con la rappresentazione simulacrale dei documenti, ci si trova a confrontarci con un'esperienza tele-epistemica, in cui non è l'osservatore  $x$  (il ricercatore) in SLO (Spazio Locale di Osservazione), che accede all'ente  $y$  (il documento) in SRO (Spazio Remoto di Osservazione) a essere presente in SRO, ma al contrario è l'ente  $y$  in SRO a cui l'osservatore ha accesso telepisticamente a essere presente anche nello SLO dell'osservatore in quanto portatore di proprietà [11].

Seguendo le teorie sulla presenza all'interno dei processi che governano la filosofia dell'informazione e l'*infosfera*, giungiamo a poter ipotizzare che nell'interazione uomo-digitale, nell'ambito della ricerca storica, non sia più il ricercatore a recarsi in un archivio, ma che sia l'archivio a divenire presente nello spazio di lavoro dello storico.

Questo processo è facilmente comprensibile se consideriamo che grazie alla digitalizzazione dei documenti e dall'archivio, non solo lo storico può portare avanti le sue ricerche senza doversi fisicamente spostare all'interno di un archivio e senza comunque l'utilizzo di sistemi di telepresenza, ma anche e soprattutto perché il documento che ha subito la rimediazione viene visualizzato sul computer del ricercatore, il quale ne può salvare una copia, in forma di immagine, che rimane presente solo nello spazio di lavoro dello storico, che a quel punto può riutilizzarla anche se non connesso a internet e dunque all'archivio digitale. È nello SLO del ricercatore che avviene un cambiamento, dove le proprietà del documento-medium si trasferiscono interamente.

Se all'interno di un archivio storico di tipo materico le fonti sono archiviate mantenendo tra di loro una correlazione di natura fisica e logica, con la dematerializzazione delle fonti primarie e la costruzione di archivi digitali le fonti sono depositate in una organizzazione unicamente logica e non più fisica [13]; l'organizzazione archivistica viene coordinata grazie all'utilizzo di metadati e definire degli standard di metadattazione diviene fondamentale per assicurare che le fonti digitalizzate che verranno ordinate all'interno di questi sistemi di gestione, mantengano tra di loro dei vincoli specifici, imprescindibili per la fruizione dell'archivio digitale da parte di uno storico [7].

La metadattazione della fonte digitalizzata consente di attribuire informazioni specifiche alle fonti e attraverso l'utilizzo del modello RDF [14], quando i metadati sono collegati ad un vocabolario comune come nel caso di FOAF o DCMI, assumono anche un valore semantico, consentendo l'interoperabilità tra le fonti in una connessione ipertestuale di natura semantica che conduce alla costruzione di ontologie digitali tematiche all'interno dell'archivio digitale stesso [5].

Un altro processo che determina in modo sensibile il rapporto di fruizione tra un archivio digitale e un ricercatore è dato da strumenti di gestione dell'archivio digitale come OAI, Open Archival Information System, che come scrive Silvio Salza è oggi unanimemente considerato il riferimento per i sistemi di conservazione digitale<sup>1</sup>, tra le altre cose perché composto da tre livelli detti *Information Package* che devono essere gestiti in collaborazione dall'ente che produce i documenti e da quello che ne garantisce la conservazione e la fruizione.

---

1 <http://www.conservazionedigitale.org/wp/materiale-didattico/il-modello-oais/>

In modo sintetico il modello funziona tra l'interazione di:

- *Submission Information Package (SIP)*, che viene trasmesso nella fase di versamento dal produttore al deposito;
- *Archival Information Package (AIP)*, che viene generato a partire dal SIP in fase di accettazione (*Ingestion*) e poi diventa oggetto diretto della conservazione deposito;
- *Dissemination Information Package (DIP)*, che viene generato a partire dall'AIP per essere distribuito alla Comunità Designata per la fruizione

I tre livelli di Information Package sono i livelli che maggiormente ci aiutano a riflettere su quale sia e possa essere in ambito digitale la relazione tra fonte-archivio-fruitore e che se riportati in un'ottica di Public History richiamano al rapporto tra storia-storico-pubblico.

Un sistema di gestione e archiviazione digitale, come lo è OAIS, deve essere in grado di contenere e elaborare differenti dati che conducano alla definizione di quelle che possiamo sostenere essere le quattro componenti maggiormente significative nei processi di digitalizzazione e gestione delle fonti: creazione del Digital Object; Representation of Information; individuazione della Designated Community; comprensione e condivisione della Knowledge Base. Ognuna di queste componenti richiede di essere analizzata epistemologicamente dallo storico, dall'archivista o dal teorico della storia, dal momento che coinvolgono le fasi significative del lavoro di archiviazione, ricerca, analisi e diffusione delle fonti. Per ognuna di queste fasi si aggiungono e si pensa come organizzare delle informazioni che entrano in relazione diretta con le informazioni che determinano l'esistenza e il valore stesso del documento storico. I livelli informativi riferiti al documento dunque si stratificano e richiedono una differente e nuova lettura e comprensione rispetto alla consueta archiviazione, ricerca e analisi e di fonti materiche.

La digitalizzazione e la costruzione di ontologie digitali si manifesta come un processo di neo-intermediazione digitale della storia, chiunque navigherà un archivio digitale non si troverà da solo all'interno della sua ricerca, ma si confronterà con ontologie digitali capaci di accompagnare il ricercatore.

I processi di digitalizzazione che coinvolgono la disciplina della storia possono essere osservati attraverso la lente di ciò che Floridi ha definito *infosfera* [10], un luogo fisico e virtuale all'interno del quale non siamo entità isolate, quanto piuttosto agenti informativi interconnessi, che condividono con altri agenti biologici e artefatti ingegneristici un ambiente globale, costituito in ultima istanza da informazioni [11].

Secondo Floridi per riuscire a comprendere la rivoluzione che stiamo vivendo e dunque produrre pensiero in grado di descrivere questi cambiamenti, dobbiamo considerare noi stessi come degli *inforганиsmi* in un'esperienza che è descrivibile in termini di *on-life experience*.

La rivoluzione digitale modifica l'interazione dello storico con entrambe le fasi che caratterizzano il suo lavoro, quella di ricerca e analisi critica delle fonti e quella di diffusione del contenuto. Questo ci deve spingere a compiere un'analisi critica dei processi epistemologici che definiscono il rapporto tra fonte e storico, sia che si tratti di fonti materiche digitalizzate, con la conseguente facilitazione della fruizione attraverso il web, la costruzione di ontologie e i processi di interazione e selezione dati dal Web Semantico, sia per quanto riguarda la produzione e l'analisi delle fonti *digital born*.

La digitalizzazione della società ha dato un nuovo senso al concetto di scrittura e iscrizione. Ad oggi sono davvero poche le azioni di un essere umano che non passano attraverso l'utilizzo di strumenti digitali che a loro volta generano e lasciano tracce delle azioni che stiamo compiendo.

Su questo tema gli studi di Ferraris sono fondamentali e propedeutici ad affrontare la questione dell'informazione. E di sicuro interesse sono le riflessioni riguardo l'*oggetto sociale* che mi pare trovino un'interessante collocazione all'interno del dibattito metastorico su quali siano gli "oggetti" di interesse della storia come disciplina.

Le teorie di Gardiner [12], così come di Collingwood [4] sull'individualità dell'evento storico, combinate al pensiero di Dray [6] rispetto alla necessità per lo storico di comprendere il pensiero dell'agente, si possono collocare all'interno di quel dibattito che dalla seconda metà del XVIII secolo ha provato a descrivere, per utilizzare una metafora di Plank [15], la storia come trama tessuta attraverso le piccole storie, a significare come la storia si componga di un singolare collettivo.

Secondo Ferraris il web è un apparato di registrazione, come la scrittura, il denaro, l'inventario e l'archivio. Il più grande apparato di registrazione che la storia abbia conosciuto fino a qui e il più potente. Ogni azione compiuta su un piano digitale porta come conseguenza una registrazione e un nuovo livello di scrittura, di documentalità [8].

Se questa tesi la osserviamo come valida all'interno dell'*infosfera* descritta da Floridi, ci ritroviamo in un contesto in cui praticamente tutte le azioni di un essere umano che vive *on-life* producono un documento dunque una fonte osservabile.

La rivoluzione digitale credo ci chieda di dover riosservare il rapporto tra storia e storiografia, partendo dal dibattito che nella fine del XVIII secolo si è sviluppato intorno al termine *Geschichte*, nel tentativo di ricondurre in un unico spazio semantico la differenza tra *Res Gestae* e *Historia Rerum Gestarum*, dando vita all'idea del singolare collettivo che riporta le storie nella storia e che pone lo storico e il filosofo di fronte a nuove prospettive.

Osservando oggi anche la storia di domani non possiamo che condividere le parole di Ferraris e dunque porci nuove domande a partire dal presupposto che «un tempo l'oblio era la regola e la traccia l'eccezione, oggi vale l'opposto» [9].

Se questo intreccio tra teorie della filosofia della storia e rivoluzione digitale non può che apparire suggestivo, dobbiamo tenere in considerazione che all'interno di questo processo anche il lavoro dello storico sarà rivoluzionato.

La digitalizzazione ha dato vita a sistemi informatici capaci di collezionare documenti del passato e renderli al fruitore in un'apparente disintermediazione, un processo dove lo storico scompare e una neointermediazione digitale non esplicita è in grado di generare gli elementi per produrre discorso storico. Le possibilità del web semantico con la conseguente costruzione di ontologie digitali può portare alla realizzazione di archivi storici digitali autoesplicativi, dove il lavoro normalmente demandato allo storico, quello della ricerca e della spiegazione può essere sostituito da nuove intelligenze.

Se in chiave non positivista, in *Narrare la storia*, Jerzy Topolski afferma che ogni racconto storico sottende a specifiche pratiche retoriche tipiche della narrazione letteraria, la stratificazione di linguaggi e significati attribuiti alle fonti, che avviene con le operazioni di digitalizzazione e di archiviazione attraverso sistemi come OAIS [19], ci può far riflettere riguardo a processi metanarrativi che si attivano ancor prima che lo storico compia il lavoro di racconto e che vanno ad influenzare i processi di ricerca, potremmo supporre anticipando, in un ipotetico processo analitico dei sistemi di narrazione normalmente presenti nell'apparato storiografico, questioni relative alla contestualizzazione e alla spiegazione. Oltre a ciò la rivoluzione digitale ha fatto sì che lo storico del futuro potrà/dovrà confrontarsi con una mole di fonti in passato neanche pensabile.

Il nostro comportamento intelligente è chiamato a confrontarsi con quello di artefatti ingegneristici e le tecnologie digitali sembrano talora conoscere i nostri desideri meglio di noi stessi [11].

È importante comprendere che la rivoluzione digitale ci pone di fronte alla necessità di chiedersi quale sia oggi il rapporto tra storia, storico e pubblico, e in un'ottica in cui il rapporto tra ricercatore e fonte muta in base alla presenza di strumenti digitali è necessario interrogarsi su come osservare la tensione tra ricerca, spiegazione e narrazione. Se la storia è in divenire, anche le riflessioni su di essa non possono che mutare con il mutare dei tempi.

## BIBLIOGRAFIA

1. Bertucelli, Lorenzo, Alfonso Botti, e Paolo Bertella Farnetti. *Public History. Discussioni e pratiche*. Sesto San Giovanni: Mimesis, 2017.
2. Bloch, Marc. *Apologia della storia*. Torino: Einaudi, 1998.
3. Bolter, J. D., e R. Grusin. *Rimediatio. Competizione e integrazione tra media vecchi e nuovi*. Milano: Guerini, 2002.
4. Collingwood, Robin George. *Il concetto della storia*. Milano: Fabbri, 1966.
5. Di Noia, T., R. De Virgilio, E. Sciascio, e F. M. Donini. *Semantic Web. Tra ontologia e open data*. Santarcangelo di Romagna: Maggioli, 2018.
6. Dray, William. *Leggi e spiegazione in storia*. Milano: Il Saggiatore, 1974.
7. Feliciati, Pierluigi. «Gestione e conservazione di dati e metadati per gli archivi: quali standard?» In *Conservare il digitale, Riflessioni su modelli archivistici, figure professionali e soluzioni applicative*. Macerata, 2009. <http://eprints.rclis.org/13657/1/atti-testoFeliciati09.pdf>.
8. Ferraris, Maurizio. *Documentalità. Perché è necessario lasciar tracce*. Bari: Laterza, 2009.
9. Ferraris, Maurizio, e Germano Painsi. *Scienza nuova. Ontologia della trasformazione digitale*. Torino: Rossemberg e Sellier, 2018.
10. Floridi, Luciano. *La quarta rivoluzione. Come l'infosfera sta trasformando il mondo*. Milano: Raffaello Cortina, 2017.
11. ———. *Pensare l'infosfera: La filosofia come design concettuale*. Milano: Raffaello Cortina, 2020.
12. Gardiner, Patrick. *The Nature of Historical Explanation*. Oxford: Oxford University Press, 1952.
13. Guercio, Maria Rosaria. *Archivistica informatica. I documenti in ambiente digitale*. Roma: Carocci, 2000.
14. Guerrini, Mauro, e Tiziana Possemato. *Linked data per biblioteche, archivi e musei*. Milano: Editrice Bibliografica, 2015.
15. Koselleck, Reinhart. *Storia: la formazione del concetto moderno*. Bologna: CLUEB, 2009.
16. Pigliapoco, Stefano. *Progetto archivio digitale*. Torre del Lago: Civita Editoriale, 2018.
17. Rossi, P, a c. di. *La teoria della storiografia oggi*. Milano: Il Saggiatore, 1983.
18. Scanagatta, Manfredi. «Public Historian, tra ricerca e azione creativa». In *Public History. Discussioni e Pratiche*, a cura di Paolo Bertella Farnetti, Lorenzo Bertucelli, e Alfonso Botti. Milano: Mimesi, 2015.
19. Topolski, Jerzy. *Narrare la storia. Nuovi principi di metodologia storica*. Milano: Mondadori, 1997.
20. Vitali, Stefano. *Passato digitale: le fonti dello storico nell'era del computer*. Milano: Mondadori, 2004.
21. White, Hayden. *Metahistory: retorica e storia*. Milano: Meltemi, 2014.

# I bibliotecari della pubblica amministrazione tra gestione documentale e information literacy

Lucia Antonelli

Biblioteca dell'Albo nazionale dei Segretari comunali e provinciali. Ministero dell'Interno  
lucia.antonelli [at] interno.it

## SINTESI

I bibliotecari operanti all'interno di ministeri, enti e istituzioni pubbliche si occupano, oltre alle attività ordinarie e tradizionali tipiche della professione, anche della gestione e della diffusione on line del patrimonio informativo-documentale pubblico. Si tratta di una specificità che andrebbe maggiormente valorizzata e che colloca il bibliotecario della pubblica amministrazione nella sfera di azione delle *Digital Humanities*, sia in virtù delle attività di gestione della documentazione digitale, sia per le attività e i servizi finalizzati a far acquisire competenze digitali alla propria utenza di riferimento.

## PAROLE CHIAVE

documentazione di fonte pubblica on line, biblioteche e bibliotecari della pubblica amministrazione, information literacy

## INTERVENTO

### 1 INTRODUZIONE

Come tutte le biblioteche, anche quelle della pubblica amministrazione conservano, organizzano, catalogano e mettono a disposizione dell'utenza un patrimonio documentario ricco ed eterogeneo che comprende libri, riviste, banche dati, risorse elettroniche di vario tipo. Oltre a questo, una funzione tradizionalmente caratterizzante per le biblioteche delle pubbliche amministrazioni consiste nella raccolta, la conservazione, l'organizzazione e la diffusione della cosiddetta "documentazione di fonte pubblica"<sup>1</sup>, che negli ultimi anni è ampiamente disponibile on line in formato digitale.

Sebbene molte biblioteche afferenti a ministeri, enti e istituzioni pubbliche rappresentino, proprio grazie alla messa a disposizione della documentazione pubblica, uno dei canali di collegamento tra l'utenza di riferimento e le amministrazioni pubbliche, troppo spesso il ruolo dei bibliotecari che vi operano è inteso come marginale, oltre che poco o nulla valorizzato.

---

1 Su questo aspetto si veda la Presentazione di Sabino Cassese al volume a cura di Sandro Bulgarelli, Madel Crasta e Patrizia Valentini. *Le biblioteche dell'amministrazione centrale dello Stato italiano*. Roma: Associazione italiana biblioteche, 1990, p. 7



Un cambiamento di paradigma può avvenire se i bibliotecari acquisiranno maggior consapevolezza del proprio ruolo all'interno delle amministrazioni di appartenenza e se si attiveranno per rafforzare la specificità di alcune mansioni in particolare: la gestione della documentazione di fonte pubblica on line e la formazione finalizzata a far acquisire agli utenti le competenze digitali necessarie per il recupero di informazioni, dati e documenti pubblici presenti sul web.

## **2 LA DOCUMENTAZIONE DI FONTE PUBBLICA**

La documentazione di fonte pubblica, comunemente intesa come l'insieme di dati, informazioni e documenti prodotti dalle istituzioni pubbliche nell'esercizio delle loro diverse attività istituzionali, può essere suddivisa in due categorie distinte. All'interno di un primo gruppo è inclusa quella documentazione senza la quale le amministrazioni pubbliche non vedrebbero svolte le proprie attività: atti normativi, sentenze, delibere, direttive, circolari, determinazioni dirigenziali e molto altro ancora. In un secondo gruppo vanno compresi invece i rapporti, gli studi, le ricerche, le riviste, le linee guida, le raccolte statistiche, i bollettini, i dossier: in questo caso si tratta di documenti che illustrano attività svolte nell'ambito di un'istituzione o ne approfondiscono le tematiche nell'ambito del dominio di appartenenza. A questa macroclassificazione vanno aggiunti gli "Open data", che da qualche anno ampliano l'ordinamento tassonomico della documentazione di fonte pubblica e che sono caratterizzati da peculiarità tecnologico-formali a sé.

Il web è lo strumento attraverso il quale le istituzioni pubbliche mettono a disposizione della comunità il proprio patrimonio informativo-documentario e i destinatari sono i cittadini, le imprese, le istituzioni. In base alla normativa vigente<sup>2</sup>, la documentazione prodotta dalle pubbliche amministrazioni è da considerarsi "aperta", vale a dire pubblicata sui siti istituzionali, liberamente fruibile e riutilizzabile da chiunque, rilasciata in formati aperti; è necessario inoltre che tali documenti siano disponibili nel tempo, costantemente aggiornati, facili da ricercare e recuperare da parte di tutta la comunità.

## **3 IL RECUPERO ON LINE DELLA DOCUMENTAZIONE DI FONTE PUBBLICA: CRITICITÀ E SOLUZIONI**

Dunque, il panorama della documentazione pubblica è costituito da diverse tipologie informativo-documentarie, rispetto alle quali, però, si riscontrano diverse criticità in ordine alla loro diffusione e la loro accessibilità on line:

- per le norme e le sentenze, la ricerca e il recupero possono risultare difficili a causa di database caratterizzati da modalità di interrogazione poco amichevoli;
- per gli atti amministrativi, non è raro imbattersi in un patrimonio documentale obsoleto o incompleto, del quale chi naviga sul web può non avere consapevolezza;
- per gli Open data, risultano ancora poco diffusi i siti web che prevedono sezioni o portali ad essi dedicati e spesso i dataset non sono aggiornati o vengono rilasciati con livelli qualitativamente bassi in termini di apertura e interoperabilità;
- per la cosiddetta "editoria pubblica", si rileva una mancanza di uniformità e di standardizzazione in ordine alle modalità di accesso che ostacola una loro agevole fruizione via web.

---

2 D.lgs. 7 marzo 2005, n. 82 - Codice dell'amministrazione digitale e D.lgs. 14 Marzo 2013, n. 33 - Riordino della disciplina riguardante il diritto di accesso civico e gli obblighi di pubblicità, trasparenza e diffusione di informazioni da parte delle pubbliche amministrazioni.

Tali criticità rappresentano la spia di una gestione non sempre efficace del patrimonio documentario pubblico on line, che rischia sovente di essere poco valorizzato oltre che recuperato con difficoltà. Le possibili soluzioni per fronteggiare limiti e carenze informative andrebbero affidate - più di quanto non avvenga attualmente - ai bibliotecari delle pubbliche amministrazioni che, in quanto figure professionali esperte di documentazione, hanno le competenze adeguate per intervenire in diversi ambiti e limitare alcune criticità: dalla formazione all'utenza per la ricerca sulle banche dati, all'individuazione dei criteri di indicizzazione e classificazione degli Open data; dalla gestione on line dell'editoria pubblica digitale, alla strutturazione dell'architettura dei siti web istituzionali.

Alla luce delle problematiche relative alla gestione e al recupero dei documenti di fonte pubblica on line, emergono dunque due questioni di fondo strettamente connesse. Da una parte, le pubbliche amministrazioni dovrebbero valorizzare la documentazione istituzionale e, conseguentemente, definire in modo più strutturato le competenze, i ruoli e le risorse professionali demandate alla loro gestione. Dall'altra, i cittadini dovrebbero poter acquisire competenze digitali adeguate ai fini del recupero di dati e informazioni attraverso i siti web istituzionali. Per entrambe le questioni, i bibliotecari delle pubbliche amministrazioni possono diventare gli artefici del cambiamento.

#### **4 IL RUOLO DEI BIBLIOTECARI DELLE PA NELLA GESTIONE DELLE RISORSE DIGITALI**

Tradizionalmente il ruolo dei bibliotecari è quello di organizzare le informazioni bibliografiche relative ai manufatti della conoscenza. Nel contesto digitale tale ruolo non ha subito modifiche sostanziali: come per i libri e le riviste cartacee, anche per i documenti in formato digitale il bibliotecario è di fatto la figura professionale demandata alla loro organizzazione, catalogazione, classificazione, messa a disposizione e valorizzazione. Per estensione: poiché la documentazione di fonte pubblica on line è una tipologia di risorsa digitale, in quanto tale necessita anch'essa di essere ordinata e catalogata da figure professionali adeguate, in modo da essere ricercata, recuperata e utilizzata da parte di eventuali portatori di interesse. In particolare per i prodotti editoriali di una pubblica amministrazione, la loro messa a disposizione on line è una pratica che compete legittimamente ai bibliotecari, che sono in grado di utilizzare strumenti gestionali quali i "repositories", supportati da appositi *content management system* (ad esempio: D-Space, E-Prints, Omeka). Tali sistemi consentono di organizzare e descrivere le pubblicazioni con modalità assimilabili a quelle di un catalogo bibliografico, consentendo anche il recupero *full-text* on line dei documenti ricercati. I *repositories* documentali, già largamente utilizzati in ambito accademico nel rispetto dei principi dell'accesso aperto, grazie ai bibliotecari potrebbero essere adottati anche dalle pubbliche amministrazioni per la gestione on line di riviste, studi, ricerche, relazioni, monografie tematiche ed altra documentazione prodotta. In tal modo si avrebbero benefici molteplici e condivisi: per le istituzioni perché vedrebbero notevolmente valorizzata e diffusa la propria produzione editoriale, per i cittadini perché potrebbero accedere in modo più agevole ai documenti, per le biblioteche delle pubbliche amministrazioni perché implementerebbero l'offerta del patrimonio documentale a disposizione dell'utenza, infine per i bibliotecari perché acquisirebbero nuove competenze.

Ma oltre a questo, l'intervento dei bibliotecari delle pubbliche amministrazioni dovrebbe essere previsto anche in altri ambiti, connessi con la comunicazione on line e la diffusione di risorse digitali: nella progettazione dell'architettura di un sito web istituzionale e nella realizzazione dei database che gestiscono gli Open data, strutturando efficacemente le informazioni grazie a sistemi di classificazione logico-semantica.

In quest'ottica i bibliotecari delle pubbliche amministrazioni potrebbero apportare un contributo sostanziale e di sicuro impatto nella valorizzazione dell'intero patrimonio informativo digitale prodotto nell'ambito dell'istituzione di appartenenza.

## **5 I BIBLIOTECARI DELLE PA E L'INFORMATION LITERACY**

Oltre alla gestione delle risorse digitali, i bibliotecari delle pubbliche amministrazioni dovrebbero sempre di più attivarsi per offrire ai propri utenti tutti gli strumenti utili per orientarsi nelle attività di ricerca e recupero della documentazione ufficiale prodotta da istituzioni pubbliche. Tali attività rientrano nella cosiddetta "information literacy" e sono finalizzate a far acquisire agli utenti delle specifiche competenze informative, sia attraverso interventi *on demand*, sia attraverso veri e propri percorsi formativi, in presenza o a distanza.

L'*information literacy* consiste nella capacità di saper gestire la ricerca, il recupero, la valutazione e l'utilizzo di informazioni, ai fini della propria conoscenza individuale e per la creazione di altra e ulteriore conoscenza. Le biblioteche rappresentano un contesto privilegiato per l'*information literacy*, non solo perché custodiscono e mettono a disposizione della collettività informazioni, dati e documenti, ma anche e soprattutto perché sono i luoghi di quello che negli anni '90 è stato definito come il "sapere organizzato"<sup>3</sup>. La biblioteca possiede ancora tali prerogative e continua a mantenerle nonostante l'avvento di internet e dei motori di ricerca, caratterizzati dalla frammentazione e dal sovraccarico informativo, dal sapere distribuito, dall'assenza di intermediazione.

Si auspica che i bibliotecari della pubblica amministrazione possano offrire sempre più diffusamente ai propri utenti corsi sull'uso di banche dati normative e giurisprudenziali, guide per la ricerca su *opac* e su *discovery tool*, suggerimenti per ricerche in rete mirate al recupero della documentazione pubblica; il tutto calibrando l'offerta di servizi in base ai bisogni formativi e informativi eventualmente rilevati. Entrando nel merito, ad esempio, di un corso sul recupero della normativa per dirigenti e funzionari dell'amministrazione di appartenenza, il bibliotecario dovrà porre in risalto l'importanza dell'utilizzo di fonti ufficiali, attendibili e aggiornate, che rappresentano la migliore garanzia per la redazione corretta di un provvedimento amministrativo.

La diffusione di attività di *information literacy* nelle biblioteche delle amministrazioni pubbliche consentirebbe inoltre di rafforzare per gli utenti interni quelle competenze digitali ritenute ormai imprescindibili per l'attuazione di buone pratiche in una pubblica amministrazione di qualità. Le competenze informative e digitali consentono di padroneggiare le tecnologie dell'informazione e della comunicazione e permettono di acquisire competenze di *e-leadership*, caratterizzate dall'utilizzo efficace delle tecnologie digitali nell'ambito nella propria istituzione e finalizzate alla pianificazione e alla realizzazione concreta di progetti di innovazione nell'ambito di una pubblica amministrazione efficace ed efficiente, oltre che rivolta al futuro e all'innovazione.

---

3 Guido Martinotti. *Informazione e sapere*. Milano: Anabasi, 1992.

## 6 CONCLUSIONI

Per i bibliotecari di ministeri, enti e istituzioni pubbliche, una strutturata azione congiunta che coniughi *information literacy* e gestione della documentazione pubblica digitale può impattare positivamente in ambiti e livelli differenti: sui servizi bibliotecari, sull'amministrazione di appartenenza, sul ruolo delle *Digital Humanities* nel contesto pubblico, sugli utenti della biblioteca. Perché questo sia possibile, i bibliotecari dovranno però cogliere l'opportunità di rendere partecipi le amministrazioni di appartenenza di una visione strategica secondo cui anche l'utilizzo di sistemi sofisticati di gestione documentale e l'attivazione di corsi di alfabetizzazione digitale contribuiscono a mantenere vivi i principi di efficacia, efficienza ed economicità dell'azione amministrativa e a costruire le basi per una cittadinanza informata, digitale e consapevole. Il compito più arduo, ma non impossibile, sarà quello di rendere diffusamente assodato che il bibliotecario è un artefice del cambiamento e che, in quanto umanista digitale, è in grado di fornire alla comunità di riferimento tutti gli strumenti più adatti per fronteggiare le "magnifiche sorti e progressive" della contemporaneità.

## BIBLIOGRAFIA

1. Alberani, Vilma. *Le pubblicazioni ufficiali italiane*. Roma: Associazione Italiana Biblioteche, 1995.
2. Antonelli, Lucia. «Pubblicazioni on line della Pubblica Amministrazione e accesso aperto: stato dell'arte di una "casa divisa"». *AIDA Informazioni: rivista di Scienze dell'informazione* 26, n. 3/4 (2008): 131–144.
3. Ballestra, Laura. *Information literacy*. Roma: Associazione Italiana Biblioteche, 2020.
4. Bolt, Nancy. *Guidelines for Libraries of Government Departments*. A cura di Suzanne Burge. The Hague: IFLA Headquarters, 2008. <https://www.ifla.org/files/assets/hq/publications/professional-report/106.pdf>.
5. Bulgarelli, Sandro, e Madel Crasta. *Le biblioteche dell'amministrazione centrale dello Stato italiano*. A cura di Patrizia Valentini. Roma: Associazione Italiana Biblioteche, 1990.
6. Cavalieri, Piero. *Documenti e dati pubblici sul web: guida all'informazione di fonte pubblica in rete*. A cura di Fernando Venturini. Bologna: Il Mulino, 2004.
7. Faloci, Silvia, a c. di. *Documentazione di fonte pubblica. Le biblioteche come servizio per gli amministratori locali*. Roma: Associazione Italiana Biblioteche, 2009.
8. Lana, Maurizio. *Introduzione all'information literacy. Storia, modelli, pratiche*. Milano: Editrice Bibliografica, 2020.
9. Melis, Guido. *Le biblioteche dell'amministrazione centrale in Italia: contributi al dibattito e indagini sui siti web e sulla professione*. A cura di Fernando Venturini. Roma: Associazione Italiana Biblioteche, 2004.
10. ———. «Passato, presente e futuro delle biblioteche dell'amministrazione pubblica». In *Percorsi e luoghi della conoscenza. Dialogando con Giovanni Solimine su biblioteche, lettura e società*, a cura di Giovanni Di Domenico, Giovanni Paoloni, e Alberto Petrucciani. Milano: Editrice Bibliografica, 2016.
11. Merloni, Francesco, a c. di. *L'informazione delle pubbliche amministrazioni*. Rimini: Maggioli, 2002.
12. Pochesci, Laura. «Le biblioteche ministeriali: da strumenti dell'amministrazione a biblioteche pubbliche». *Le Carte e la Storia* 9, n. 1 (2003): 118–121.
13. Rocchi, Andrea. *Nuovi modelli organizzativi nella pubblica amministrazione dall'e-government all'open data*. Soveria Mannelli: Rubettino, 2015.
14. Roncaglia, Gino. «E-democracy: l'importanza della documentazione nella negoziazione redazionale». *Bollettino AIB (1992-2012)* 48, n. 2–3 (2008): 213–224. <https://bollettino.aib.it/article/view/5523/5268>.
15. Schreibman, Susan, e Ray Siemens. *A companion to digital humanities*. A cura di John Unsworth. Oxford: Blackwell, 2004. <http://www.digitalhumanities.org/companion/>.
16. Venturini, Fernando. «Le biblioteche istituzionali fra isolamento e nuove opportunità». *Economia della Cultura* 13, n. 3 (2003): 329–334.
17. ———. «L'editoria pubblica in rete tra autonomia e controllo: dal "feudalesimo" alla "federazione"?» *AIB Studi* 58, n. 2 (2018): 225–39.

# The DCH in the Italian cultural system. The necessity of a change

Mauro De Bari<sup>1</sup>, Nicola Barbuti<sup>2</sup>

<sup>1</sup> Università di Bari – mauro.debari [at] uniba.it

<sup>2</sup> Università di Bari – nicola.barbuti [at] uniba.it

M. De Bari edited the Par. 2 and the References; N. Barbuti edited the Par. 1.

## ABSTRACT

Digital Transformation is changing the daily life of the communities, but a clear definition of what can be defined as Digital Cultural Heritage (DCH) still misses. Moreover, in this period, Cultural Institutions (CI) have been under pressure because of the COVID-19 pandemic. This paper proposes a digital strategy based on a central role of the cultural heritage users. To this goal, a first-rate need is to hire new professionals provided with digital knowledge and skills.

## KEYWORDS

Digital Cultural Heritage (DCH), Cultural Institutions (CI), COVID-19

## TALK

### 1 INTRODUCTION

The recent COVID-19 pandemic has modified worldwide the relation between people and cultural heritage, increasing the use of digital media to supply the lack of physical enjoyment. The full cultural ecosystem is still inaccessible: museums, archives, libraries, archaeological and monumental sites are closed because they cannot guarantee safe access to their places. Moreover, they haven't planned useful digital strategies or programs for fostering the enjoyment of heritage.

Recently, several purposes in the digital cultural field have been hypothesised. The Italian cultural system is living a quick change, and we need to understand how CI could address the digital challenges fostering the creation of DCH.

But the question is: does DCH exist? In the last years, a first proposal for defining what we could identify as cultural among the digital floating "magma" have been drafted as follows: "*Digital Cultural Heritage is the ecosystem of processes, Born Digital and Digitized entities, virtual phenomena certified and validated as meeting the requirements in since their genesis and over time, which will become testimonies, manifestations and expressions of the evolutionary processes that identify and connote each community, socio-cultural context, simple or complex ecosystem of the Digital Age, assuming the function of memory and source of knowledge for future generations*" [2].

This definition claims the topic of a recent Horizon Call: "*So far, digitisation focused mainly on capturing the visual appearance of individual objects, collections or sites. There is a real need to establish a*

*comprehensive picture of the studied assets, capturing and re-creating not only visual and structural information, but also stories and experiences (stored in language data), together with their cultural and socio-historical context, as well as their evolution over time”*.<sup>1</sup>

Digital Humanities (DH) too must address the challenges of digital preservation and the opportunities coming from CI. The 2019 founded *Istituto Centrale per la digitalizzazione del Patrimonio Culturale*<sup>2</sup> has not suggested a useful strategy that guarantees an integration between heritage and digital processes. The modern society should choose a new approach to tackle the challenge of how best to convey the DCH. The goal of institutions is to ponder an innovative emotional offer [1]: the new DCH connected with the tradition. The key is not to destroy the old cultural systems, but to develop processes that unify tradition, innovation and long-term necessities. As it is argued in “*Musei del Futuro*”,<sup>3</sup> in the Italian museums tradition the digitization lacks because a standard that organizes different realities doesn't exist. Their main issue is the necessity of a ubiquitous culture in which the digitization plays a fundamental role, and this needs to improve the National higher education system for providing the future generation with digital culture.

## 2 THE ITALIAN DISCOVERY OF THE DIGITAL

The cultural ecosystem needs processes to facilitate the production of digital cultural artefacts. During the pandemic, many cultural Institutions attempted to offer digital services aimed at encouraging the enjoyment of their CH. Italian Institutions have offered several proposals to promote something they consider similar to DCH<sup>4</sup>. However, several issues have emerged because of the conclamant confusion existent in terms of digitization.

Digital artefacts created by augmented, mixed and immersive reality processes are rare. What is considered innovative and avant-garde in Italy, in other Countries is practice. Still, the legitimate question is whether these digital cultural services offered by galleries, libraries, museums which have received an enormous consensus during this peculiar context, can receive the same attention at the end of the closure imposed by law? One of the hypotheses born from the MIBACT was the creation of a *Netflix of culture*<sup>5</sup>. However, the idea of payment for digital cultural services could be a winning action?

What will happen when the closing of CI will end? Users will have the opportunity to choose between analogue and digital content: is it sure that Italian users will choose innovation?

A first answer has been offered by the analysis of “*Soluzioni museali*”<sup>6</sup>.

The organization conducted an open-ended and multiple-choice survey using social platforms. The results show that today is still difficult to build customers loyalty to online offers. Assuming that users are familiar with the digital, they request a more stimulating and engaging offer, especially if a payment is required.

---

1 <https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/opportunities/topic-details/dt-transformations-12-2018-2020>

2 <https://www.beniculturali.it/ente/istituto-centrale-per-la-digitalizzazione-del-patrimonio-culturale-digital-library>

3 [http://www.project-musa.eu/it/wp-content/uploads/sites/5/2018/03/Musei-del-futuro\\_Mu.SA-it.pdf](http://www.project-musa.eu/it/wp-content/uploads/sites/5/2018/03/Musei-del-futuro_Mu.SA-it.pdf)

4 <https://www.tribune.com/arti-visive/arte-contemporanea/2020/03/forestoacasa-musei-covid-19-contenuti-online-guida/>

5 [https://www.ansa.it/sito/notizie/politica/2020/04/18/franceschini-pensiamo-a-una-netflix-della-cultura\\_ea51312e-43dc-4812-81d1-42518db342c7.html](https://www.ansa.it/sito/notizie/politica/2020/04/18/franceschini-pensiamo-a-una-netflix-della-cultura_ea51312e-43dc-4812-81d1-42518db342c7.html)

6 [https://drive.google.com/drive/folders/1JiEnYf0GChT3p5aMEPBgKW6HnHZN\\_6Tx](https://drive.google.com/drive/folders/1JiEnYf0GChT3p5aMEPBgKW6HnHZN_6Tx)

The Italian offer is limited in terms of production and processes used, except for some rare cases. An example based on innovation and user interaction is offered by the project “*Crowddreaming. An Erasmus+ project about digital cultural heritage*”<sup>7</sup>. The users experiment on multiple levels, starting from the interaction of young people with the practice of Crowddreaming and Museater methodologies, meshing the museum and the theatre in an original way of conceiving the DCH [3].

The idea of community lacks in Italian museums: the user doesn’t feel himself like a part of a bigger system. Italians try to be *prosumer* [4], but the actual state-of-the-art doesn’t permit it. The audience has a marginal role in terms of consideration. Sometimes, CI are synonymous with elitist knowledge [5]: most of the public museums, libraries and archives open more for academic finalities rather than to offer a good user experience. The Director of the *Reggia di Caserta* stated that the primary function of the museums is the academic research activity.<sup>8</sup> This dominant conception has led to underestimating the importance of new technologies,<sup>9</sup> implementing the inequality of knowledge in terms of DCH.

Institutions must improve their offer with a multifaceted staff with digital abilities. In this, the digital and its processes can be facilitators. It is crucial to understand which abilities must learn the professionals that will work inside CI to avoid the externalization of digital work outside their realities. Recently, the Department of Humanities DISUM of University of Bari “A. Moro”, in collaboration with important international partners, has launched the BIBLIO<sup>10</sup> project. The project aims to identify two different levels of professionals with digital capacities in the librarian field. The COVID-19 pandemic has confirmed that the digital is a need more than before. The cultural system should promote a flexible strategy to improve those realities that try to invest in the digital. As BIBLIO project is researching good practices to promote the training of professionals, Italian CI should activate academics and researchers to ponder a competent category of digital cultural professionals capable of maximizing the productivity, the attractivity and the professionalism in digital Italian Institutions processes. This solution will influence the user category too, that will be a functional part of the cultural system.

---

7 <https://crowddreaming.eu/it/>

8 [http://www.project-musa.eu/it/wp-content/uploads/sites/5/2018/03/Musei-del-futuro\\_Mu.SA-it.pdf](http://www.project-musa.eu/it/wp-content/uploads/sites/5/2018/03/Musei-del-futuro_Mu.SA-it.pdf) p. 23

9 [http://www.project-musa.eu/it/wp-content/uploads/sites/5/2018/03/Musei-del-futuro\\_Mu.SA-it.pdf](http://www.project-musa.eu/it/wp-content/uploads/sites/5/2018/03/Musei-del-futuro_Mu.SA-it.pdf) p. 12

10 <https://www.biblio-project.eu>



## REFERENCES

1. Alelis, Genevieve, Ania Bobrowicz, and Chee Siang Ang. 'Comparison of Engagement and Emotional Responses of Older and Younger Adults Interacting with 3D Cultural Heritage Artefacts on Personal Devices'. *Behaviour & Information Technology* 34, no. 11 (2015): 1064–1078. <https://kar.kent.ac.uk/50319/1/Comparison%20of%20engagement%20and%20emotional%20responses%20of%20older%20and%20younger%20adults%20interacting%20with%203D%20cultural%20heritage.pdf>.
2. Barbuti, Nicola. 'Thinking Digital Libraries for Preservation as Digital Cultural Heritage: By R to R 4 Facet of FAIR Principles'. *International Journal on Digital Libraries*, 2020, 1–10. <https://link.springer.com/article/10.1007/s00799-020-00291-7>.
3. Barbuti, Nicola, Giuliano De Felice, Annalisa Di Zanni, Paolo Russo, and Altheo Valentini. 'Creating Digital Culture by Co-Creation of Digital Cultural Heritage: The Crowddreaming Living Lab Method'. *Umanistica Digitale* 9 (2020): 19–34. [http://aiucd2019.uniud.it/wp-content/uploads/2019/01/BoA-2019\\_PROVV.pdf](http://aiucd2019.uniud.it/wp-content/uploads/2019/01/BoA-2019_PROVV.pdf).
4. Bravi, Laura, Federica Murmura, and Gilberto Santos. 'Attitudes and Behaviours of Italian 3D Prosumer in the Era of Additive Manufacturing'. *Procedia Manufacturing* 13 (2017): 980–986. [https://www.researchgate.net/publication/320243961\\_Attitudes\\_and\\_behaviours\\_of\\_Italian\\_3D\\_prosumer\\_in\\_the\\_Era\\_of\\_Additive\\_Manufacturing](https://www.researchgate.net/publication/320243961_Attitudes_and_behaviours_of_Italian_3D_prosumer_in_the_Era_of_Additive_Manufacturing).
5. Visser, Anna Maria. 'Quale Pubblico per i Musei Italiani?' *Archeologia Viva*, 2016.

# Informatica Umanistica e Cultura Digitale. La sfida epistemologica

Dino Buzzetti<sup>1</sup>, Antonella De Ninno<sup>2</sup>, Domenico Fiormonte<sup>3</sup>

<sup>1</sup> Università di Bologna – dino.buzzetti [at] gmail.com

<sup>2</sup> ENEA TECFIS-DIM – antonelladeninno [at] enea.it

<sup>3</sup> Università di Roma Tre – domenico.fiormonte [at] uniroma3.it

## SINTESI

This paper examines the relationship between the two main components of the scientific practice of the members of AIUCD, i.e. Informatica Umanistica and Cultura Digitale. We see these two components as complementary and not as opposing fields. In this view, our goal is to propose a common epistemological ground between them, based on the new scientific paradigm emerged in the development of quantum mechanics. We envisage a plausible connection between the availability of new indeterminist types of formalisation and the indeterministic paradigm that emerges from quantum mechanics. After a brief characterisation of the new paradigm and our experience in interdisciplinary dialogues, their impact on the social sciences and humanities and their cultural sway on the digital ecosystem are exposed. The new approaches to formalisation and modelling ensuing from the connectionist turn in the Artificial Intelligence are then examined and shown to be consistent with the new epistemological paradigm. We argue that an a priori refusal of formalisation in the humanities, and in the analysis of their interpretive practices, would still be subordinate to the old superseded paradigm and would seriously hinder the overcoming of the “two cultures” divide.

## PAROLE CHIAVE

informatica umanistica, cultura digitale, paradigma epistemologico, fisica quantistica, studi testuali, svolta connessionista in AI, formalizzazione, interpretazione

## INTERVENTO

### 1 INTRODUZIONE

Il tema del convegno pone il problema degli effetti sociali ed economici delle pratiche di ricerca in campo umanistico che si rivolgono al digitale, un problema importante che interroga la comunità di studiosi che si riconosce in questa Associazione sull'identità che la tiene unita e sugli indirizzi della propria attività scientifica.

Il presente contributo riunisce alcune esperienze diverse, l'esperienza del progetto internazionale New Humanities<sup>1</sup> e quella di un gruppo di studiosi che si sono da tempo impegnati in una riflessione sui metodi computazionali applicati alla ricerca nelle discipline umanistiche. Pur nelle differenze, anche profonde, che sussistono, entrambe le esperienze condividono l'obiettivo di una "new epistemic for a new paideia" ([3], 416), cioè il comune fondamento epistemologico su cui si basa la ricerca metodologica delle due ragioni sociali della nostra Associazione, l'Informatica Umanistica e la Cultura digitale—un fondamento epistemologico che configura questa base comune come una meta-riflessione sulle pratiche di ricerca multidisciplinari e sull'applicazione di metodi computazionali nelle discipline dell'area umanistica. Lo scopo della nostra proposta è dunque duplice: da un lato illustrare come la fisica quantistica possa fornire una cornice epistemologica per le scienze umane, dall'altro usare questa stessa cornice per ripensare l'idea di formalizzazione e rappresentazione computazionale dei fenomeni storico-letterari.

## 2 INDETERMINISMO E SCIENZE UMANISTICHE

Il modello ermeneutico delle discipline umanistiche è di per sé indeterministico—a parte la pretesa di "oggettività" di Wilhelm Dilthey ([14], cfr. *infra*, § 3). Gli scambi interdisciplinari o multidisciplinari tra le discipline umanistiche e le scienze naturali sono resi possibili solo dalla comune accettazione del paradigma indeterministico, senza la quale non si supererebbe la separazione tra le "due culture" [24]. Gli umanisti che si oppongono alla formalizzazione sono implicitamente subalterni al modello classico perché considerano ogni tipo di formalizzazione deterministico e non ermeneutico, mentre gli studiosi di scienze naturali, amanti del rigore formale—spesso inteso anche da loro secondo il modello deterministico classico—hanno difficoltà ad accettare un approccio ermeneutico che non si presenti come procedura formale, benché anch'esso formalizzabile come procedura analoga a quella dei modelli connessionisti di tipo adattativo (cfr. *infra*, § 3). Dunque, a nostro parere, il paradigma indeterministico costituisce la base comune per fondare la pratica interdisciplinare e superare la separazione tra le "due culture". Solo su questa base comune, ermeneutica e formalizzazione possono non venire più considerate come pratiche antitetiche e tra loro inconciliabili.

## 3 IL PARADIGMA EPISTEMOLOGICO DELLA FISICA QUANTISTICA

Uno degli obiettivi principali del progetto New Humanities sin dalla sua creazione (2011) è stato quello di elaborare una nuova cornice interdisciplinare che fornisse un terreno di incontro e dialogo fra le discipline umanistico-sociali e le scienze della natura. Nel 2013 fu organizzato un seminario con Emilio Del Giudice, Mauro Bergonzi, Domenico Fiormonte, Massimo Marraffa e Michele Lucantoni che affrontava il problema della coscienza in relazione alla fisica quantistica [13]. Purtroppo, con la scomparsa di Del Giudice nel 2014, non è stato possibile proseguire questo lavoro. Tuttavia oggi ci pare più che mai necessario riannodare i fili di quel dialogo, proprio per mostrare come, a partire dalla fisica quantistica, sia possibile elaborare una "epistemologia ponte", ovvero mostrare i limiti (e i danni) epistemologici della divisione fra le "due culture" che ha prodotto l'attuale e divisivo assetto delle scienze.

---

1 [www.newhumanities.org](http://www.newhumanities.org)

La visione riduzionista della scienza, che considera i sistemi complessi come somma delle parti che li compongono e delle loro relazioni di “corto raggio”, non è affatto basata sull’evidenza scientifica ma su un pregiudizio ontologico e profondamente ideologico. In particolare, la meccanica quantistica introduce una visione della realtà nella quale viene a mancare il concetto di “corpo isolato”. È del tutto irrealistico descrivere un oggetto o una azione a prescindere dalla considerazione di tutte le interazioni che si sviluppano nell’ambiente che li ospita. In linea di principio non sarebbe lecito neanche limitarsi alle interazioni con i corpi o i campi più prossimi perché la descrizione corretta deve tenere in conto qualsiasi interazione anche della più debole intensità. La grande maggioranza degli eventi che ci coinvolgono può essere comodamente spiegata considerando un numero limitato di interazioni, quindi la fisica classica continua ad essere valida in moltissimi casi; tuttavia un numero non piccolo di eventi significativi rimangono del tutto incomprensibili a meno di abbandonare il pregiudizio del “corpo isolato”.

L’approccio riduzionista presenta l’ovvio vantaggio di una immediata semplificazione del processo conoscitivo: ogni complessità è ridotta fino alla sua prima interpretazione nota. Abbiamo così l’effimera sensazione di padroneggiare i sistemi complessi attraverso la conoscenza dei loro componenti anche se il processo di “ricostruzione” della complessità spesso rimane fuori dall’interpretazione della realtà.

La rivoluzione epistemologica della meccanica quantistica ci presenta un Universo in cui il processo riduzionista perde ogni senso a causa della non separabilità dei corpi e della loro intrinseca fluttuabilità per cui anche la distinzione tra il corpo e il suo movimento perde di significato. Questo approccio rende impossibile o quasi, separare il messaggio dal contesto ed apre uno scenario rivoluzionario ma non inedito nel processo conoscitivo. L’indeterminismo della meccanica quantistica, quindi, lungi dall’essere solo una precisa e formalizzata relazione matematica tra osservabili, ci comunica una “qualità” della realtà, anzi, forse la più importante delle “qualità”: la sua non-separabilità dall’osservatore. Tutto ciò non implica affatto l’inconoscibilità della realtà quanto piuttosto rappresenta un forte suggerimento a diffidare del processo riduzionista che, in cambio di una apparente semplificazione della descrizione porta inevitabilmente a perdere informazioni importanti con il risultato di scambiare la realtà con il suo toy-model.

Emilio Del Giudice, in un dialogo con Felice Accame sulla teoria della conoscenza, affermava: «Una famosa frase di Cartesio afferma che tanto meglio conosciamo un oggetto quanto più lo dividiamo in parti. Questo è il sogno dell’atomismo. Se riesco a sbriciolare una persona e a numerare uno dopo l’altro tutti gli atomi di cui è fatta, avrò una conoscenza perfetta di questa persona. Ma se qualcuno mi mostra tutti gli atomi di cui è composto un essere umano, che qui chiameremo Concettina, questi non mi dicono se Concettina è una donna allegra o triste; se è una donna che si commuove oppure no; se è bella o brutta» [10]. Il modo di pensare e “conoscere” degli umanisti è prevalentemente incline a considerare gli oggetti di studio come sistemi interconnessi e non isolatamente, nel modo in cui vengono considerati nella fisica classica. Anche per questo le scienze umanistico-sociali non possono rimanere insensibili alla rivoluzione epistemologica della fisica quantistica. Ma quali “applicazioni” è possibile immaginare alle specifiche metodologie di ricerca umanistica?

Le possibili ricadute del modello epistemologico basato sulla meccanica quantistica di fatto sono molteplici. Per esempio, nelle scienze del testo, in particolare nelle discipline che studiano la fluidità testuale, sia a livello di tradizione (diacronia), sia a livello del singolo manufatto (sincronia), si potrebbe pensare a modelli interpretativi (ermeneutica e critica letteraria) e ricostruttivi (filologia) che includano e riflettano la dinamicità dei processi di scrittura, non considerando più il testo come “corpo isolato” (il testo illusoriamente fisso della stampa gutemberghiana), ma come un processo culturale che agisce nell’ambiente e ne viene continuamente influenzato e trasformato [1].

## **4 LA SVOLTA CONNESSIONISTA E IL PROBLEMA DELLA FORMALIZZAZIONE**

Il tema della formalizzazione risulta centrale per la comprensione del rapporto tra Informatica Umanistica e Cultura Digitale. Furono il «programma logicista» proposto di Jean-Claude Gardin [15] e il suo «inevitabile riferimento all’epistemologia» [16] a ribadire l’importanza decisiva della formalizzazione per l’applicazione di procedure computazionali alla ricerca umanistica. Il procedimento di formalizzazione concepito da Gardin si fondava sulle cosiddette «schematizzazioni» (un concetto introdotto dal logico svizzero Jean-Blaise Grize), ossia su quell’«esercizio che mira a isolare le operazioni cosiddette ‘di logica naturale,’ correntemente praticate nell’argomentazione del linguaggio ordinario» (31; cfr. [18]); tali «esercizi di messa in forma logica» (31) sono necessari, secondo Gardin, per costruire, attraverso una «serie di operazioni» di questo tipo, inferenze affidabili che conducano dai «dati messi in campo per fondare le ipotesi» alle ipotesi stesse, «o viceversa» ([15], 68). Inoltre, per ciò che qui importa, Gardin sosteneva che «queste due componenti», i dati che fondano le ipotesi e le schematizzazioni, «corrispondono rispettivamente alle basi di fatti e alle basi di regole con cui vengono alimentati i motori di inferenza nei sistemi esperti» (68).

Ma sta proprio nel riferimento ai ‘sistemi esperti’ il limite della proposta di Gardin, che con ciò finiva per risultare subalterna a un modello di formalizzazione deterministico. L’adozione dello strumento computazionale nasceva da una precisa scelta metodologica, che si fondava sull’«omologia» tra l’«architettura» dei sistemi esperti e quella delle schematizzazioni [16]. Secondo Gardin, il ricorso ai sistemi esperti si fondava sul «valore aggiunto» presumibilmente ottenuto, «sul piano epistemologico» (37), dall’applicazione dalle «basi di regole dei sistemi esperti» nel sottrarre «le operazioni dell’interpretazione» dei dati «agli oscuri percorsi del libero arbitrio e dell’intuizione» [15]. Con l’applicazione di tali regole di inferenza Gardin restava però legato al modello classico del ragionamento deterministico ipotetico-deduttivo, che è proprio delle scienze naturali.

Va detto tuttavia che lo stesso Gardin non nascondeva le sue «riserve» sull’«aspetto meccanicistico» (70) del «paradigma del calcolo, nell’accezione informatica del termine» (68) e non trascurava un «secondo aspetto dell’approccio logicista», che ammette «la molteplicità delle interpretazioni» e legittima la dimensione «ermeneutica» della ricerca in campo umanistico (74), legata a suo giudizio all’analisi di «contesti particolari» (75). Ciò detto, si deve però osservare che la legittimazione del criterio ermeneutico non è di per sé sufficiente a superare il punto di vista deterministico e la separazione tra le cosiddette ‘due culture.’

Questa separazione viene addirittura teorizzata da Dilthey con la celebre distinzione tra le *Naturwissenschaften* e le *Geisteswissenschaften*, che si fonda sull'affermazione della radicale 'differenza' dei loro rispettivi metodi. Hans Georg Gadamer ha infatti messo in luce l'«intima incoerenza del pensiero di Dilthey», il cui sforzo per «comprendere» e «giustificare» le scienze umane facendo riferimento direttamente alla vita e all'esperienza vissuta, «rimane legato» a un «concetto di oggettività» che è quello a cui tendevano le spiegazioni deterministiche delle scienze naturali del tempo. Gadamer fa risalire l'origine di quest'idea di oggettività al «concetto cartesiano di scienza, del quale», a suo giudizio, Dilthey «non era capace di disfarsi» [14]. L'ancoramento al modello classico delle scienze naturali da un lato conduce Gardin a concepire in modo deterministico il concetto di formalizzazione e, dall'altro, porta Dilthey ad assegnare alla pratica ermeneutica, come fine ultimo, il raggiungimento di un tipo di oggettività identico a quello delle scienze della natura. Dilthey quindi non si accorge, come ribadisce il fisico Hugh Everett, dal punto di vista quantistico, «che ogni teoria fisica è sostanzialmente solo un modello per il mondo dell'esperienza». e che «si deve rinunciare a ogni speranza di trovare qualcosa come 'la teoria corretta'». La gran parte dei fenomeni che ricade sotto la nostra esperienza ammette la corrispondenza con vari modelli e, limitatamente alla completa conoscenza di un certo numero di fatti, ciò vale anche per le teorie: questo è il caso della fisica classica che si limita alla descrizione del mondo macroscopico (e sub-luminare) in cui «il modello tende a venire identificato con la 'realtà' stessa e la natura di modello della teoria viene oscurata» [12]. Il processo di costruzione della 'teoria corretta' richiede, al limite, una conoscenza infinita della realtà e pertanto sfugge alla portata del ragionamento logico-deduttivo, ragion per cui, quando si vuole descrivere fenomeni e processi che includono fatti non conosciuti o conoscibili la scelta è tra la rinuncia alla descrizione o il ricorso al pensiero analogico. Oggi, infatti, abbandonato il modello classico, si sono affermate nuove possibilità di formalizzazione in grado di rimuovere gli «ostacoli epistemici (epistemic barriers)» [2] che non permettono di accantonare la separazione tra le «due culture (two cultures)» (416). Il caso più rilevante, soprattutto per le conseguenze che produce nella pratica stessa delle Digital Humanities, è costituito dalla recente svolta connessionista dell'Intelligenza Artificiale, che rende possibili nuove procedure euristiche nell'analisi dei dati. L'impiego di architetture a reti neurali ha favorito l'adozione di nuove forme di inferenza che hanno permesso di superare i limiti del modello deduttivo su cui si fondavano i processi di elaborazione dell'informazione messi in opera dai sistemi esperti. Infatti, i «sistemi esperti», costruiti come motori di inferenza per basi di conoscenza specializzate, «producevano solo risultati molto limitati» rispetto a quanto era possibile ottenere con i metodi connessionisti che venivano proposti da «scuole di pensiero» considerate fino a quel momento «eterodosse e devianti»; il proposito di «arricchire ad infinitum repertori di regole esplicite per tener conto delle migliaia di sottigliezze della percezione, del linguaggio e del ragionamento umano è apparso sempre più come un compito impossibile, irragionevole e inefficace [7; 11]» [6].

In sostanza, occorre nuove procedure in grado di tener conto della presenza di contesti diversi e non classificabili a priori. Tuttavia, non solo la quantità delle connessioni ma, soprattutto la loro "solidità" sono fondamentali nei processi di elaborazione delle informazioni [9]. In condizioni di forte correlazione tra i nodi di una rete emerge un regime altamente coeso la cui insorgenza è prevedibile analiticamente. Reti che interagiscono debolmente presentano una risposta agli stress simile a quella di un solido (rigidità, fragilità), mentre un sistema in regime di coesione si comporta come un gel (plasticità, resilienza).

Da un punto di vista euristico, però, l'invenzione di nuove «macchine induttive»—come sono chiamate da Cardon et al.—portava a «definizioni sensibilmente diverse dell'intelligenza, del ragionamento e della predizione» che possono venire descritte non tanto come genericamente induttive, ma più propriamente come forme di abduzione, ossia come quel tipo di inferenza, descritto da Peirce<sup>2</sup>, che porta a concludere che un certo caso preso in esame dipende da una certa ipotesi sulla base di altri risultati già noti. Analogamente, nell'ambito di quella che viene chiamata «ingegneria della conoscenza» [25] l'abduzione è stata riconosciuta come un tipo di inferenza «per collegare le prove alle ipotesi» [26] e in effetti l' 'apprendimento automatico supervisionato' ottenuto da macchine con «architetture a reti neurali» (438) è teso a verificare che certi dati osservati siano conformi a determinate ipotesi.

Un aspetto cruciale dell'inferenza abduttiva comporta il controllo sul processo inferenziale in cui consiste e quindi è dotato di portata autoreferenziale. Secondo Peirce, ogni inferenza implica un «controllo cosciente dell'operazione» [19] e quindi, sul piano logico, una relazione metalinguistica intrinsecamente autoreferenziale tra una conseguenza e la regola, o *leading principle* [21], da cui dipende. Su questa base, l'introduzione delle reti neurali ha reso possibile l'implementazione di modelli cosiddetti 'adattativi' per l'analisi dei processi interni ai sistemi naturali di auto-organizzazione, quali i processi biologici e quelli propri di tutte le attività umane come, per esempio, l'attività linguistica. L'approccio connessionista tiene espressamente conto di questo aspetto euristico essenziale, teorico e autoreferenziale, del processo di inferenza e lo fa «facendo risiedere il successo della previsione nel mondo stesso, erinno-vando così la promessa adattativa delle macchine riflessive della cibernetica: fare sistema con l'ambiente da calcolare per implementare un nuovo tipo di cicli di retroazione» o *feedback* [6].

I sistemi artificiali adattativi che si basano su questi principi sono sistemi che contengono «regole che determinano le condizioni di possibilità di altre regole» [3]. Infatti, nuove connessioni, o connessioni non considerate precedentemente come rilevanti, che si producono nel corso del tempo tra gli elementi, le strutture e i processi locali dell'insieme dei dati contestuali al fenomeno considerato, rendono necessaria la trasformazione delle ipotesi generali del sistema. Tali connessioni «permettono ai modelli di generare regole in modo dinamico» (56) e «trasformano abduttivamente i principi regolatori sovraordinati da cui dipende il funzionamento del sistema» [5]. Ciò rende i modelli computazionali ricavati da sistemi adattivi di questo tipo «strumenti adeguati per la ricostruzione formale dei processi interpretativi del testo» [4].

Infatti, la natura indeterminata della relazione tra la sintassi e la semantica nel linguaggio naturale dà origine al ciclo autoreferenziale di regole di ordine superiore che permettono l'introduzione di nuove regole all'interno del sistema. Per finire, basti qui ricordare, come esempi concreti di applicazione di modelli adattativi, i lavori di «semantica vettoriale» di Michael Gavin [17] per lo studio dei contesti rilevanti nei processi di interpretazione, o i risultati ottenuti da di Peter de Bolla [8] nel campo della cosiddetta «conceptual history».

---

2 Secondo Peirce, l'induzione, che inferisce una «regola» generale da un «caso» e da un «risultato» ripetuto un certo numero di volte, «non è l'unico caso di inversione di un sillogismo deduttivo (quello che inferisce un «risultato» da una «regola» e da un «caso»), perché esiste anche «l'inferenza di un caso da una regola e da un risultato», e «questo tipo di inferenza viene chiamato fare un'ipotesi» [20], ovvero «abduzione», cioè un tipo di inferenza che «per la stessa definizione del termine conduce a un'ipotesi» [22].

## 5 EPILOGO. PER UN SUPERAMENTO DELLA SEPARAZIONE DELLE “DUE CULTURE” (E DELLE DUE COMPONENTI, IU E CD, DELLA NOSTRA ASSOCIAZIONE)

L'assenza di integrazione tra le due componenti delle Digital Humanities, l'Informatica Umanistica e la Cultura Digitale, un'integrazione senza la quale ne verrebbe trascurata la dimensione teorica che è di natura essenzialmente epistemologica, come si è cercato di argomentare, le porterebbe a svolgere una funzione essenzialmente subalterna, prevalentemente comunicativa e ancillare rispetto alle scienze sociali, da un lato, e alle scienze formali e computazionali dall'altro.

### BIBLIOGRAFIA

1. Benozzo, Francesco. *Etnofilologia. Un'introduzione*. Napoli: Liguori, 2010.
2. Bergonzi, Mauro, Francesco Fiorentino, Domenico Fiormonte, Laura Fortini, Ugo Fracassa, Michele Lucantoni, Massimo Marraffa, e Teresa Numerico. «The New Humanities Project—Reports from Interdisciplinarity». *Humanities* 3, n. 3 (settembre 2014): 415–41. <https://doi.org/10.3390/h3030415>.
3. Buscema, Massimo. «The General Philosophy of Artificial Adaptive Systems (AAS)». In *ARCHEOSEMA: Artificial Adaptive Systems for the Analysis of Complex Phenomena: Collected Papers in Honour of David Leonard Clarke*. Borgo San Lorenzo: All'insegna del Giglio, 2014.
4. Buzzetti, Dino. «Alle origini dell'Informatica Umanistica: Humanities Computing e/o Digital Humanities.» *Il museo virtuale dell'informatica archeologica*, Rendiconti dell'Accademia Nazionale dei Lincei, Classe di Scienze morali, storiche e filologiche, 9, n. 30 (2019).
5. ———. «Dell'ineluttabile 'paradossalità' del testo: la conferma digitale». In *Letteratura e altre rivoluzioni: per Raul Mordenti*. Roma: Bordeaux, 2020.
6. Cardon, Dominique, Jean-Philippe Cointet, e Antoine Mazières. «La revanche des neurones». *Reseaux* n° 211, n. 5 (16 novembre 2018): 173–220. <https://www.cairn.info/revue-reseaux-2018-5-page-173.htm?ref=doi>.
7. Collins, Harry M. *Artificial Experts: Social Knowledge and Intelligent Machines*. Cambridge: MIT Press, 1990.
8. De Bolla, Peter. *The Architecture of Concepts: The Historical Formation of Human Rights*. New York: Fordham University Press, 2013.
9. Dehmamy, Nima, Soodabeh Milanlouei, e Albert-László Barabási. «A Structural Transition in Physical Networks». *Nature* 563, n. 7733 (novembre 2018): 676–80. <https://doi.org/10.1038/s41586-018-0726-6>.
10. Del Giudice, Emilio. *L'anima passionale della ragione scientifica*. Milano: Biblion, 2019.
11. Dreyfus, Hubert L. «Why Heideggerian AI failed and how fixing it would require making it more Heideggerian». *Artificial Intelligence* 18, n. 171 (2007).
12. Everett, Hugh III. «The Theory of the Universal Wave Function.» In *The Many-worlds Interpretation of Quantum Mechanics*. Princeton: Princeton University Press, 1973.
13. Fiormonte, Domenico. *La coscienza. Un dialogo interdisciplinare e interculturale*. Roma: Istituto di Studi Germanici, 2018.
14. Gadamer, Hans Georg. *Il problema della coscienza storica*. Napoli: Guida, 1988.



15. Gardin, Jean-Claude. «Calcul et narrativité dans les publications archéologiques». *Archeologia e Calcolatori* 10 (1999).
16. Gardin, Jean-Claude, e Maria Novella Borghetti. *L'architettura dei testi storiografici: Un'ipotesi*. Bologna: CLUEB, 1995.
17. Gavin, Michael. «Vector Semantics, William Empson, and the Study of Ambiguity». *Critical Inquiry* 44, n. 4 (2018).
18. Grize, Jean-Blaise. «Logique mathématique, logique naturelle et modèles». In *Formalisierung in den Geistes-wissenschaften / Sciences humaines et formalisation*, 1974.
19. Peirce, Charles Sanders. *Collected Papers (vol. I-VIII)*. Cambridge: Harvard University Press, 1958.
20. ———. «Deduction, Induction, and Hypothesis». *Popular Science Monthly* 13 (1878).
21. ———. «On the Algebra of Logic». *American Journal of Mathematics* 3, n. 1 (1880).
22. ———. «The Proper Treatment of Hypotheses: a Preliminary Chapter, toward an Examination of Hume's Argument against Miracles, in its Logic and in its History». MS [ R ] 692, 1901.
23. Robin, Richard S. *Annotated Catalogue of the Papers of Charles S. Pierce*. Amherst, 1967.
24. Snow, Charles Percy. *The Two Cultures*. New York: Cambridge University Press, 2014.
25. Tecuci, Gheorghe, Dorin Marcu, Michai Boicu, e David A. Schum. *Knowledge Engineering: Building Cognitive Assistants for Evidence-Based Reasoning*. New York: Cambridge University Press, 2016.
26. Tecuci, Gheorghe, David A. Schum, Dorin Marcu, e Mihai Boicu. *Intelligence Analysis as Discovery of Evidence, Hypotheses, and Arguments: Connecting the Dots*. New York: Cambridge University Press, 2016.

# Dalla Digital History alla Digital Public History

Marcello Ravveduto  
Università di Salerno – mravveduto [at] unisa.it

## SINTESI

Dopo una disamina delle definizioni di Digital History (DHy) e Public History (PH), il paper propone una definizione della Digital Public History (DPHy) come ibridazione delle due discipline. La DPHy conferma le caratteristiche della DHy come risposta delle Digital Humanities alla ricerca storica ma allo stesso tempo, in quanto ibrido, è attenta ai processi di uso e comunicazione della storia tanto nel web quanto nei social network coniugando analisi computazionali e ricerche qualitative per re-intermediare il sapere storico negli ambienti disintermediati e interpretare il senso comune della storia dei pubblici interconnessi.

## PAROLE CHIAVE

digital history, public history, digital public history, social network, facebook

## INTERVENTO

Il contributo si propone di tracciare, all'interno delle Digital Humanities, l'identità della Digital Public History (DPHy) rispetto alla Digital History (DHy).

Quest'ultima è un settore di ricerca che usa metodologie e tecniche informatiche finalizzate al trattamento automatico o semiautomatico di dati quantitativi per l'elaborazione di strumenti digitali applicati alla ricerca storica. Il campo d'azione è vastissimo: dalla edizione digitale delle fonti alla creazione di database con eventuale geo-referenziazione; dall'analisi automatica delle concordanze sui testi storiografici alla elaborazione di timeline collaborative; dal digital storytelling alla creazione di archivi tematici condivisi e annotati. Questa stratificazione mostra quanto la DHy sia una disciplina derivata dalle Digital Humanities per rispondere alla diffusione del digitale nella pratica storica.

Nei progetti di DHy l'aggettivo digital ha quasi sempre, però, una valenza duplice: non riguarda solo lo "strumento" informatico, ma anche il mondo digitale come luogo di pubblicazione e condivisione delle risorse, o anche come luogo di raccolta e creazione di risorse utili al discorso storico. In parole povere un progetto di storia digitale trova origine e si realizza nel web acquistando immediatamente una connotazione "pubblica": consente un ampio e diffuso accesso a una conoscenza storica che nel passato era riservata a un numero limitato di persone. Ma basta la presenza nel Web per trasformare una ricerca di DHy in un progetto di Digital Public History (DPHy)? La risposta è negativa. Facciamo un passo indietro allungando lo sguardo verso la Public History (PH).

Potremmo definirla quel campo di azione in cui lo storico non solo insegna o divulga il sapere scientifico ma lo applica ai problemi dibattuti nell'arena pubblica «in contatto diretto con l'evoluzione della mentalità e del senso delle appartenenze collettive» [6]. L'obiettivo del public historian è una «maggiore presenza della storia come fattore interpretativo del presente attraverso una diffusa conoscenza del passato nella società» (Noiret, 2012a). In prima istanza la PH sembra non avere nulla di “digitale”; in realtà si avvale di molteplici supporti mediatici per coinvolgere il pubblico e le «audience interconnesse» (boyd, 2008) nella pratica della storia. La Rete è entrata con prepotenza nella diffusione del sapere storico con siti, tools e app che stanno modificando il rapporto con il passato attraverso la “preponderanza” (o se vogliamo la “liberalizzazione”) delle memorie individuali e collettive. Nel conflitto reale/virtuale tra storia e memoria il public historian si offre come agente di rimediazione del sapere che si muove in diversi ambienti costruendo una storiografia accessibile, creando fonti di comunità, realizzando siti materiali e immateriali che aumentano la consapevolezza della storia nell'arena pubblica al servizio di committenti pubblici e privati [6]. Quando il public historian agisce nella Rete assomiglia al passeur francese: la figura professionale che interconnette il sapere storico con quello informatico [2] e adopera il linguaggio ITC e quello umanistico risolvendo le questioni epistemologiche poste dalla duplice convivenza. Questo è possibile perché la base di partenza non è la tecnica in sé, ma la necessità di dare nuove risposte al bisogno di storia attraverso le infrastrutture cibernetiche (Noiret, 2012b).

Riassumendo: se per DHy si intende la commistione tra Storia e Informatica e per PH il coinvolgimento del pubblico nella interpretazione della Storia, anche tramite la Rete, la DPHy, allora, è quel settore di studi e ricerche della storia digitale orientato alle «audience interconnesse».

Il rischio, tuttavia, è ricondurre i progetti di DHy o DPHy al solo campo storico dimenticando che la Digital History appartiene al campo delle Digital Humanities poiché abbraccia più settori disciplinari di ambito umanistico e informatico. A questo punto bisogna considerare un ulteriore fattore: un progetto di DPHy non è mai solo storia digitale ma anche sempre storia con il digitale. Se da un lato «usa e domina le tecnologie» con «pratiche cognitive specifiche degli storici», dall'altro implementa la ricerca, l'insegnamento e la comunicazione dei risultati scientifici (Noiret, 2019). L'ibridazione con la PH rende la DPHy qualcosa che va oltre l'applicazione alle discipline storiche di «metodologie, strumenti computazionali e tecniche informatiche tese al trattamento automatico o semi-automatico dei dati, i quali vengono visualizzati e restituiti allo studioso attraverso analisi di tipo quantitativo» (Paci, 2019). Nell'esercizio della DPHy lo storico per essere “digitale” non deve necessariamente confrontarsi con analisi di tipo quantitativo; anzi il più delle volte si trova, come un qualsiasi utente ma con una specifica competenza, a fare «storia con il digitale». Esistono numerosi campi in cui il digitale entra nel mestiere di storico senza comportare «un trattamento automatico o semi-automatico dei dati» o, quanto meno, ne costituisce solo una parte. Del resto, DHy e DPHy appartengono al «sistema sociale della Storia» (Ortoleva, 1995): un insieme di parti interconnesse attraverso il digitale che lo storico deve dominare imparando a usare strumenti e medium. Chi fa storia digitale impiegando i social può sfruttare l'analisi dei big data, ma deve innanzitutto conoscere e governare il mezzo: usare i social non significa compiere solo un lavoro computazionale, ma anche veicolare informazioni, interagire con gli utenti e, infine, raccogliere gli esiti della ricerca tanto con la comunità scientifica, quanto con la rete dei contatti. Allo stesso modo, procedere all'edizione digitale di una fonte storica significa non solo conoscere la “codifica dei testi” ma elaborare la “giusta codifica” per il recupero e il confronto delle informazioni al fine di consentire le opportune analisi (Thaller, 1993),

Significa, inoltre, ragionare su come l'utente può o deve accedere all'informazione, quali informazioni veicolare e come, valutando l'influenza che l'edizione avrà sull'interpretazione storica della fonte tanto da parte del singolo quanto dalle comunità interessate (scientifica, locale, tematica ecc.).

Pertanto è lecito domandarsi «se continuare o meno a riferirci, per fare storia digitale oggi al campo – vero ombrello generalista – dell'umanistica digitale» o sia «meglio tradurre l'umanistica digitale nelle singole discipline, per non confondere strumenti, metodi e domande»? Una domanda che induce un'ulteriore riflessione: se da un lato «come storici abbiamo bisogno di creare contenuti» e avere chiara «l'originalità dei nostri metodi, compiti e obiettivi finali nell'ambito del digitale», dall'altro non possiamo farlo utilizzando «strumenti digitali diversi da quelli usati da altri umanisti che promuovono soprattutto l'esegesi, l'analisi e la codifica del testo» (Noiret, 2019).

In altri termini lo storico digitale può riconoscere come “suoi” alcuni strumenti e metodi che condivide con altri umanisti e che rispondono anche ad altre domande oltre quelle dalla ricerca storica. Pensiamo, per esempio, all'analisi del testo: gli strumenti e il metodo sono gli stessi di filologi e linguisti, ciò che cambia è lo sguardo ovvero la necessità di rispondere, in maniera efficace e corretta, alla domanda di storia. Ma questo può accadere, determinando avanzamenti di pratiche, solo se gli storici partecipano, con le loro competenze, a équipe di umanisti digitali. È sulla originalità e specificità della domanda storica – non sugli strumenti – che si dovrebbe insistere, puntando alla collaborazione interdisciplinare confrontandosi con i pubblici sulla base di contenuti non computazionali.

Proviamo ad applicare lo “sguardo” della DPHy a Facebook. Il social network è stato paragonato a una nazione e il suo proprietario, Zuckerberg, a un capo di stato. Una nazione multietnica, multiculturale e multilinguistica. Il cittadino/utente esibisce la sua identità e tutela la sua privacy. Il profilo è lo spazio personale privato, mentre il wall della home page è lo spazio pubblico collettivo. La separazione richiama le caratteristiche delle società borghesi nella definizione del nation building: impegno pubblico (il lavoro e le attività sociali) e momento privato (la famiglia, gli affetti). Le due sfere sono collegate dalla politica ovvero dalla condivisione in pubblico degli ideali e degli interessi personali e di classe, al fine di trovare consenso tra i simili (likes). Questa azione, equiparabile all'attività politica della civiltà industriale, è presente nei social attraverso lo sharing di contenuti che comunica alle networked communities gli ideali e gli interessi personali dell'utente/cittadino, senza più distinzione di classe. Il commento, invece, tende a riprodurre la libertà di parola, così come i gruppi e le pagine fan ricalcano la libertà di associazione. Facebook è una nazione delle nazioni in cui la sfera pubblica è un coagulo esperienziale dove passato e presente si scontrano con storia e memoria e temi globali si intrecciano con valori ultralocali.

È qui che la DPHy può intervenire per analizzare l'esuberanza di un “passato” digitalizzato che riallaccia memorie individuali e collettive alle identità locali e nazionali delle società globalizzate [7]. Sfruttando le potenzialità del social casting e applicando la shared authority (Frisch, 1990) la DPHy può creare una storia condivisa con le «audience interconnesse» aprendo il varco a progetti di DHy non computazionali.

## BIBLIOGRAFIA

1. Cohen, Daniel J., Michael Frisch, Patrick Gallagher, Steven Mintz, Kirsten Sword, Amy Murrell Taylor, William G. Thomas, e William J. Turkel. «Interchange: The promise of digital history». *The Journal of American History* 95, n. 2 (2008): 452–491.
2. Cooper-Richet, Diana, Jean-Yves Mollier, e Ahmed Silem. *Passeurs culturels dans le monde des médias et de l'édition en Europe (XIXe et XXe siècles)*. Villeurbanne: Presses de l'Ensibb, 2005.
3. Criscione, Antonio. «Sopravviverà la storia all'ipertesto?» *Memoria e Ricerca* 12 (2003).
4. Cullen, Charles T., Peter B. Hirtle, David Levy, Clifford A. Lynch, e Jeff Rothenberg. *Authenticity in a Digital Environmen*. Arlington: Council on Library and Information Resources, 2020.
5. Gallini, Stefania, e Serge Noiret. «La historia digital en la era del Web 2.0. Introducción al dossier Historia digital». *Historia Critica* 43 (2011).
6. Noiret, Serge. «“Public History” e “Storia Pubblica” nella Rete». *Ricerche Storiche* 2–3 (2009): 275–327.
7. ———. «Storia pubblica digitale». *Zapruder. Storie in movimento* 36 (2015): 15–17.
8. Salvatori, Enrica. «Digital (Public) History: la nuova strada di una antica disciplina». *RiMe. Rivista dell'Istituto di Storia dell'Europa Mediterranea* 1 (2017): 57–94.
9. Siemens, Ray. «Communities of Practice, the Methodological Commons, and Digital Self-Determination in the Humanities». *Digital Studies/le Champ Numérique* 5 (2016). <http://doi.org/10.16995/dscn.31>.

**Sessione X**  
Ipazia Alessandrina



# Il MouseGuffin come supporto di verità fittizia

Luca Bandirali  
Università del Salento – luca.bandirali [at] unisalento.it

## SINTESI

La rivoluzione digitale ha fortemente influenzato le tecniche narrative del cinema e della televisione. In questo intervento si analizza un nuovo tipo di oggetto che caratterizza i mondi narrativi contemporanei: è un oggetto digitale, intrinsecamente irrilevante ma drammaturgicamente utile, che eredita la funzione del MacGuffin di hitchcockiana memoria.

## PAROLE CHIAVE

narrazioni digitali, filosofia digitale, mondi finzionali, film studies, television studies

## INTERVENTO

### 1 INTRODUZIONE

Nella teoria di Kendall L. Walton [12], le opere di finzione sono strutturate come “giochi infantili di far finta”. Tali giochi necessitano di *supporti*, ossia di oggetti (come le bambole) che generano verità fittizie in un contesto sociale: “facciamo che i tronconi sono orsi”, è uno degli esempi di gioco proposti da Walton.

Le narrazioni dell'era digitale costruiscono mondi caratterizzati dal consueto principio di scostamento minimo dal mondo reale [10]: dunque nei film, nelle serie TV, nei videogiochi collocati nel presente si assume che la storia si svolga in un ambiente la cui estensione ontologica e la cui dotazione tecnologica coincidono essenzialmente con quella del mondo reale. In questa visione delle storie, le innovazioni tecnologiche fanno per lo più da sfondo: le automobili un film del 1950 sono diverse da quelle di un film del 2020 ma svolgono la stessa funzione in entrambi i mondi narrativi. A un grado di innovazione maggiore della tecnologia corrispondono nuove situazioni narrative. Per esempio, c'è un cinema prima e dopo il telefono cellulare: un film come *Perfetti sconosciuti*, in cui alcuni amici a cena mettono sulla tavola i propri smartphone nella speranza di non aver niente da nascondere, si può immaginare soltanto ambientato nella nostra epoca; il film si è prestato a numerosi remake ambientati in altri paesi (dalla Spagna alla Corea del Sud) ma non in altre epoche storiche.



Tra l'oggetto di sfondo (l'automobile) e l'oggetto in primo piano (il telefono), ci sono cose che hanno uno statuto più sfumato; cose che svolgono una funzione nel gioco narrativo, come i supporti di Walton e portano la storia in una certa direzione, pur essendo sostanzialmente sostituibili, come i tronconi di Walton. I contenuti delle storie intesi come "things of the world" [4] cambiano nel tempo e influenzano le forme narrative, per cui ci chiederemo anche in che modo nuove cose del mondo hanno determinato nuovi racconti.

## 2 IL MACGUFFIN E LA SUPERCAZZOLA

Nel cinema hollywoodiano classico, lo statuto di questi oggetti è stato esplicitato da Alfred Hitchcock. Prima in una conferenza alla Columbia University nel 1939, poi in una celebre intervista con il giovane collega François Truffaut del 1966, il regista inglese spiega:

Si può immaginare una conversazione tra due uomini su un treno. L'uno dice all'altro: "Che cos'è quel pacco che ha messo sul portabagagli?". L'altro: "Ah quello, è un MacGuffin". Allora il primo: "Che cos'è un MacGuffin?". L'altro: "È un marchingegno che serve per prendere i leoni sulle montagne scozzesi". Il primo: "Ma non ci sono leoni sulle montagne scozzesi". Quindi l'altro conclude: "Bene, allora non è un MacGuffin!" Come vedi, un MacGuffin non è niente.

Nel film *Notorius*, per esempio, il MacGuffin era l'uranio trafugato dai nazisti, qualcosa che serviva allo story concept come obiettivo, senza che l'uranio in sé fosse indispensabile. Il MacGuffin ha avuto una tale notorietà da essere incluso nell'*Oxford Dictionary of Phrase and Fable*, dove viene descritto come un oggetto necessario al plot e decisivo per i personaggi di un film ma irrilevante in sé. Blindato da tecnicismi, il MacGuffin rientra nel linguaggio *gibberish* o *gobbledygook*, vale a dire una forma comunicativa che suona autorevole e importante, ma di cui non si comprende il senso. A questo ordine di discorsi, il filosofo Maurizio Ferraris ha dedicato il libro *Intorno agli unicorni* [5], in cui si dibatte approfonditamente di un parente italiano del MacGuffin, la Supercazzola di *Amici miei* (1975), entrata nel 2015 nel dizionario Zingarelli con il significato di "parola o frase senza senso, pronunciata con serietà per sbalordire e confondere l'interlocutore". Per Ferraris il passaggio dall'ontologia (le cose che esistono) all'epistemologia (la conoscenza intorno alle cose) è mediato dalla tecnologia; la diffusione della tecnologia nell'era digitale consente, secondo Ferraris, di esercitare "una competenza senza comprensione" che genera sostanzialmente supercazzole o anche, citando Marcuse [9], "formule magico-rituali".

## 3 IL MOUSEGUFFIN

In questo intervento ci occupiamo di un nuovo tipo di oggetto generato dalla rivoluzione digitale che sbalordisce e confonde (come la Supercazzola), suona autorevole e importante (come il gibberish), è necessario al plot ma irrilevante in sé (come il MacGuffin). Lo chiameremo MouseGuffin. L'idea è che la tecnologia informatica offra alle narrazioni audiovisive del presente un repertorio di oggetti digitali funzionali allo scioglimento rapido di nodi narrativi.

Il Mouse-Guffin è tendenzialmente qualcosa che il narratore finge di conoscere e lo spettatore finge di comprendere. Gli oggetti digitali convocati nei film e nelle serie tv sono in tal senso numerosi: firewall, password, virus, rootkit, trojan, bug, deep web, bit coin e via dicendo. L'uso di questi oggetti digitali è spesso demandato a personaggi specializzati (l'ingegnere informatico come *official hero* e l'hacker come *outlaw hero*) ma anche dai personaggi semplicemente *geek* o *nerd*.

La specificità di questi oggetti sta nel favorire alcune svolte narrative senza troppe spiegazioni: il mandante di un omicidio digita svelto un codice alfanumerico e trasferisce bitcoin a un sicario; il sicario sta per andare a compiere l'omicidio su commissione ma la porta della sua camera d'albergo viene bloccata da un poliziotto informatico che entra nel sistema di sicurezza dell'hotel, e via dicendo.

Ci chiederemo dunque, nello scenario contemporaneo del "far finta" nel senso di Walton, come funziona questo nuovo supporto della finzione, questo MacGuffin dell'era digitale.

Esempi cinematografici e seriali degli schemi narrativi orientati dal MouseGuffin saranno tratti dai film *Domino* (Brian De Palma, 2019), *Skyfall* (Sam Mendes, 2012), *Mission: Impossible III* (J. J. Abrams, 2006), *Codice: Swordfish* (Dominic Sena, 2001) e dalle serie televisive *House of Cards* (2013-2018), *NCIS* (2003-), *Castle* (2009-2016), *Felicity* (1998-2002), *CSI: Crime Scene Investigation* (2000-2015).

## BIBLIOGRAFIA

1. Auerbach, David. «What's Up With All the "Deep Web" Stuff on House of Cards?» *Slate Magazine*, 20 febbraio 2014. <https://slate.com/technology/2014/02/house-of-cards-season-2-whats-up-with-all-the-deep-web-stuff.html>.
2. Bandirali, Luca, e Enrico Terrone. *Filosofia delle serie TV: dalla scena del crimine al trono di spade*. Mimesis, 2012.
3. ———. *Il sistema sceneggiatura*. Lindau, 2009.
4. Eldridge, Richard. «Form and Content: An Aesthetic Theory of Art» 25, n. 4 (1985): 303–16.
5. Ferraris Maurizio, Alberto. *Intorno agli unicorni*. Il Mulino, 2018.
6. Ferraris, Maurizio, e Germano Painsi. *Scienza nuova. Ontologia della trasformazione digitale*. Torino: Rosenberg & Sellier, 2018.
7. Ferraro, Guido. «Teorie della narrazione», 2015.
8. Knowles, Elizabeth. *The Oxford dictionary of phrase and fable*. Oxford: Oxford University Press, 2006.
9. Marcuse, Herbert. *L'uomo ad una dimensione*. Torino: Einaudi, 1967.
10. Pavel, Thomas G. *Mondi di invenzione: realtà e immaginario narrativo*. Einaudi, 1992.
11. Truffaut, François. *Il cinema secondo Hitchcock*. Vol. 76. Il saggiautore, 2009.
12. Walton, Kendall L. *Mimesi come far finta. Sui fondamenti delle arti rappresentazionali*. Sesto San Giovanni: Mimesis, 2011.



# Digit-filosofia o filosofia del digitale?

Enrico Terrone

Università di Genova – enrico.terrone [at] unige.it

## SINTESI

Solitamente, quando si parla di digital humanities, non si intende che le discipline umanistiche debbano includere le tecnologie digitali fra i loro oggetti di studio. In certi casi, peraltro, la pretesa sarebbe assurda, considerato che vi sono discipline umanistiche, si pensi ad esempio alla filologia o all'archeologia, che vertono su epoche che non hanno nulla a che fare con il digitale. L'idea delle digital humanities è piuttosto quella di un'integrazione delle tecnologie digitali nel metodo delle discipline umanistiche, anziché nel loro oggetto. Si tratta insomma di condurre ricerche in ambito umanistico avvalendosi sistematicamente del potenziale insito nei dati e negli algoritmi che le tecnologie digitali mettono a nostra disposizione [6].

Se però si cerca di includere la filosofia nel novero delle digital humanities il discorso si complica. Il metodo filosofico sembra infatti costitutivamente refrattario a una trasformazione che, per il tramite della nuove tecnologie, dovrebbe condurre a una nuova disciplina che si potrebbe denominare “digital philosophy” o—volendo schivare l'anglicismo—“digit-filosofia”. Fin che si tratta di considerare il digitale come oggetto di studio di una “filosofia del digitale” tutto sembra procedere per il meglio, ma quando si cerca di integrare il digitale nel metodo filosofico, passando così alla “digit-filosofia”, ci si imbatte in una difficoltà peculiare, che il mio saggio si propone di discutere.

## INTERVENTO

### 1 FILOSOFIA DI X E X-FILOSOFIA

Una peculiarità della filosofia, in rapporto alle altre discipline umanistiche, è la sua capacità di vertere su qualsiasi cosa. Prendete un X a vostra scelta, e la filosofia di X sarà un'impresa perfettamente lecita. Pare arduo, ad esempio, concepire una sociologia dell'elettrone—per non dire una filologia dell'elettrone—mentre la filosofia dell'elettrone trova agevolmente il suo posto in una linea di ricerca sui costituenti elementari della realtà che da Leucippo e Democrito arriva sino agli attuali dibattiti sulla metafisica delle particelle. Una tale versatilità differenzia la filosofia non solo dalle discipline umanistiche ma anche da quelle scientifiche: sebbene la fisica abbia ancora più titoli della filosofia per occuparsi degli elettroni, una fisica dell'arte pare, almeno allo stato attuale, un'impresa insensata, mentre la filosofia dell'arte è un ambito di ricerca perfettamente legittimo.

Non tutti però ritengono evidente che la filosofia possa occuparsi di qualsiasi cosa. Una volta un collega mi ha obiettato che proporre una filosofia dello scolapasta non sarebbe così diverso dal mettersi uno scolapasta in testa. In effetti, sono entrambi modi di avere uno scolapasta in testa, uno letterale e uno figurato. Fin qui gli si può anche dare ragione. La sostanza dell'obiezione mi sembra però basarsi su un equivoco. Il fatto che non vi siano opere filosofiche il cui tema primario è lo scolapasta non significa che i filosofi non si occupino dello scolapasta: significa piuttosto che se ne occupano indirettamente, all'interno del discorso sul genere ontologico al quale lo scolapasta appartiene, quello degli artefatti.

In effetti, la filosofia si è occupata a più riprese degli artefatti, in un arco storico che dalla teoria delle quattro cause di Aristotele arriva sino ai lavori di Peter Kroes [4] e Amie Thomasson [7]. Il fatto che non ci si occupi dello scolapasta in quanto tale ma soltanto dello scolapasta in quanto artefatto è una mera conseguenza del fatto che le caratteristiche che rendono uno scolapasta filosoficamente rilevante —il suo svolgere una funzione in base a una struttura, per dirne una—sono condivise con una varietà di altri artefatti quali ad esempio tavoli, sedie, coltelli e cucchiari. La filosofia dello scolapasta dunque c'è, basta andarla a cercare all'interno della filosofia degli artefatti.

Quel che in tutta evidenza non c'è, e mi pare bene che non ci sia, è una “scolapasta-filosofia”—o “co-lander-philosophy”, per chi predilige gli anglicismi. Questa sì che equivarrebbe a presentarsi a una conferenza con uno scolapasta in testa. Per quanto oggetto degnissimo di indagine filosofica, lo scolapasta non sembra aver nulla da dire quanto al metodo della filosofia. Non è detto però che tale conclusione sia generalizzabile. Vi sono altri oggetti di indagine che, a differenza dello scolapasta, sembrano avere la capacità di sconfinare dall'oggetto al metodo della filosofia. La scienza e l'arte risultano casi emblematici a tal proposito.

Ci si può chiedere, a questo punto, se non possa valere un discorso analogo per quanto riguarda il medium digitale. La lunga ma necessaria premessa ci ha così portato al cuore della questione. Posto che la ricerca filosofica abbia una sua specificità, in quale misura questa risulta compatibile con le novità metodologiche che dovrebbero sancire il passaggio dalla filosofia tradizionale alla digit-filosofia? In quale misura l'informatica può contribuire a rigenerare l'attività di analisi e argomentazione che è costitutiva della filosofia? Qual è la lacuna del metodo filosofico che le tecnologie digitali andrebbero finalmente a colmare? Affrontare questi interrogativi significa chiedersi se la filosofia può davvero trovare il suo posto nel quadro delle digital humanities.

## 2 FILOSOFIA E DIGITAL HUMANITIES

Al cuore del progetto delle digital humanities c'è l'idea di avvalersi della potenza di calcolo dell'informatica per meglio gestire le moli di dati che tradizionalmente costituiscono gli “input” delle discipline umanistiche. La filologia, ad esempio, può servirsi dell'informatica per meglio studiare le moli di dati che costituiscono i testi, e analogamente la sociologia può avvalersi dei computer per acquisire e analizzare le moli di dati sui comportamenti sociali. Ma la filosofia non sembra basarsi su moli di dati nel senso in cui vi si basano discipline umanistiche come la filologia o la sociologia. La ricerca filosofica non risulta dunque particolarmente penalizzata da limiti e carenze computazionali della mente umana, e pertanto non è chiaro quali benefici la filosofia potrebbe trarre dal superamento di tali limiti e dalla compensazione di tali carenze mediante l'utilizzo dei computer.

Un possibile controesempio a questo scetticismo nei confronti del potenziamento digitale del metodo filosofico viene dall'ambito storiografico. La storia della filosofia, a differenza di altri ambiti filosofici, comporta il trattamento di moli di dati costituite dalle opere dei filosofi e dagli eventuali altri documenti che ne supportano la contestualizzazione e la comprensione. La possibilità di analizzare tali dati mediante algoritmi informatici può dunque portare preziosi benefici alla ricerca storiografica, e pertanto la storia della filosofia trova sicuramente un suo posto nel quadro delle digital humanities [1]. In tal senso, la storia della filosofia funziona proprio come altre forme di storiografia quali ad esempio la storia della medicina, la storia dell'arte o la storia della letteratura [5].

Questa affinità risulta però, come minimo, sospetta. Lo storico dell'arte non è certo, in quanto tale, un artista, né lo storico della letteratura è un poeta, e difficilmente accetteremmo di farci curare da uno storico della medicina anziché da un medico. Perché mai, dunque, uno storico della filosofia dovrebbe essere considerato un filosofo? Si tratta di una questione annosa che in questa sede non può certo essere affrontata come meriterebbe. Tuttavia, la mera controversia sul considerare la storia della filosofia come forma genuina di filosofia è sufficiente per depotenziare l'argomento che fa derivare la possibilità della digit-filosofia dal fatto che gli storici della filosofia sia avvalgono con successo delle tecnologie digitali. Tale successo, infatti, potrebbe derivare proprio dal fatto che la storia della filosofia non è una forma genuina di ricerca filosofica.

Un altro possibile controesempio allo scetticismo sulla digit-filosofia viene da un programma di ricerca in forte ascesa che prende il nome di filosofia sperimentale [3]. L'antecedente fondamentale della filosofia sperimentale è il medesimo della filosofia tradizionale: Socrate che se ne va in giro per Atene a chiedere ai suoi concittadini il loro parere su nozioni quali il coraggio, la pietà o la virtù. La filosofia, tuttavia, si è storicamente costituita a partire dalla confutazione, da parte di Socrate, delle concezioni proposte dai suoi concittadini, e dalla conseguente ricerca di una concezione più convincente in quanto più razionale. La filosofia sperimentale esplora invece un'alternativa metodologica alla quale Socrate non sembrava particolarmente interessato: ordinare e contare le varie risposte degli ateniesi, alla ricerca di quelle con frequenza massima, come se la soluzione ai problemi filosofici si celasse nell'orientamento della maggioranza dei membri di una comunità.

Cruciale nel metodo della filosofia sperimentale risulta pertanto la somministrazione di questionari ad ampi gruppi di cavie. Il che significa che, a differenza di quanto accade nella filosofia tradizionale, l'acquisizione e il trattamento di moli di dati svolge un ruolo essenziale, creando così le condizioni per un fruttuoso connubio con le tecnologie informatiche. La filosofia sperimentale sembra dunque avere tutte le carte in regola per entrare a far parte delle digital humanities.

La replica dello scettico a questa obiezione è simile a quella già considerata nel caso della storiografia filosofica: è controverso che si tratti di forme genuine di filosofia. Come la storia della filosofia sembra metodologicamente più affine alla storia che non alla filosofia, così la filosofia sperimentale ha le parvenze di un'indagine sociologica su temi di interesse filosofico. Siano tutti d'accordo sul fatto che andare in giro a chiedere alle persone di risolvere equazioni non conta certo come matematica, quanto piuttosto come indagine filosofica sulle conoscenze matematiche. Ma allora perché mai interrogare le persone su questioni filosofiche dovrebbe contare come filosofia anziché come indagine sociologica sulle opinioni filosofiche?

Anche in questo caso, la discussione sullo status filosofico della filosofia sperimentale va ben oltre la portata del mio intervento. Quel che mi interessa in questa sede è soltanto evidenziare un punto molto semplice: gli ambiti filosofici che sembrano trovare posto nel quadro delle digital humanities—la storiografia filosofica da una parte, la filosofia sperimentale dall'altra—non sono forme paradigmatiche di filosofia bensì casi controversi, il che getta seri dubbi sulla possibilità di includere, in quel quadro, la filosofia nel suo complesso.

Finora mi sono soffermato su quello che ho ipotizzato essere il tratto essenziale delle digital humanities, ossia l'utilizzo dei computer per l'acquisizione e il trattamento di moli di dati, e ho sostenuto che non sembra essere un aspetto particolarmente rilevante per la filosofia. Si potrebbe però obiettare che le tecnologie digitali comportano altre novità culturali che sarebbero in grado di rivoluzionare il nostro modo di fare filosofia. Si pensi ad esempio alla straordinaria accessibilità delle risorse bibliografiche resa possibile da siti come Google Scholar o JStor, oppure alla questione dell'open access e ai possibili cambiamenti nei processi di "peer review" che si potrebbero ottenere mediante il passaggio dalle riviste tradizionali ai "repository" digitali [2]. Oppure si pensi al ruolo "neo-illuministico" di siti enciclopedici come Wikipedia o, per restare in ambito strettamente filosofico, la Stanford Encyclopedia of Philosophy. O ancora si pensi all'enorme potenziale di blog, piattaforme e social network per quanto riguarda la didattica, la divulgazione e la rilevanza pubblica della filosofia, come peraltro testimoniano la quantità e la vivacità delle discussioni filosofiche che hanno luogo in quei contesti (Aeon, Daily Nous, Leiter Reports, Aesthetics for Birds—giusto per citare qualche sito).

Nessuno di questi elementi mi sembra però capace di influenzare il metodo filosofico al punto da creare le condizioni per l'avvento della digit-filosofia. Sicuramente le tecnologie digitali hanno rivoluzionato le forme di comunicazione, come lo avevano fatto in precedenza grandi invenzioni quali la stampa, la telefonia e la radiofonia. Al cuore del metodo filosofico vi è però la produzione di idee, non la loro comunicazione, perlomeno se si ritiene che la polemica di Socrate contro i sofisti sia un momento fondamentale dell'istituzione della filosofia come pratica di ricerca.

La comunicazione rappresenta indubbiamente una fase importante della pratica filosofica, ma esorbita dal metodo filosofico, al centro del quale si trovano piuttosto la riflessione, l'analisi e l'argomentazione. La rivoluzione digitale nell'ambito della comunicazione non sembra dunque una ragione sufficiente per parlare di digit-filosofia, per le stesse ragioni per cui l'invenzione della stampa non portò a parlare di print-filosofia, né l'invenzione della radio ha portato alla radio-filosofia, sebbene ci restino memorabili momenti di filosofia radiofonica come ad esempio il dibattito fra Frederick Copleston e Bertrand Russell sull'esistenza di Dio. Le tecnologie digitali permettono alle idee filosofiche di circolare in modalità impensabili nel secolo scorso, ma la produzione di tali idee continua ad attenersi a principi metodologici che risalgono a più di duemila anni fa.

### 3 QUALE FILOSOFIA PER IL DIGITALE?

I ragionamenti fin qui proposti non escludono definitivamente la possibilità della trasformazione della filosofia in digit-filosofia, ma presi nel loro insieme sembrano suggerire un sano scetticismo al tal riguardo. La filosofia ha una sua specificità che le distingue da quelle “humanities” che transitano senza troppe difficoltà nelle “digital humanities”. Si potrebbe addirittura sospettare che la filosofia, proprio in virtù di tale specificità, si ponga al di là della tradizionale distinzione fra discipline umanistiche e scientifiche. Tale specificità, ho sostenuto, è legata al fatto che la filosofia non ha un suo dominio specifico di oggetti ma può occuparsi, potenzialmente, di qualsiasi oggetto.

La relazione fra filosofia e digitale, nella prospettiva che propongo, merita di essere considerata soprattutto nei termini di una filosofia del digitale. Senza bisogno di rivoluzionare il suo metodo, la filosofia può mettere a disposizione il suo repertorio di risorse concettuali al fine di fare luce sulle tecnologie che hanno rivoluzionato il nostro mondo. Quel che mi sembra veramente proficuo nel connubio fra filosofia e digitale, in tal senso, è la combinazione di un metodo millenario con un oggetto rivoluzionario, non l'estensione della rivoluzione dall'oggetto al metodo.

Un tema ricorrente nella cultura del Novecento è stato la fine della filosofia, o perlomeno della metafisica come filosofia prima. Da qui il tentativo di rimpiazzare la filosofia tradizionale con nuove forme di pensiero che potevano essere—come si è visto—riconducibili all'arte e alle scienze naturali oppure, in decenni successivi, a scienze umane emergenti quali l'antropologia e la semiotica. Sul finire del Novecento tuttavia, e soprattutto nel primo ventennio del nuovo secolo, la filosofia sembra essere tornata a godere di buona salute, e persino la metafisica si è rimessa in moto tornando ad attingere all'inesauribile sorgente aristotelica. Si fa sempre più diffusa la consapevolezza che la tradizionale cassetta degli attrezzi del filosofo non è niente affatto da rottamare, anzi, può risultare decisiva per comprendere come sta cambiando il mondo. Particolarmente importante, in tal senso, è la nozione di dipendenza ontologica o “grounding”, mediante la quale i filosofi possono connessioni illuminanti fra campi del sapere che restano solitamente disconnessi. L'essere, come notoriamente sosteneva Aristotele, si dice in molti modi, e dunque possiamo attribuire l'esistenza sia alle particelle elementari sia agli stati mentali sia alle norme sociali sia a una varietà di altre cose. Posto che ogni cosa ha un suo modo di esistenza, la questione davvero cruciale non è tanto “che cosa esiste?” bensì “in che modo l'essere di una certa cosa dipende da quello di altre?”.

In quest'ottica, il digitale si rivela un campo straordinariamente interessante per la ricerca filosofica, ponendoci di fronte a nuove entità e sfidandoci a trovare le relazioni di dipendenza che le costituiscono come tali. Da che cosa dipende l'esistenza di un programma informatico? E quella del web? E quella della realtà virtuale? Si tratta di oggetti inseriti in trame di dipendenza complesse, in cui svolgono un ruolo decisivo tanto i fatti fisici e matematici quanto quelli psicologici e sociali. La filosofia, con la sua visione a volo d'uccello sulla totalità dei fatti, si trova dunque in una posizione privilegiata per comprendere la natura degli oggetti che il digitale ha aggiunto al catalogo dell'universo.

La morale della favola è che per capire la rivoluzione digitale non serve rivoluzionare la filosofia: è sufficiente mettere a punto gli strumenti sviluppati nel corso dei secoli in ambiti filosofici tradizionali quali la metafisica, l'ontologia, l'epistemologia, l'etica, l'estetica, l'ermeneutica.



Quando nella filosofia dei giorni nostri si parla, ad esempio, di ontologia del cyberspazio, di metafisica del web, di etica dell'intelligenza artificiale, di estetica del virtuale, o di ermeneutica digitale, non si sta incorporando il digitale nel metodo della filosofia. Ci si serve piuttosto di quest'ultimo—in tutto il suo antico splendore—per analizzare una nuova classe di oggetti.

Quali scelte morali è giusto codificare negli algoritmi dell'intelligenza artificiale? Che tipo di artefatti sono questi algoritmi, e che tipo di artefatti sono i computer che li eseguono? I videogiochi e la realtà virtuale vanno considerati come nuove forme d'arte? Come cambia la nostra esperienza del mondo quando viene mediata dalle tecnologie digitali? Sono tutte questioni alle quali la filosofia può dare risposte illuminanti senza bisogno di rinunciare ai suoi metodi tradizionali, anzi, proprio in virtù di tali metodi.

Che ne è dunque, in ultima analisi, della relazione fra filosofia e digital humanities? La mia conclusione è che anche questa relazione non vada pensata in termini di inclusione, bensì di riflessione. Anziché entrare a far parte delle digital humanities, al cui interno non è ben chiaro quale ruolo dovrebbe svolgere, la filosofia farebbe meglio a riflettere dall'esterno sulla natura delle digital humanities medesime. Nella fattispecie, si tratterebbe di mettere a fuoco, mediante le possenti lenti concettuali dell'epistemologia, il contributo delle tecnologie digitali alla formazione del sapere. È di una filosofia *delle* digital humanities, non di una filosofia *nelle* digital humanities, che abbiamo bisogno.

## BIBLIOGRAFIA

1. Betti, A., H. van den Berg, Y. Oortwijn, e C. Treijtel. «History of Philosophy in Ones and Zeros». In *Methodological Advances in Experimental Philosophy*, a cura di M. Curtis e E. Fischer, 295–332. London: Bloomsbury, 2019. <https://doi.org/10.5040/9781350069022>.
2. Heesen, Remco, e Liam Kofi Bright. «Is Peer Review a Good Idea?» *The British Journal for the Philosophy of Science*, 2019. <https://doi.org/10.1093/bjps/axz029>.
3. Knobe, Joshua, e Shaun Nichols. «Experimental Philosophy». In *The Stanford Encyclopedia of Philosophy*, a cura di E. N. Zalta, 19 dicembre 2017. <https://plato.stanford.edu/archives/win2017/entries/experimental-philosophy/>.
4. Kroes, Peter. *Technical Artefacts: Creations of Mind and Matter*. A Philosophy of Engineering Design. Berlin: Springer, 2012.
5. Moretti, Franco. *Distant Reading*. London: Verso, 2013.
6. Terras, Melissa, e Julianne Nyhan. *Defining Digital Humanities: A Reader*. A cura di Edward Vanhuotte. Farnham: Ashgate, 2013.
7. Thomasson, Amie L. «Public Artifacts, Intentions, and Norms». In *Artefact Kinds: Ontology and the Human-Made World*, a cura di Maarten Franssen, Peter Kroes, Thomas A.C. Reydon, e Pieter E. Vermaas, 45–62. Synthese Library. Berlin: Springer, 2014. [https://doi.org/10.1007/978-3-319-00801-1\\_4](https://doi.org/10.1007/978-3-319-00801-1_4).

# Lacuna di autorità e costruzione dell'ignoranza attiva

Fabio Ciraci  
Università del Salento – fabio.ciraci [at] unisalento.it

## SINTESI

Gli elementi che minano l'ecologia dell'infosfera – disinformazione, cattiva informazione, fake-news, post-verità, deep-fake – prosperano in rete a causa di numerosi fattori, alcuni dei quali si configurano come le condizioni di contesto per il diffondersi dell'inquinamento informativo, in special modo su piattaforme digitali e reti sociali. A fianco al noto fenomeno dell'information overload (casuale o indotto), la mancanza di autorevolezza delle fonti scientifiche e istituzionali costituisce un fattore preponderante per il proliferare del ciarpame informativo. Il disaccordo guidato dall'interesse all'interno della comunità scientifica ha posto in essere uno smantellamento dell'autorevolezza della scienza agli occhi dell'opinione pubblica, decretandone inevitabilmente lo screditamento e un utilizzo capzioso e fazioso dei risultati scientifici. Il presente paper si pone l'obiettivo di analizzare il ruolo svolto dal formarsi di questa “lacuna di autorevolezza”, di analizzarne le cause, alla luce di una nuova nozione di ignoranza, che non è definita più semplicemente come “mancanza di conoscenza”, ma come produzione di un'ignoranza attiva.

## ABSTRACT

The elements that undermine the ecology of the infosphere - disinformation, misinformation, fake-news, post-truth, deep-fake - thrive on the web due to numerous factors, some of which are configured as the context conditions for the spread of information pollution, especially on digital platforms and social networks. Alongside the well-known phenomenon of information overload (casual or induced), the lack of authority of scientific and institutional sources is a preponderant factor for the proliferation of trash. The disagreement driven by interest within the scientific community has resulted in a dismantling of the authority of science for the public opinion, inevitably decreeing a discrediting of science and a specious and biased use of scientific results. This paper aims to analyze the role played by the formation of this “lack of authority”, to analyze its causes, in the light of a new notion of ignorance, which is no longer simply defined as lack of knowledge, but as the production of active ignorance.

## PAROLE CHIAVE

authority, disinformation, misinformation, ignorance, agnotology

## INTERVENTO

La natura complessa delle informazioni che circolano sulle piattaforme social è ambigua ed insidiosa, spesso capziosa. Inoltre, se la conoscenza, come afferma Floridi, racchiude sempre la verità fattuale, la non-conoscenza è invece un fenomeno più complesso e bivalente: da un lato può derivare da una conoscenza distorta o una fonte mendace, che ha un potere negativo sulle azioni umane, perché la condiziona sulla base di falsi contenuti, credenze o pseudo-verità; dall'altro lato, può essere una questione di *semplice ignoranza* (non conoscenza della verità), quindi di mancanza di verità. Una cosa, infatti, è dichiarare di non aver visto un ladro commettere un delitto (omissione colposa), un'altra è mentire affermando che il criminale è Tizio anziché Caio (falsa dichiarazione). Inoltre, ignoranza e credenza non coincidono, ma si sostengono a vicenda, si nutrono reciprocamente. Ad esempio, prima di Copernico non si sapeva che la terra fosse un pianeta in orbita attorno al sole (ignoranza); e si credeva che la terra fosse al centro dell'universo (credenza). Sebbene le due tesi si sostengano a vicenda, non coincidono. In effetti, il geocentrismo è anche compatibile con la teoria del sistema di Tycho Brahe (anch'esso errato).

Pertanto, oltre alle false informazioni diffuse involontariamente (*misinformation*) o intenzionalmente (*disinformation*), l'infosfera [10] è condizionata da vere e proprie lacune informative (ignoranza, *ignorance as uninformation*) e da ammassi di credenze o pseudo-verità, che rimangono tali fino a prova della loro inesattezza o insensatezza, oppure assurgono a rango di post-verità, inscalfibili anche se sottoposte a confutazione scientifica. A tutto ciò si aggiunge che le informazioni false o cattive non sono allo stato, per così dire, *pure*: non si tratta cioè di menzogne chiare e distinte. Ovvero, non si tratta di risolvere un'espressione formale secondo la logica booleana, poiché le singole proposizioni non sono enucleabili e atomiche, dissolte e risolte, ma esistono insieme ad altro, come grumi concettuali, in cui vero e falso convivono, in un sistema granulare tipico dell'informazione in rete (Roncaglia), nella forma della plausibilità, del *verosimile*. Infatti, se le informazioni fossero palesemente vere o false, nessun utente cadrebbe nella trappola delle fake-news.

In tale contesto, un ruolo significativo è svolto dalla comunità scientifica e dalla credibilità che le si riconosce, come fonte autorevole di verità. La divergenza tra diversi punti di vista, che nella storia è il sale della scienza, può però essere manipolata a favore di false costruzioni di realtà, dovute a verità convenienti, condizionate cioè da interessi economici o scientifici o più semplicemente di parte. Con ciò si genera un fenomeno di discredito che può essere definito come "lacuna di autorità". I punti dogmatici di divergenza o disaccordo tra presunti scienziati (o individui che, in virtù di una più o meno flebile fama mediatica si accreditano presso l'opinione comune come luminari) e la comunità scientifica internazionale generano un fenomeno di demolizione dell'autorità in senso ampio (istituzionale, sociale e scientifico). Di ignoranza non solo come vuoto informativo ma soprattutto come "costruzione sociale", quindi in senso attivo, si discute in Kourany e Carrier [13]. I curatori usano il termine *agnologia* (termine coniato invero da Robert Proctor) per riferirsi allo studio di ignoranza intenzionalmente prodotta, "Creata, mantenuta e manipolata" da una scienza sempre più condizionata dalla politica e dagli affari, come costruzione sociale giustapposta.

L'agnotologia riguarda la costruzione attiva dell'ignoranza attraverso la progettazione e l'interpretazione distorte di esperimenti e studi empirici, ad esempio la "falsa pubblicità" dei negazionisti del cambiamento climatico; la costruzione "virtuosa" dell'ignoranza, ad esempio limitando la ricerca sulle differenze cognitive legate alla razza e al genere; oppure, l'ignoranza come sottoprodotto non intenzionale delle scelte fatte nel processo di ricerca, quando regole, incentivi e metodi incoraggiano un'enfasi sugli effetti benefici (o nocivi) e commerciali dei prodotti chimici industriali e quando certi concetti e persino gli interessi di certi gruppi non possono essere sfruttati in un data struttura concettuale. Secondo questo nuovo approccio, insomma, l'ignoranza è molto più complessa di quanto si sia sinora pensato. L'ignoranza non è cioè solo il vuoto che precede la conoscenza o la privazione che risulta da un'attenzione parziale. È anche – anzi e soprattutto - qualcosa di socialmente costruito e attivo: la confusione prodotta, ad esempio, quando interessi particolari bloccano l'accesso alle informazioni o addirittura creano disinformazione (p. 3). Ferrier [7] aveva già fatto riferimento al tema di una forma di ignoranza attiva. L'autore non solo usa per primo il termine *epistemologia* in senso moderno, ma usa per primo il termine *Agnoiologia* (il libro è diviso in tre sezioni: dopo una lunga Introduzione, seguono le sezioni su Epistemologia, Agnoiologia e Ontologia). In un certo senso, l'agnologia di Proctor sembra essere influenzata dalla agnoiologia di Ferrier, che scrive "L'ignoranza è un difetto intellettuale, imperfezione, privazione o mancanza" [7]. Tuttavia, pur avendo una natura difettosa, costituisce una barriera contro la conoscenza, che svolge un ruolo attivo. Su questo argomento Arfini [3].

Un altro fenomeno rilevante, a tal proposito, è generato da una forma di ignoranza attiva derivata dalla *proiezione di autorità*, che ha luogo quando uno scienziato viene chiamato a discutere di questioni che esulano dalle proprie specifiche competenze scientifiche e dall'ambito dei propri studi. In tal caso, l'autorevolezza conquistata nel proprio ambito viene proiettata sulle opinioni dello scienziato che è legittimato a disquisire di temi e argomenti non pertinenti al suo campo di indagine, senza cioè adottare alcun metodo scientifico. Nell'ambito delle piattaforme sociali questo fenomeno moltiplica il suo potere in senso virale, polarizzando la discussione, generando cioè un'oggettivazione di opinioni frequenti e un ricorso all'*argumentum ab auctoritate* che abbassa il livello critico della discussione.

In tal senso, la cornice di delegittimazione scientifica in atto, tanto sulle piattaforme sociali quanto nel paese reale, si riflette anche sui comportamenti degli stessi scienziati, con conseguenze catastrofiche per il credito e l'autorevolezza della comunità scientifica: Venendo meno a un *aplomb* istituzionale e professionale improntato al rigore scientifico, lo scienziato ricerca il plauso e il consenso si piega a una polarizzazione delle opinioni e, per dirlo con Freud, ha luogo una "alleanza del sintomo", sicché egli si allinea a una fazione polarizzata dell'opinione comune. In parte, la proiezione dell'autorità è effetto del condizionamento performante dei new-media che tendono a spettacolarizzare il messaggio, con effetti di semplificazione e polarizzazione delle opinioni.

## Ringraziamenti

Il presente contributo sviluppa un'indagine condotta in occasione del saggio Per una teoria critica del digitale: fake-news e postverità alla luce della logica della verosimiglianza, per il numero monografico Filosofia e digitale, "Rivista di Filosofia", ISSN0015-1823, Mimesis, 2020, curato assieme a Cristina Marras e Riccardo Fedriga. Si tratta quindi di ricerche che ho discusso, in tempi e luoghi diversi, non solo con i curatori del monografico, ma anche con altri amici e studiosi (alcuni dei quali hanno contribuito come autori): Luca Bandirali, Enrico Terrone, Mario Bochicchio, Roberto Limonta, Luigi Catalani, Federico Meschini, con i quali da lungo tempo, ormai, esiste un ricco e proficuo confronto sui temi relativi al ruolo svolto dalla filosofia nella teoria dell'informazione e, più in generale, della cultura digitale. A tutti loro va il mio grato riconoscimento.

## BIBLIOGRAFIA

1. Andina, Tiziana. «Truth, lies, and post-truth». In *Post-Truth, Philosophy and Law*, a cura di Angela Condello e Tiziana Andina, 1–13. London: Routledge, 2019.
2. Appel, Markus. *Die Psychologie des Postfaktischen: Über Fake News, „Lügenpresse“, Clickbait & Co.* Berlin-Heidelberg: Springer, 2020.
3. Arfini, Selene, Selene Arfini, e Di Cecco. *Ignorant cognition*. Berlin: Springer International Publishing, 2019.
4. Cameron, Glen T., e Kuen-Hee Ju-Pak. «Information pollution?: Labeling and format of advertorials». *Newspaper Research Journal* 21, n. 1 (2000): 65–76.
5. Del Vicario, Michela, Gianna Vivaldo, Alessandro Bessi, Fabiana Zollo, Antonio Scala, Guido Caldarelli, e Walter Quattrociocchi. «Echo chambers: Emotional contagion and group polarization on facebook». *Scientific reports* 6 (2016): 37825. <https://doi.org/https://doi.org/10.1038/srep37825>.
6. Ferraris, Maurizio. *Postverità e altri enigmi*. Bologna: Il Mulino, 2017.
7. Ferrier, James Frederick. *Institutes of Metaphysic: The Theory of Knowing the Mind*. London: Blackwood and Sons, 1854.
8. Floridi, Luciano. «Artificial intelligence, deepfakes and a future of ectypes». *Philosophy & Technology* 31, n. 3 (2018): 317–321.
9. ———. *Il verde e il blu: Idee ingenue per migliorare la politica*. Milano: Raffaello Cortina, 2020.
10. ———. *La rivoluzione dell'informazione*. Torino: Codice, 2012.
11. ———. *Pensare l'infosfera: La filosofia come design concettuale*. Milano: Raffaello Cortina, 2020.
12. Hanin, Mark L. «Theorizing Digital Distraction». *Philosophy & Technology*, n. I–XX (2020): 1–12. <https://doi.org/10.1007/s13347-020-00394-8>.
13. Kourany, Janet, e Martin Carrier. *Science and the Production of Ignorance: When the Quest for Knowledge Is Thwarted*. Cambridge: MIT Press, 2020.

14. Lazer, David MJ, Matthew A. Baum, Yochai Benkler, Adam J. Berinsky, Kelly M. Greenhill, Filippo Menczer, Miriam J. Metzger, Brendan Nyhan, Gordon Pennycook, e David Rothschild. «The science of fake news». *Science* 359, n. 6380 (2018): 1094–1096. <https://doi.org/https://doi.org/10.1126/science.aao2998>.
15. Magnani, Lorenzo. *Abductive Structure of Scientific Creativity*. Berlin: Springer, 2017.
16. Pala, Davide. «A political and deliberative virtue? The epistemic trust in trustworthy epistemic authorities». In *Post-Truth, Philosophy and Law*, di Angela Condello e Tiziana Andina, 112–35. London: Routledge, 2019.
17. Quattrococchi, Walter, e Antonella Vicini. *Misinformation. Guida alla società dell'informazione e della credulità*. Milano: Franco Angeli, 2016.
18. Roncaglia, Gino. «Tra granularità e complessità: contenuti digitali e storia della rete». *Nuovi annali della Scuola speciale per archivisti e bibliotecari* 31 (2017): 349–361.
19. Vosoughi, Soroush, Deb Roy, e Aral Sinan. «The spread of true and false news online». *Science* 359, n. 6380 (2018): 1146–1151. <https://doi.org/https://doi.org/10.1126/science.aap9559>.
20. Zuboff, Shoshana, e Paolo Bassotti. *Il capitalismo della sorveglianza: il futuro dell'umanità nell'era dei nuovi poteri*. Roma: Luiss University Press, 2019.



# Filosofia e digitale: le scienze filosofiche in Wikipedia

Luigi Catalani  
Università della Basilicata – luigi.catalani [at] unibas.it

## PAROLE CHIAVE

Wikipedia, filosofia, enciclopedia, Wikimedia

## INTERVENTO

Il contributo approfondisce il rapporto tra le scienze filosofiche e l'enciclopedia libera, che il 15 gennaio 2021 festeggia i suoi primi venti anni. L'intervento analizza, in particolare, il Progetto Filosofia della versione in lingua italiana di Wikipedia – che è il luogo di coordinamento per gli utenti che vogliono dedicarsi alla produzione e al miglioramento delle voci concernenti questa disciplina – ma indaga anche il Portale Filosofia e i portali collegati, il sistema di monitoraggio e l'opera di manutenzione delle voci, il luogo di discussione degli utenti che si occupano di filosofia, la presenza di contenuti filosofici nelle altre piattaforme dell'ecosistema wikimediano. Il contributo sottopone al vaglio gli standard per la stesura delle biografie dei filosofi, le norme per la citazione delle fonti esterne, l'efficacia delle operazioni di categorizzazione degli autori e delle opere e i criteri per la compilazione delle liste di voci cosiddette “indispensabili”. L'intervento riflette sull'opportunità di tradurre contenuti enciclopedici di interesse filosofico da altre versioni linguistiche e pone a confronto il grado di copertura di autori e concetti filosofici nelle principali versioni linguistiche di Wikipedia. L'analisi del Portale Filosofia, il principale punto di accesso ai contenuti enciclopedici di natura filosofica, divisi per categorie (filosofia, filosofi, opere filosofiche), è lo spunto per riflettere sulla capacità dell'enciclopedia di orientare la consultazione verso le discipline classiche, le discipline tematiche, i metodi e gli strumenti dell'indagine filosofica, le principali posizioni e teorie, le scuole e le correnti filosofiche. Il contributo dedica particolare spazio ad una delle attività peculiari di Wikipedia, ossia il monitoraggio, in virtù del quale gli utenti hanno la possibilità di valutare le voci secondo diversi indici. Poiché le valutazioni proposte favoriscono il processo di miglioramento qualitativo delle pagine, consentendo di individuare aspetti critici ed effettuare interventi mirati, nell'intervento vengono analizzati i parametri in base ai quali è valutata la qualità delle voci e viene descritto lo stato dell'arte dell'azione di monitoraggio effettuata dagli utenti attivi nel Progetto Filosofia. Vengono inoltre passate in rassegna le voci filosofiche in vetrina o di qualità, al fine di individuarne i tratti comuni e il rispetto degli standard qualitativi.



Il contributo si sofferma poi sullo spazio di discussione degli utenti che si occupano di filosofia, denominato “La botte di Diogene”: è il luogo in cui si possono porre domande, inserire comunicazioni e coordinare il lavoro intorno alle voci di filosofia presenti in Wikipedia. Si individuano i tratti principali delle discussioni e il numero di utenti coinvolti. Infine, l’intervento offre una panoramica delle risorse filosofiche presenti negli altri progetti della galassia Wikimedia: le opere contenute in Wikisource, le citazioni presenti in Wikiquote, i manuali disponibili in Wikibooks, i lemmi presenti in Wikizionario, le dispense raccolte in Wikiversità, le immagini disponibili su Commons e gli elementi presenti in Wikidata. Sulla base dei dati raccolti, si evidenziano punti di forza, criticità e margini di miglioramento dell’enciclopedia libera in riferimento alla diffusione del pensiero filosofico.

## BIBLIOGRAFIA

1. Bisogno, Armando. «Wikipedia loves science. Perché Wikipedia ha bisogno dell’Università (e viceversa) in Wikipedia in Academia». In *Atti del convegno internazionale*, a cura di Corrado Petrucco, in corso di pubblicazione.
2. Bynum, Terrell Ward, e James Moor. «La fenice digitale: come i computer stanno cambiando la filosofia». *Apogeo*, 2000.
3. Catalani, Luigi. «Biblioteche e Wikimedia: strategie comuni per l’accesso aperto alla conoscenza e la costruzione collaborativa del sapere libero». *JLIS. it: Italian Journal of Library, Archives and Information Science. Rivista italiana di biblioteconomia, archivistica e scienza dell’informazione* 8, n. 3 (2017): 100–114.
4. ———. *Come educare al sapere libero nella biblioteca scolastica*. Editrice Bibliografica, 2020.
5. ———. «Costruire un manuale universitario con gli studenti su Wikibooks in Wikipedia in Academia». In *Atti del convegno internazionale*, a cura di Corrado Petrucco, 2019.
6. ———. «I progetti Wikimedia per l’apprendimento delle competenze informative e digitali in biblioteca, a scuola, nelle università». *AIB studi* 57, n. 2 (2017).
7. ———. «La biblioteca come ambiente di produzione di risorse educative aperte». *Biblioteche oggi Trends* 6, n. 1 (2020): 20–28.
8. ———. «“Omnia Numerorum Videntur Ratione Formata”. A ‘Computable World’ Theory in Early Medieval Philosophy». In *International Conference on the History and Philosophy of Computing*, 131–140. Springer, 2015.
9. Dalby, Andrew. *The world and Wikipedia: How we are editing reality*. Siduri, 2009.
10. Davis, Martin. *The universal computer: The road from Leibniz to Turing*. CRC Press, 2018.
11. Di Giandomenico, Mauro. «L’uomo e la macchina trent’anni dopo. Filosofia e informatica ieri ed oggi.» In *Atti del Convegno nazionale della Società Filosofica Italiana*. Laterza, 2000.
12. Floridi, Luciano, a c. di. *Infosfera. Etica e filosofia nell’età dell’informazione*. Torino: Giappichelli, 2009.
13. ———. *The Blackwell Guide to Philosophy of Computing and Information*. Blackwell, 2004.
14. Fondazione Centro studi filosofici. *Enciclopedia filosofica*. Milano: Bompiani, 2006.
15. Lilh, Andrew. *La rivoluzione di Wikipedia: come un gruppo di illustri sconosciuti ha creato la più grande enciclopedia del mondo*. Codice, 2010.

16. Mancini, Marco. *Introduzione a: Giorgio R. Cardona, Introduzione all'etnolinguistica*. UTET, 2006.
17. Martinelli, Luca. «Wikidata: la soluzione wikimediana ai linked open data». *AIB Studi* 1, 2016, 75–85. <https://aibstudi.aib.it/article/view/11434>.
18. Orlowitz, Jake. «The Wikipedia Library: la più grande enciclopedia ha bisogno di una biblioteca digitale e noi la stiamo costruendo». *JLIS. it: Italian Journal of Library, Archives and Information Science. Rivista italiana di biblioteconomia, archivistica e scienza dell'informazione* 9, n. 3 (2018): 1–15. <https://www.jlis.it/article/view/12505/11342>.
19. Pagallo, Ugo. *Introduzione alla filosofia digitale. Da Leibniz a Chaitin*. Giappichelli, 2005.
20. Vrandečić, Denny, e Markus Krötzsch. «Wikidata: a free collaborative knowledgebase». *Communications of the ACM* 57, n. 10 (2014): 78–85. <https://doi.org/https://dl.acm.org/doi/10.1145/2629489>.

# Filosofia e Digitale: un confronto dialettico e multiprospettico tra teorie e pratiche

Cristina Marras  
ILIESI-CNR – cristina.marras [at] cnr.it

## SINTESI

L'utilizzo di strumenti informatici nel campo delle ricerche storiche, filosofiche e filologiche ha negli anni disegnato un panorama problematico e complesso. La filosofia, come le altre discipline umanistiche, ha attraversato e condiviso l'evoluzione delle tecnologie per la ricerca, anche in filosofia gli sforzi si sono concentrati nella creazione di sempre più completi archivi e biblioteche digitali di fonti testuali, sui problemi di edizione, sulla codifica di testi con linguaggi standard e su motori di ricerca sempre più raffinati per l'accesso, il recupero, l'archiviazione e il riuso delle informazioni e dei documenti. Ciononostante, in ambito filosofico gli studiosi sembrano guardare alla tecnologia per lo più come supporto per la ricerca, ed è difficile identificare una comunità di “filosofi digitali” aggregata (§1). Questo contributo prende le mosse da una ricognizione della presenza della filosofia nelle riviste di umanistica digitale e si propone una riflessione sull'interdisciplinarietà delle metodologie attraverso l'analisi di alcuni progetti che hanno tracciato dei percorsi e dei modelli di ricerca di ‘filosofia digitale’ (§2). Esperienze che mettono in luce come il ruolo della filosofia sia anche quello di scardinare le dicotomie (materiale/digitale, fatti/valori, descrittivo/normativo) (§3) ostacolo ad una concreta sinergia e mutua trasformazione tra ricerca umanistica e pratiche computazionali, come si sottolineerà nelle conclusioni (§4).

## PAROLE CHIAVE

filosofia digitale, modelli, dicotomie, multiprospettivismo, dialettica

# INTERVENTO

## 1 INTRODUZIONE

L'utilizzo di strumenti informatici nel campo delle ricerche storiche, filosofiche e filologiche ha disegnato negli anni un panorama problematico e complesso. La filosofia, come le altre discipline umanistiche, ha attraversato e condiviso l'evoluzione delle tecnologie per la ricerca; anche in filosofia, gli sforzi si sono concentrati nella creazione di archivi e biblioteche digitali di fonti testuali, sui problemi di edizione, sulla codifica dei documenti con linguaggi standard e su motori di ricerca sempre più raffinati per l'accesso, il recupero, l'archiviazione e il riuso delle informazioni e dei documenti [1; 2].

Ciononostante, la tecnologia è prevalentemente "relegata" a facilitare l'accesso agli archivi e ai repository e alla pubblicazione e diffusione dei risultati della ricerca, inoltre, è difficile identificare una comunità di "filosofi digitali" aggregata. Questo è anche confermato dalla tipologia di progetti e da una ricognizione della presenza della filosofia nelle riviste di umanistica digitale. Questo contributo prende le mosse da questa ricognizione e intende discutere l'interdisciplinarietà delle metodologie nella relazione tra filosofia e scienze computazionali analizzando alcuni progetti che hanno elaborato specifici modelli di ricerca digitale in filosofia (§ 2). Attualmente, l'ambito della ricerca filosofica digitale sembra dividersi in due settori: uno relativo ai progetti 'digitali' di filosofia e uno che discute teoricamente categorie, paradigmi, modelli. Meno frequentemente i secondi accompagnano in modo integrale e integrato i primi. In un recente volume collettaneo che indaga proprio i rapporti e i limiti nella mutua trasformazione tra ricerca filosofica e pratiche computazionali [3] la questione che si pone, e che vien ripresa in questo contributo, è come concettualizzare e tematizzare l'euristica contenuta nella relazione tra teoria filosofica e pratiche computazionali a partire da concrete esperienze di ricerca. Dalle attuali discussioni sembra configurarsi un compito per la filosofia, quello di un lavoro per un responsabile confronto costruttivo e dialettico al fine di superare le dicotomie che si sono generate dalla radicale contrapposizione tra ciò che viene considerato 'naturale' e ciò che è 'artificiale' (§3). Il contributo si conclude sottolineando l'importanza di assumere un approccio multiprospettico nella ricerca filosofica digitale per la ridefinizione e l'innovazione dei paradigmi e delle metodologie di ricerca in sinergia con le sue pratiche [7; 8; 9; 14] (§4).

## 2 FILOSOFIA E UMANISTICA DIGITALE: TEORIE E PRATICHE

Nelle riviste di umanistica digitale la presenza della filosofia, sia come riflessione teorica sul e con il digitale, sia come presentazione di progetti di ricerca o di casi studio che lo assumono a oggetto di ricerca, è limitata. Si è condotto un primo spoglio di occorrenze del termine ‘filosofia’, sulla lista delle riviste di Digital Humanities (DH) realizzata da Gianmarco Spinaci, Giovanni Colavizza, Silvio Peroni<sup>1</sup>. In particolare, al momento, ci si è focalizzati su *Digital Humanities Quarterly*<sup>2</sup>, *Digital Scholarship in the Humanities*<sup>3</sup>, *Digital Studies / Le camp numerique*<sup>4</sup>, *Umanistica Digitale*<sup>5</sup>, e su una lista selezionata delle riviste di DH<sup>6</sup>, oltre che sulla lista di pubblicazioni selezionata da Alliance of Digital Humanities Organizations (ADHO)<sup>7</sup>. Il numero di occorrenze conferma quanto detto. Solo per fare un esempio estratto da questa ricognizione (il dettaglio completo dei risultati della ricerca è in corso di elaborazione per la pubblicazione), la rivista *Umanistica Digitale* (8 numeri all’attivo dal 2017) che offre un panorama abbastanza accurato del dibattito e dei temi di ricerca presenti in Italia nell’ambito degli studi umanistici digitali, su circa 60 articoli pubblicati, solo 3 toccano direttamente temi filosofici (dati del luglio 2020). Anche tra i progetti di ricerca non sono molti quelli digitali di filosofia capaci di bilanciare e integrare ricerca teorica e storico-filosofica con lo sviluppo di strumenti computazionali. Colpisce come, a dispetto della portata innovativa di alcuni progetti, le dinamiche di integrazione tra i due campi richiedano ancora ulteriori riflessioni, studi dedicati e pratiche integrate, nonostante questa necessità sia ormai rilevata da tempo [6, 10]. Nella presentazione prenderò pertanto brevemente in considerazione alcuni progetti che, a partire da diverse finalità, contribuiscono a riflettere e proporre modelli di ricerca filosofica collaborativa e sinergica con teorie e pratiche computazionali. In particolare, presento il modello di lavoro e alcune primi risultati del progetto in corso *L’iter Italicum di G. W. Leibniz: coordinate per una mappa digitale*. Il progetto parte dal volume dello studioso André Robinet, *G. W. Leibniz, Iter italicum (Mars 1689-Mars 1690). La dynamique de la République des Lettres* (Olschki 1988) per ricostruire il viaggio italiano del filosofo tedesco. Il testo è uno scrigno di informazioni intrecciate (contatti, luoghi, riferimenti bibliografici, corrispondenze, temi di discussione etc...) organizzate in ordine cronologico seguendo le tappe del viaggio, ma tale ricchezza rende anche difficile la fruizione dei dati, capire la rete intellettuale e scientifica dei contatti particolarmente intricata per un filosofo come Leibniz, avere una visione di insieme della complessità dei temi e dei contenuti correlati.

---

1 <https://zenodo.org/record/3406564>

2 <http://www.digitalhumanities.org/dhq>

3 <https://academic.oup.com/dsh>

4 <https://www.digitalstudies.org/>

5 <https://umanisticadigitale.unibo.it/>

6 <https://digitalhumanities.berkeley.edu/resources/digital-humanities-journals>

7 <https://adho.org/publications>

In questa prima fase di lavoro anziché digitalizzare l'intero volume, si è scelto di utilizzare gli indici (persone, concetti, luoghi, accademie scientifiche, testi) come chiavi di accesso ai contenuti e, a partire dalla mappa del viaggio si è elaborata una serie di mappe concettuali al fine di restituire la multilinearità e la stratificazione delle informazioni. Ciò consente anche di avere un contesto di accesso immediato per la verifica, l'aggiornamento e l'integrazione delle informazioni via via che il lavoro di ricerca procede e si sviluppa. La ricerca storico filosofica è condotta a partire da un modello di lavoro (ecosistema) che concettualizza i processi di interscambio tra filosofia e pratiche computazionali (euristica), mette insieme le strategie di accesso ai dati, le modalità di rappresentazione della 'conoscenza', con particolare attenzione alla modellizzazione dei dati stessi, testa gli strumenti informatici per la ricerca a uso della comunità (open, standard etc...), ottimizza le risorse, ha come finalità quella di rendere i dati F.A.I.R. L'obiettivo, oltre a consentire agli studiosi l'accesso a delle informazioni talvolta poco fruibili e aprire possibili nuove interpretazioni, è quello di tematizzare l'euristica contenuta nella relazione tra filosofia e pratiche computazionali nella ricerca e non continuare a definire e ribadire perimetri disciplinari e categoriali [11; 12].

### **3 UN ECOSISTEMA PLURILINGUE E DIALETTICO**

L'esperienza di ricerca in progetti filosofici digitali mostra come filosofia e pratiche computazionali, teoria e pratica sono spesso considerati come polarità irriducibili: riflettere sia sul valore dell'interdisciplinarietà e sulla cross-fertilizzazione dei confini disciplinari e sugli orizzonti che si aprono, rappresenta dunque un test critico per la comunità di "filosofi digitali". La rivoluzione digitale ha messo in luce la necessità di riorganizzare anche sul piano strutturale e infrastrutturale la formazione e la ricerca, le modalità di condurla, diffonderla e condividerla. Occorre notare anche che, guardando agli ultimi tristi avvenimenti legati alla pandemia, anziché considerare la 'rivoluzione digitale' una opportunità, si cerca di appiattire i nuovi modelli e strumenti su consueti approcci e paradigmi, perpetrando opposizioni irriducibili o manichee non certo nuove nella storia del pensiero.

Il dibattito si è fatto sempre più animato e 'la questione digitale' è entrata nelle agende pubbliche, politiche, accademiche, con discussioni che si costruiscono su una serie di dicotomie, familiari al dibattito filosofico, tra quelle più frequenti si possono menzionare: Oggettivo/Soggettivo, Giustificazione/Scoperta, Assoluto/Relativo, Fatti/Valori, Descrittivo/Normativo, Meccanico/Creativo, Conoscenza/Credenza, Tecnocrazia/Democrazia, Materiale/Digitale, Naturale/Artificiale, Scienza Pura/Scienza Applicata. Anche solo alcune di queste polarità mostrano come utilizziamo, descriviamo e analizziamo quanto accade con un uso collettivo di modelli cognitivi condivisi basati su una schematizzazione dicotomica di conoscenza, valori ed emozioni. Il dibattito sulle questioni di preoccupazione pubblica o scientifica non viene affrontato dialetticamente ma sempre come scelte alternative.

Proprio su queste dicotomizzazioni bisognerebbe lavorare filosoficamente per depotenziare il carattere esclusivo delle alternative che suppongono di esaurire lo spazio degli atteggiamenti possibili di fronte a un determinato insieme di fenomeni. Almeno in linea di principio tra i poli vi è lo spazio per una posizione intermedia, uno spazio concettuale che si apre tra le polarità. Attuare un confronto dialettico 'dal di dentro' tra filosofia e digitale significa abbandonare le radicalizzazioni, rendere dinamiche le posizioni, le obiezioni e gli argomenti.

Considerare il dibattito filosofico sul digitale come un elemento costitutivo per l'evoluzione delle idee significa aprire la strada all'innovazione; non un dibattito per dimostrare di avere ragione riducendosi a una dimensione eristica, ma un dibattito che assuma i caratteri della controversia, una intensa e specifica attività polemica in cui l'obiettivo non è vincere sull'avversario a partire da irriducibili posizioni contrapposte, ma trovare una soluzione a un problema attraverso il coordinamento di punti vista differenti (filosofia multiprospettica) [4; 5].

## **4 CONCLUSIONI: FILOSOFIA DIGITALE MULTIPROSPETTICA**

La filosofia nel e per il digitale può contribuire a rendere permeabili i confini disciplinari, definire nuovi metodi per la ricerca e mettere a fuoco alcune domande e questioni fondamentali transdisciplinari: modellizzazione della conoscenza, riflessione sul linguaggio e la terminologia, formalizzazione qualitativa e quantitativa dei processi, rappresentazione della conoscenza, ridiscussione delle premesse e dei paradigmi di ricerca, temi etici, infrastrutture. Una breve e sommaria rassegna dei progetti vuole mettere in evidenza come originalità e innovazione non siano sufficienti per garantirne la reale sostenibilità e sviluppo. Indubbiamente il supporto economico è un fattore fondamentale, la ricerca e la tecnologia hanno infatti un costo, ma altrettanto fondamentale è che ci sia un concreto coinvolgimento delle istituzioni che li promuovono e delle comunità scientifiche che li animano, li popolano di contenuti, sviluppano sinergicamente le tecnologie e, in particolare che si investa nella formazione [13]. Solo quando la comunità scientifica sentirà la responsabilità della cura delle risorse online, non le darà per date e solo come funzionali alla ricerca, ma le vedrà, come una opportunità di collaborazione e di creazione di conoscenza condivisa, solo allora potremmo immaginare di entrare anche in filosofia in una dimensione della ricerca aperta e innovativa, con la creazione di nuovi modelli ermeneutici, e con un rinnovamento delle forme discorsive, narrative e di comunicazione [15].

## BIBLIOGRAFIA

1. Bandirali, Luca, Fabio Ciracì, Luigi Catalani, Francesca Di Donato, Riccardo Fedriga, e Cristina Marras. «Filosofia e Digitale?/Digital Philosophy?» *AIUCD 2019*, 2019, 21.
2. Buzzetti, Dino. «Che cos' è, oggi, l'informatica umanistica? L'impatto della tecnologia». *Dall'Informatica umanistica alle culture digitali: Atti del convegno di studi (Roma, 27-28 Ottobre 2011) in memoria di Giuseppe Gigliozzi 2* (2012).
3. Buzzetti, Dino, Claudio Gnoli, e Paola Moscati. «The Digital Humanities Role and the Liberal Disciplines». In *Proceedings of the Third AIUCD Annual Conference on Humanities and Their Methods in the Digital Ecosystem*, 1–8, 2014.
4. Dascal, Marcelo, Amnon Knoll, e Daniel Cohen. «'Cognitive systemic dichotomization' in public argumentation and controversies», 2011.
5. Davisson, Amber, e Paul Booth. *Controversies in digital ethics*. Bloomsbury Publishing USA, 2016.
6. Di Iorio, Paolo. *HyperNietzsche. Modèle d'un hypertexte savant sur Internet pour la recherche en sciences humaines. Questions philosophiques, problèmes juridiques, outils informatiques.*, 2000. <http://www.diorio.info/puf/index.html>.
7. Dyson, George. *La cattedrale di turing. Le origini dell'universo digitale*. Codice, 2012.
8. Floridi, Luciano. *Pensare l'infosfera: La filosofia come design concettuale*. Raffaello Cortina Editore, 2020.
9. ———. *The fourth revolution: How the infosphere is reshaping human reality*. OUP Oxford, 2014.
10. Hrachovec, Herbert, e Alois Pichler. *Philosophy of the Information Society: Proceedings of the 30th International Ludwig Wittgenstein-Symposium in Kirchberg, 2007*. Vol. 7. Walter de Gruyter, 2013.
11. Liburdi, Annarita. «Materiali per una storia dell'ILIESI», 2018.
12. Marras, Cristina. «Exploring Digital Environments for Research in Philosophy: Metaphorical Models of Sustainability». In *Proceedings of the Third AIUCD Annual Conference on Humanities and Their Methods in the Digital Ecosystem*, 1–4, 2014.
13. ———. «'Filosofia, ricerca e innovazione. Considerazioni a margine del documento 'Orientamenti per l'apprendimento della filosofia nella società della conoscenza'»», 2018.
14. Moor, James, e Terrell Ward Bynum. *The Digital Phoenix how computers are changing philosophy*. Oxford: Oxford University Press, 1998.
15. Roncaglia, Gino. *La quarta rivoluzione: sei lezioni sul futuro del libro*. Gius. Laterza & Figli Spa, 2011.





# **Sessione XI**

## Nelson Mandela



# Per una democrazia partecipata: gli storici, wiki e le citizen humanities

Deborah Paci

Università di Venezia Ca' Foscari | Università di Modena e Reggio Emilia  
deborah.paci [at] unive.it – deborah.paci [at] unimore.it

## PAROLE CHIAVE

digital History, wiki, citizen humanities

## INTERVENTO

L'espressione Citizen Science (CS), "scienza dei cittadini" o "scienza partecipata", indica la partecipazione e il coinvolgimento attivo dei cittadini in attività di ricerca scientifica. Secondo il White Paper on Citizen Science for Europe la CS esprime il contributo effettivo offerto alla scienza da soggetti non necessariamente appartenenti a istituzioni scientifiche o culturali [5]. Una comunità di cittadini di diversa estrazione sociale e politica partecipa in modalità volontaria al reperimento e all'analisi dei dati, contribuendo così a implementare gli orizzonti della ricerca scientifica. La CS è contraddistinta dalla sua vocazione alla condivisione e all'inclusione: ciò la rende una modalità innovativa di svolgere attività di ricerca e al contempo rende effettiva la democratizzazione della conoscenza. Questa modalità di scienza partecipata ha diversi ambiti di applicazione, il cui livello di sviluppo dipende dal contesto culturale e sociale in cui vengono formulate le ipotesi di ricerca. Per quanto la CS sia basata sul volontariato, è condizione necessaria l'istituzione di un sistema di relazioni tra quanti decidono di partecipare alla realizzazione dei progetti di CS. Rispetto ai fondi – necessariamente limitati – che le istituzioni culturali possono stanziare, il contributo alla ricerca dei volontari, proprio per l'ampiezza delle risorse umane mobilitate, può risultare di gran lunga superiore a quello che viene fornito dai professionisti del settore nei luoghi tradizionalmente deputati alla ricerca. Perché un progetto di CS possa essere coronato dal successo è necessario che i soggetti coinvolti siano in grado di adottare soluzioni innovative, ma al contempo rispettose del rigore scientifico nella progettazione delle varie fasi della ricerca. Mentre la CS include le scienze naturali, come la biologia, la chimica e la fisica e la citizen social science si occupa delle società, le citizen humanities si applicano alle discipline storiche, letterarie, linguistiche e filosofiche [1].

La Digital Public History, per la sua vocazione pubblica, partecipativa e collaborativa in un contesto digitale, si presenta come uno dei campi di applicazione privilegiata dalle citizen humanities. Il valore aggiunto delle citizen humanities dipende primariamente da due condizioni alla base della Digital Public History: la partecipazione degli storici che non provengono dall'accademia alla creazione di progetti collaborativi e la comunicazione dei risultati della ricerca storica al pubblico della rete.

L'intervento intende riflettere su come la scrittura collaborativa, resa possibile dagli strumenti wiki, possa costituire un ottimo punto di partenza per una democrazia partecipata che si situi nel solco delle citizen humanities. L'ipotesi è che un progetto di citizen humanities che sia proficuo debba prevedere il coinvolgimento di più attori – non, dunque, escludere gli addetti ai lavori come gli storici di professione – nonché creare una linea di continuità con le pratiche delle public and digital humanities. Accademici, istituzioni culturali, responsabili delle politiche culturali e cittadini dovrebbero stabilire una collaborazione al fine di dare vita a progetti in cui si fondano le citizen humanities e la digital public history nel solco di una democrazia partecipata. Il Manifesto Digital Humanities 2.0 [6] presenta evidenti connessioni con le citizen humanities, dal momento che attesta come le tecnologie digitali abbiano cambiato le modalità di diffusione della conoscenza, a partire dalle sedi in cui si esprime il discorso scientifico. Poiché le università non sono più le uniche custodi della cultura e della conoscenza, si sta imponendo una conoscenza locale di cui sono depositari i cittadini, i quali profittano delle tecnologie digitali e della cassa di risonanza del web per condividere il sapere della comunità di riferimento. Ciò comporta un mutamento di prospettiva: anziché porre l'accento sul contributo dei cittadini alla conoscenza scientifica, risulta più proficuo immaginare progetti che connettano discipline storiche e digital humanities per rendere effettive le citizen humanities [2]. Come sostenuto da Muki Haklay [3] si possono distinguere quattro tipologie di attività di CS a seconda del grado di coinvolgimento dei soggetti partecipanti: il primo livello è quello del crowdsourcing, che contempla tutti quei progetti di CS in cui i cittadini agiscono come “sensors”, ossia si limitano a reperire i dati sulla base di indicazioni precise; il secondo livello viene chiamato “distributed intelligence” e fa riferimento all'iniziativa dei cittadini che mettono in campo le loro competenze per fornire l'interpretazione dei dati raccolti; il terzo livello è quello della condivisione, la cosiddetta “participatory science”, che implica il coinvolgimento dei cittadini nelle fasi di definizione delle ipotesi di ricerca e della metodologia più indicata; infine il livello “extreme” contempla tutte le possibili attività che possono svolgere i cittadini: dalla raccolta dei dati, all'interpretazione dei risultati e alla formulazione dei quesiti di ricerca. I progetti di citizen humanities che rientrano nel campo di applicazione della Digital Public History prevedono sinora il primo livello del crowdsourcing [4], quello che si riferisce alla citizen science contributiva, dove una comunità di cittadini raccoglie fonti che possono provenire, ad esempio, da archivi privati o dal web. Può trattarsi di corrispondenza epistolare, di documenti audiovisivi, fonti iconografiche, fonti born digital o digitalizzate. Questa mole documentaria diviene un patrimonio della cittadinanza che partecipa attivamente alla diffusione della conoscenza. Il primo livello risulta però eminentemente classificatorio e autoreferenziale perché si limita a fornire dati alla comunità degli storici. Nella citizen science collaborativa, quella dell'intelligenza distribuita, rientrano invece tutti quei progetti wiki che grazie alla scrittura collaborativa consentono ai contributori di presentare contenuti storici suffragati da fonti sulla base di una linea interpretativa. Uno degli strumenti per la gestione collaborativa dei progetti è rappresentato dal wiki, che consente di creare progetti in formato testuale, ipertestuale e ipermediale, in modalità collaborativa. Chi adopera gli strumenti wiki ha la possibilità di intervenire sulle stesse pagine per modificarle, eliminarle, implementarle e, cosa ancora più interessante, può visionare la sequenza storica delle modifiche.

Per certi versi si potrebbe sostenere come Wikipedia rappresenti una voce fuori dal coro rispetto ai progetti wiki che raggiungono il livello dell'intelligenza distribuita. Il cosiddetto "punto di vista neutrale", uno dei cinque pilastri su cui si fonda Wikipedia, esclude infatti che i contributori propongano un'interpretazione ma implica uno sforzo di metodo per raggiungere una "neutralità interpretativa". Questo contraddice il metodo storico e i dettami della disciplina, dal momento che il lavoro di ricerca dello storico consiste proprio nel ricostruire il passato attraverso l'analisi e l'interpretazione delle fonti. Le voci di carattere storico non possono prevedere una neutralità benché Wikipedia abbia una vocazione enciclopedica. Questo è ancor più lampante quando si trattano argomenti che hanno una rilevanza nel dibattito pubblico. Poste queste necessarie premesse bisogna riconoscere alla Wikimedia foundation di aver avviato alcune iniziative che vanno nella direzione di facilitare un dialogo tra le citizen humanities e la digital public history. Possiamo, ad esempio, citare, "wikipediano in residenza", che permette all'utente di trascorrere un periodo di formazione presso l'istituzione culturale di cui si intende fornire una serie di voci. In questo modo il wikipediano potrà formarsi e potrà acquisire quello spirito critico e interpretativo che è il fondamento del metodo storico. Il terzo livello della "scienza partecipata" – presupposto per una democrazia partecipata – in prospettiva, è particolarmente indicato per i progetti wiki: perché siano poste le condizioni per una "participatory science" è però necessario che siano compiute alcune azioni volte a dare una formazione ai cittadini. In primo luogo occorre costituire un gruppo di lavoro all'interno della comunità scientifica che fornisca il supporto formativo adeguato alle comunità locali e alle scuole intenzionate a dar vita a progetti wiki. Wikisource è la biblioteca digitale che ospita testi di pubblico dominio importati, formattati e riletti da volontari, ma anche documenti digitalizzati di vario tipo. Questo progetto, per fare un esempio tra i tanti, può servire come una palestra per le citizen humanities. Ma perché ciò avvenga occorre che gli storici – insegnanti e docenti universitari – contemplino nella loro didattica, non solo una parte di laboratorio dedicata agli strumenti wiki, ma anche il Webquest, ossia quel metodo didattico che chiama gli studenti ad effettuare ricerche in rete sotto la supervisione dell'insegnante.

## BIBLIOGRAFIA

1. Adamson, Joni. «Gathering the desert in an urban lab: designing the citizen humanities». In *Humanities for the environment: integrating knowledge, forging new constellations of practice*, a cura di Joni Adamson e Michael Davis, 106–19. London: Routledge, 2016.
2. Belknap, Geoffrey. «From Citizen Science to Citizen Humanities – 19th Century History in the Digital Age — University of Leicester». Blog Entry, 2015. <https://www2.le.ac.uk/offices/press/think-leicester/science-and-environment/2015/from-citizen-science-to-citizen-humanities-2013-19th-century-history-in-the-digital-age>.
3. Haklay, Muki. «Citizen Science and Volunteered Geographic Information – overview and typology of participation». a cura di Daniel Z. In Sui, Sarah Elwood, e Michael F. Goodchild, 105–22. Springer, 2013.
4. Hedges, Dunn. *Academic crowdsourcing in the humanities. Crowds, communities and co-production*. Amsterdam: Elsevier, 2018.
5. Serrano Sanz, Fermin, Teresa Holocher-Ertl, Barbara Kieslinger, Francisco Sanz García, e Cândida G Silva. *White Paper on Citizen Science for Europe*. Bruxelles: European Commission, 2014. [http://www.socientize.eu/sites/default/files/white-paper\\_0.pdf](http://www.socientize.eu/sites/default/files/white-paper_0.pdf).
6. UCLA (publisher). «A Digital Humanities Manifesto » The Digital Humanities Manifesto 2.0», 2009. <http://manifesto.humanities.ucla.edu/2009/05/29/the-digital-humanities-manifesto-20/>.

# Digital divide? Un PRoF per ogni persona!

Flavia Marzano

Stati Generali dell'Innovazione – flavia.marzano [at] gmail.com

## SINTESI

Il divario digitale di cui soffre l'Italia è chiaramente evidenziato dai dati di alcuni rapporti europei. Analizziamo in particolare l'Indice di Digitalizzazione dell'Economia e della Società 2020 (DESI). Presentiamo alcune importanti esperienze di Pubbliche Amministrazioni italiane volte a superare tale divario e a garantire la cittadinanza digitale.

Sono indicate inoltre gli obiettivi governativi e i relativi interventi previsti sul tema.

## PAROLE CHIAVE

digital divide, cittadinanza digitale, competenze digitali

## INTERVENTO

### 1 INTRODUZIONE

Le vicende degli ultimi tempi hanno dimostrato l'importanza e l'urgenza di una reale trasformazione digitale sia per le imprese che per le pubbliche amministrazioni e di conseguenza per tutti i cittadini.

Questo implica un idoneo livello di infrastrutture e di connettività oltre che la creazione di un ampio spettro di servizi online da parte della Pubblica Amministrazione.

Purtroppo anche i dati più recenti dimostrano che gli italiani non sempre hanno adeguate competenze digitali.

Il problema può essere affrontato a vari livelli e soprattutto va garantita l'alfabetizzazione digitale per dotare di migliori capacità di accesso ai servizi online gruppi che per motivi di età, genere, provenienza, condizioni socio-economiche e culturali hanno maggiori difficoltà.

### 2 I DATI DEL DIGITAL DIVIDE ITALIANO

Parlando di digital divide di solito si intende il divario infrastrutturale e purtroppo, come ben evidenziato dal Grafico 1, l'Italia per quanto riguarda la "fibra a casa" (FTTH) è davvero ancora molto indietro rispetto al resto dei paesi europei anche se per quanto riguarda la connettività in senso più ampio è in linea con la media europea posizionandosi, nell'Indice di Digitalizzazione dell'Economia e della Società 2020 (d'ora in poi DESI), al 17mo posto e nel campo del 5G è al di sopra della media europea del 20% con il 60%.



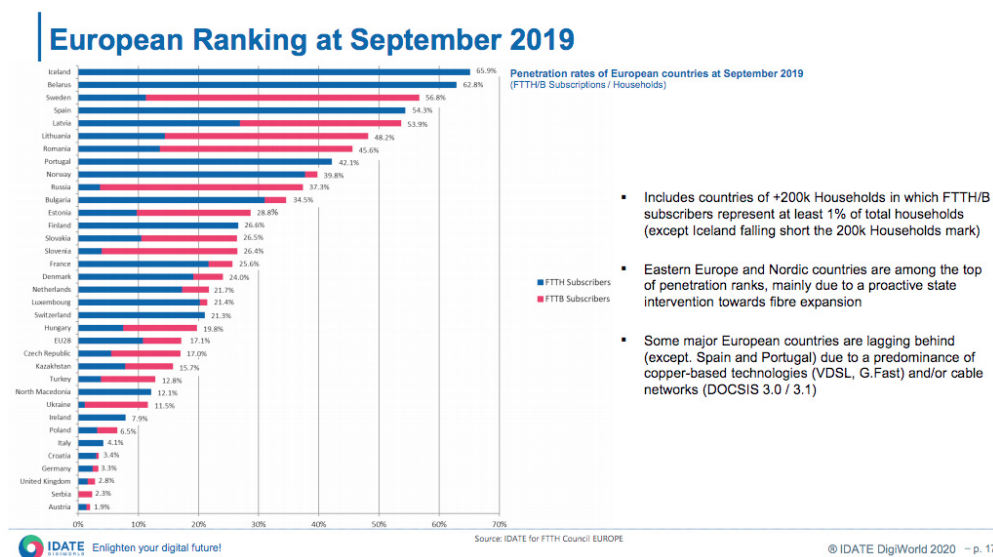


Grafico 1

L'Italia è poi superiore alla media europea per completezza dei servizi online, servizi pubblici digitali per le imprese e open data come evidenziato in Tabella 1.

|   | DESI 2018  | Italia     | DESI 2020  | UE         |
|---|------------|------------|------------|------------|
|   | valore     | valore     | valore     | valore     |
| <b>5a1 Utenti e-government</b>  | <b>30%</b> | <b>37%</b> | <b>32%</b> | <b>67%</b> |
| % degli utenti di Internet tenuti a presentare moduli                   | 2017       | 2018       | 2019       | 2019       |
| <b>5a2 Moduli precompilati</b>  | <b>33</b>  | <b>48</b>  | <b>48</b>  | <b>59</b>  |
| Punteggio (da 0 a 100)  | 2017       | 2018       | 2019       | 2019       |
| <b>5a3 Livello di completezza dei servizi online</b>                    | <b>89</b>  | <b>91</b>  | <b>92</b>  | <b>90</b>  |
| Punteggio (da 0 a 100)  | 2017       | 2018       | 2019       | 2019       |
| <b>5a4 Servizi pubblici digitali per le imprese</b>                     | <b>81</b>  | <b>82</b>  | <b>94</b>  | <b>88</b>  |
| Punteggio (da 0 a 100) - iniziative nazionali e transnazionali comprese | 2017       | 2018       | 2019       | 2019       |
| <b>5a5 Open Data</b>  | <b>NA</b>  | <b>NA</b>  | <b>77%</b> | <b>66%</b> |
| % del punteggio massimo   |            |            | 2019       | 2019       |

Tabella 1

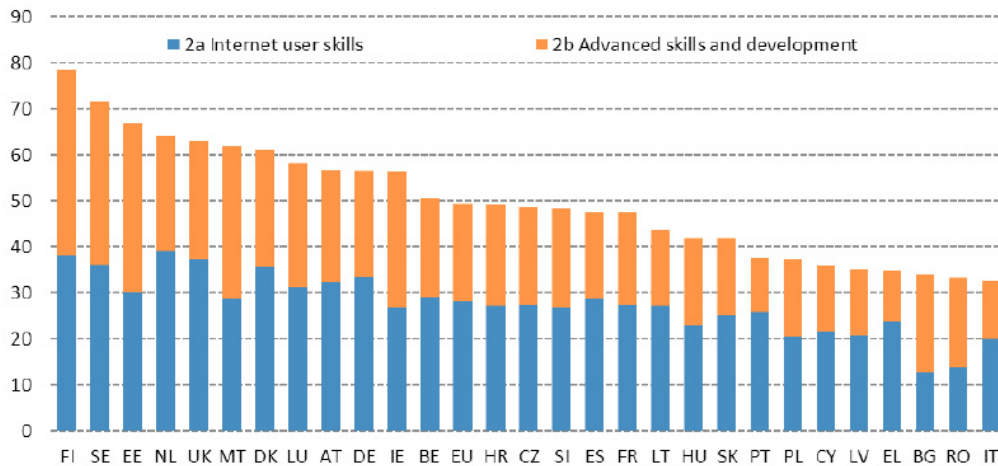
Per quanto riguarda invece il divario digitale relativo a servizi (della Pubblica Amministrazione e per le imprese) nel DESI si vede con estrema chiarezza che l'Italia (al quartultimo posto) ha, rispetto alla media europea, livelli di competenze digitali di base e avanzate davvero molto bassi.

Resta infatti scarsissimo l'uso dei servizi pubblici digitali (e-government) che sono molto al di sotto della media europea: la posizione in classifica del paese è al 26° posto su 28 e anche le imprese italiane presentano ritardi nell'utilizzo di tecnologie come il cloud computing e i big data, così come per quanto riguarda l'adozione del commercio elettronico e la percentuale di individui che non hanno mai usato internet in Italia è il 17% contro la media europea del 9% e solo il 74% degli italiani usa abitualmente Internet.

Infine il numero di specialisti nel settore ICT è molto al di sotto della media UE (2,8% vs 3,9% sull'occupazione totale), i laureati ICT in Europa sono il 3,6% e in Italia solo l'1% mentre per le donne la situazione è ancora più grave: donne ICT sul totale delle occupate 1,4% in EU e solo 1,0% In Italia. La situazione in sintesi è quella rappresentata dai Grafici 2 e 3.



Figure 45 Human capital dimension (Score 0-100), 2019



Source: DESI 2020, European Commission.

Grafico 2

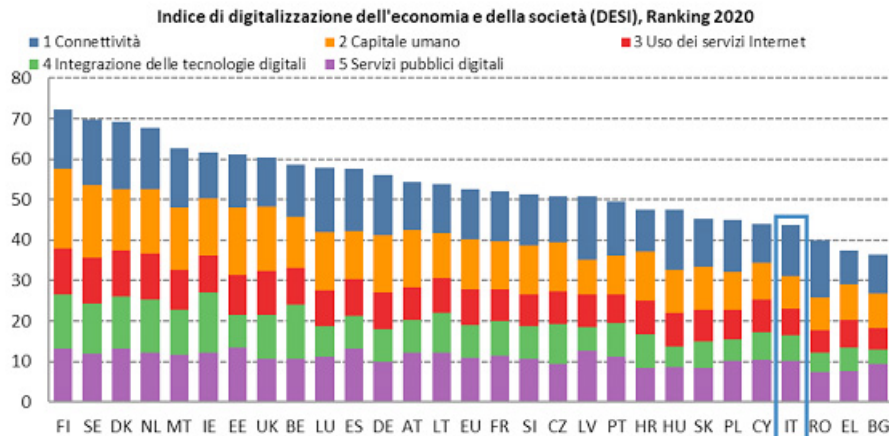


Grafico 3

### 3 QUALI SOLUZIONI? ALCUNE ESPERIENZE ITALIANE

Per poter fruire di servizi digitali e per l'esercizio dei diritti diventa quindi necessario prevedere momenti formativi e "punti di accesso assistiti" dove i cittadini possano essere supportati non solo per accedere ai servizi digitali ma anche per essere in qualche modo formati per diventare davvero cittadini digitali.

Diverse sono le esperienze in Italia a partire dalle Regioni che hanno declinato in maniera diversa i vari interventi, di seguito ne descriviamo alcune.

La Toscana ha attivato già nel 2004 i PAAS, Punti di Accesso Assistito ai servizi online e a internet, che hanno le seguenti funzioni prioritarie: "qualificare la domanda di accesso alla rete, ampliando le competenze digitali dei cittadini, alfabetizzando coloro che non hanno conoscenze digitali, favorendo l'accesso alle tecnologie e ai contenuti da esse veicolati, garantendo il diritto di accesso alla rete per tutti e soprattutto alla popolazione svantaggiata che rischia l'esclusione dal mondo digitale (ICT); diffondere e promuovere l'offerta dei servizi online per i cittadini che le Pubbliche Amministrazioni stanno realizzando sulla rete, facilitando e sostenendone l'accesso, orientando gli utenti, mediando rispetto ai contenuti e alle interfacce."

La Regione Emilia-Romagna ha predisposto i Punti Pane e Internet come "base per la diffusione della conoscenza sui servizi online per i cittadini, ... per l'esercizio dei diritti di cittadinanza digitale".

In Regione Umbria i DigiPASS sono luoghi fisici attrezzati "per l'accesso assistito alle risorse digitali, alle tecnologie ed alla crescita delle competenze digitali."

A livello comunale Roma Capitale ha progettato nel 2016 i PRoF, Punti Roma Facile, spazi assistiti "dove i "facilitatori digitali", operatori e volontari appositamente formati, sono a disposizione degli utenti che hanno bisogno di indicazioni e consigli relativi all'uso del computer, alla navigazione in rete e all'accesso ai principali servizi online.

Il servizio si rivolge in particolare a chi ha poca familiarità con Internet e con le nuove tecnologie e a chi voglia imparare ad accedere agilmente ai servizi pubblici on line di Roma Capitale e di altre amministrazioni, rendendosi gradualmente autonomo nell'uso delle nuove tecnologie."

I PRoF servono "ad accrescere le competenze digitali, a partecipare più attivamente alla vita sociale della comunità e a semplificare il rapporto con la pubblica amministrazione.

Presso la rete dei Punti Roma Facile è possibile essere supportati nella richiesta di attivazione dell'identità digitale (SPID), avere un primo approccio con i principali strumenti informatici, imparare a utilizzare i servizi digitali di Roma Capitale (prenotazione on line di appuntamenti e servizi, certificati anagrafici e iscrizioni scolastiche, pagamento dei tributi e delle contravvenzioni, ecc.) e quelli di altre amministrazioni pubbliche.

I Punti Roma Facile sono attivi nel territorio di tutti i Municipi. L'assistenza è gratuita e su prenotazione, con appuntamenti di circa 30 minuti. In ogni Punto vengono, inoltre, proposti periodici seminari aperti a tutti per approfondire, con l'aiuto di esperti, argomenti specifici riguardanti la cultura e le competenze digitali."

In questo contesto il Progetto “Giovani per l’innovazione” ha previsto la formazione di alcuni ragazzi del servizio civile per metterli in grado di svolgere l’attività di facilitatori digitali nei PRoF. I volontari supportano “coloro che hanno bisogno di indicazioni, spiegazioni e consigli relativi all’uso delle applicazioni e dei software più comuni, alla navigazione in rete e all’accesso ai principali servizi online. Il progetto ha, da una parte, il fine di promuovere la cultura digitale e l’uso responsabile delle nuove tecnologie, dall’altra, vuole favorire e diffondere l’accesso ai servizi online.”

Quest’ultima esperienza è stata ripresa nel Piano Triennale ICT 2020-2022, Obiettivo OB.8.2 “Rafforzare le competenze digitali per la PA e per il Paese e favorire l’inclusione digitale” che prevede, entro Dicembre 2020, l’impostazione del progetto “di Servizio Civile Digitale e identificazione della fase pilota, per promuovere, nell’ambito del Servizio Civile Universale coordinato dal Ministro per le Politiche Giovanili e lo Sport, progetti di facilitazione digitale e di supporto alla trasformazione digitale della PA e all’uso di servizi pubblici digitali”.

A livello europeo, il Digital Competence Framework, noto anche come DigComp, offre uno strumento per migliorare la competenza digitale dei cittadini. Oggi, essere competenti digitalmente significa che le persone devono avere competenze in tutte le aree di DigComp. AgID (l’Agenzia per l’Italia Digitale), ha tradotto il documento DigComp 2.1 “Il quadro di riferimento per le competenze digitali dei cittadini” con otto livelli di padronanza ed esempi di utilizzo.

Anche per implementare quanto definito in DigComp, un contesto non ancora sufficientemente esplorato, ma che sarebbe invece importante sviluppare è quello della formazione accademica dei formatori, ossia di tutti quei soggetti che avrebbero l’importante compito di educare quella componente della popolazione che per età o per storia personale non ha competenze digitali ad acquisirle in maniera adeguata. Questo tipo di formazione ha un carattere palesemente interdisciplinare in quanto, se da un lato per questo compito occorrono le competenze tipiche che si acquisiscono in ambito umanistico (scienza della formazione, pedagogia, sociologia), dall’altro è essenziale che il formatore abbia competenze digitali sufficientemente ampie e articolate non solo come utente ma in quanto formatore. Questo tipo di competenze si collocano in maniera del tutto naturale nell’alveo delle cosiddette digital humanities.

## ULTERIORI RIFERIMENTI

1. DigComp - <https://ec.europa.eu/jrc/en/digcomp>
2. DigComp 2.1 Il quadro di riferimento per le competenze digitali dei cittadini  
[https://www.agid.gov.it/sites/default/files/repository\\_files/digcomp2-1\\_ita.pdf](https://www.agid.gov.it/sites/default/files/repository_files/digcomp2-1_ita.pdf)
3. DigiPASS e LivingLAB  
<https://www.regione.umbria.it/agenda-digitale/digipass>
4. The Digital Economy and Society Index (DESI)  
<https://ec.europa.eu/digital-single-market/en/desi>
5. FTTH Council Europe – Panorama Markets at September 2020  
<https://www.ftthcouncil.eu/documents/FTTH%20Council%20Europe%20-%20Panorama%20at%20September%202019%20-%20Webinar%20Version%20.pdf>
6. PAAS, Punto di Accesso Assistito ai Servizi online  
<https://www.regione.toscana.it/documents/10180/339418/PAAS+-+Scheda+informativa/80351069-ff6e-4f8c-8a41-d02c6c1d0b04?version=1.3>
7. Pane e Internet Cittadini 100% digitali - <https://www.paneeinternet.it/>
8. Piano Triennale ICT - <https://pianotriennale-ict.italia.it/piano/>
9. Punti Roma Facile - <https://www.comune.roma.it/web/it/partecipa-punti-roma-facile.page>
10. SPID - <https://www.spid.gov.it/>

# **Il modello partecipativo di Roma Capitale: esperienze di democrazia diretta e nuovi diritti digitali**

Antonella Caprioli

Dipartimento Partecipazione, Comunicazione e Pari Opportunità di Roma Capitale  
antonella.caprioli [at] comune.roma.it

## **SINTESI**

Roma Capitale ha accettato la sfida della democrazia partecipata mettendo in campo, nel corso degli ultimi due anni, nuovi strumenti - normativi, organizzativi e tecnologici - per consolidare e tradurre in pratica un'effettiva cultura della partecipazione, basata sul coinvolgimento dei cittadini nei processi decisionali legati alla gestione del territorio. Il contributo descrive tale percorso e le metodologie utilizzate, a partire dalla revisione dello Statuto e dei Regolamenti fino alla prima sperimentazione di Bilancio Partecipativo a livello capitolino, passando attraverso la creazione di nuove piattaforme telematiche e la promozione di iniziative per il contrasto al divario digitale.

## **PAROLE CHIAVE**

partecipazione, e-democracy, bilancio partecipativo, democrazia partecipata

## **INTERVENTO**

### **1 INTRODUZIONE**

Il dibattito pubblico sul valore della partecipazione popolare e sugli strumenti con i quali si declina negli ordinamenti contemporanei ha assunto, negli ultimi tempi, un peso specifico sempre più significativo. Sebbene si tratti di processi nati a partire già dagli anni Sessanta, l'introduzione di pratiche partecipative nella gestione della cosa pubblica ha conosciuto una prima importante accelerazione soltanto sul finire del secolo scorso, con l'affermazione su ampia scala delle esperienze di origine latino-americana, in particolare dei modelli adottati da alcune città brasiliane, come il Bilancio Partecipativo di Porto Alegre del 1989. Un processo di «democratizzazione della democrazia», come è stato efficacemente definito in letteratura, che ha successivamente investito, con tempistiche e ordini di grandezza differenti, il Nord America e l'Europa (Francia, Spagna e Italia soprattutto), trovando ispirazione in principi e idee comuni, ma connotandosi in modo diverso secondo il contesto (nonché la cornice giuridico-normativa) in cui tali pratiche si sono via via calate.

L'obiettivo di questo paper è illustrare le modalità con cui si è evoluto, negli ultimi anni, il concetto di partecipazione popolare nelle politiche gestionali-amministrative di Roma Capitale, secondo un modello che si è caratterizzato per una profonda trasformazione del rapporto fra “amministratori e amministrati” e per un maggiore (e più consapevole) coinvolgimento dei cittadini nei processi decisionali, attraverso l'utilizzo di strumenti - anche digitali - di democrazia diretta e partecipata.

## **2 PARTECIPAZIONE E CITTADINANZA DIGITALE**

L'amministrazione capitolina ha dato avvio a una rimodulazione degli strumenti con cui si realizza tradizionalmente la partecipazione popolare. Dal punto di vista normativo le novità più rilevanti sono state introdotte dallo Statuto, così come modificato dall'Assemblea Capitolina con due distinte deliberazioni nel gennaio del 2018. Il nuovo testo segna un punto di svolta, perché ridisegna l'intero impianto concettuale della partecipazione cittadina e introduce istituti inediti come il Bilancio partecipativo, il referendum propositivo, la petizione popolare elettronica. Il cambiamento di prospettiva è evidente, con l'Amministrazione che si pone obiettivi piuttosto ambiziosi. A cominciare da quello di favorire, con il supporto del digitale, l'impegno diretto dei cittadini nello sviluppo civile, economico e culturale della città, e da quello, più strettamente operativo, di migliorare la qualità dei propri atti attraverso l'utilizzo di forme di consultazione che indirizzino le azioni da adottare, in un'ottica collaborativa e di confronto con gli appartenenti alla comunità locale. È proprio in questo filone che si collocano alcuni progetti avviati nell'ultimo triennio da Roma Capitale. Tra i primi, in ordine cronologico, l'implementazione dell'Area Partecipa all'interno del portale istituzionale: si tratta di uno spazio a tema dove i cittadini possono trovare informazioni e strumenti utili per l'esercizio dei diritti di partecipazione nonché tutti gli aggiornamenti sulle iniziative, i progetti e i processi partecipativi promossi dall'Amministrazione, sia a livello centrale che in ambito municipale. Affinché il digitale possa esprimere in pieno la propria funzione di supporto a tali processi, tuttavia, diventa fondamentale l'impegno delle Pubbliche Amministrazioni nello sviluppo di adeguate politiche di contrasto al digital divide. Roma Capitale ha lavorato molto anche sotto questo profilo, portando avanti due importanti progetti, coordinati dal Dipartimento Partecipazione, Comunicazione e Pari Opportunità: i Punti Roma Facile e la Scuola Diffusa per la Partecipazione e la Cittadinanza Digitale. I primi sono spazi assistiti di supporto al cittadino nei quali si propongono attività di promozione della cultura digitale. L'obiettivo, in particolare, è quello di fornire ai cittadini gli strumenti di base necessari per orientarsi nell'uso delle tecnologie e metterli nelle condizioni di fruire correttamente dei principali servizi online della Pubblica Amministrazione, a partire da quelli offerti da Roma Capitale. Sono attivi su tutto il territorio capitolino e attualmente se ne contano 24, dislocati perlopiù nelle sedi municipali e in misura ridotta anche nelle Biblioteche e nei Centri Sociali Anziani. Avviata in fase sperimentale nel 2017, anche la Scuola Diffusa si pone obiettivi analoghi. Intende, infatti, favorire la cittadinanza attiva attraverso le opportunità offerte dal digitale e si concretizza in momenti formativi gratuiti e aperti a tutti: corsi, seminari e workshop incentrati sui temi delle nuove tecnologie, erogati anche in modalità diverse dalla formazione d'aula (scambio intergenerazionale, peer education, integrazione aula-online). Gli incontri della Scuola Diffusa - che si avvale, per il suo funzionamento, di un Comitato Scientifico per la programmazione delle attività - si svolgono principalmente nell'ambito della rete dei Punti Roma Facile e delle Biblioteche del circuito di Roma Capitale, ma possono tenersi anche nelle scuole e in altri spazi pubblici o contesti associativi.

### 3 IL BILANCIO PARTECIPATIVO

Il Bilancio Partecipativo ha rappresentato il momento culminante del processo di costruzione della e-democracy capitolina fin qui descritto. Se il primo passo, come detto, è stato quello di rafforzare i principi della partecipazione telematica a livello statutario, in seconda battuta si è provveduto alla regolamentazione di dettaglio con il Regolamento del Bilancio Partecipativo di Roma Capitale e il relativo Disciplinare, che hanno posto le basi per una metodologia chiara e replicabile, seppur con un certo carattere di sperimentaltà. A livello pratico la prima sfida è stata la predisposizione della piattaforma telematica, chiamata a coniugare molteplici istanze: da un lato la semplicità e l'usabilità per il cittadino, dall'altro i vincoli posti dalla normativa sulla privacy e le regole del portale istituzionale, nel quale il Bilancio è stato incardinato. I soggetti ammessi al processo sono stati gli stessi che lo Statuto riconosce quali titolari dei diritti di partecipazione: cittadini residenti; i cittadini non residenti che esercitano in città la propria attività prevalente di studio o lavoro; stranieri maggiorenni, legittimamente presenti sul territorio per ragioni di studio o lavoro. Queste ultime due tipologie rientrano nella macrocategoria dei city user, fruitori importanti di servizi e spazi urbani in una metropoli che costituisce un polo di attrazione naturale a livello nazionale e internazionale. Altri protagonisti del processo sono stati i Municipi, espressioni di quella parte dell'Amministrazione più vicina al territorio e più competente nella costruzione di un dialogo diretto con i cittadini e i Comitati di Quartiere. Proprio nell'ottica di garantire massima inclusività ed effettiva rappresentatività, a essi è stato dato il compito di farsi promotori di proposte condivise con la cittadinanza attraverso incontri e assemblee pubbliche, con l'idea che la dimensione online e quella offline non si escludano a vicenda, ma debbano coesistere per assicurare il coinvolgimento anche di quelle fasce di popolazione più deboli e non ancora digitalizzate. A tali proposte è stato riservato il 20% della quota base di risorse economiche destinate a ciascun territorio. Lo stanziamento complessivo è stato di 20 milioni di euro, ripartiti tra i 15 Municipi secondo una ratio che ha tenuto conto del peso demografico e dell'estensione dei diversi Municipi, con una quota di 2 milioni di euro come premialità per quei territori che si fossero dimostrati più virtuosi in termini di partecipazione. Infine, l'ambito tematico di riferimento è stato individuato nel decoro urbano: spazio, dunque, a interventi di ampio respiro in grado di valorizzare la bellezza e la fruibilità degli spazi urbani a uso collettivo attraverso un approccio multidisciplinare che prevedesse azioni integrate di recupero e riqualificazione - verde, arredi urbani, illuminazione etc. Il Bilancio Partecipativo si è svolto da giugno a ottobre 2019, in armonia con i tempi della programmazione economica capitolina. L'iter si è articolato in cinque fasi: 1. informazione, attraverso una campagna multicanale caratterizzata dall'hashtag #ROMA-DECIDE, divenuto elemento distintivo della partecipazione capitolina; 2. presentazione delle proposte e raccolta dei sostegni online; 3. valutazione tecnico-amministrativo-finanziaria; 4. votazione finale; 5. Attuazione e programmazione. Un particolare ruolo di mediazione sul territorio è stato svolto anche dai Punti Roma Facile, che hanno risposto con efficacia al loro compito di facilitazione all'uso degli strumenti telematici, favorendo un più effettivo e incisivo esercizio dei diritti di partecipazione. Le proposte ricevute dai cittadini sono state 1.481, mentre quelle passate attraverso la mediazione dei Municipi sono stati 40, per un totale complessivo di 1.521 progetti. Temi ricorrenti sono stati la riqualificazione di piazze e aree urbane, la valorizzazione di parchi e aree verdi, la creazione di aree gioco attrezzate e aree cani. Grande attenzione anche ai temi dell'accessibilità e della sicurezza e ai percorsi di mobilità alternativa ciclo-pedonale.



Su tutto, è emerso chiaro il desiderio di riappropriarsi della città, contribuendo in prima persona al processo di ri-scoperta e fruizione consapevole del territorio da parte della comunità. Le 193 proposte dei cittadini che hanno superato la prima scrematura (data dalla raccolta dei sostegni online, con soglia di sbarramento del 5%) sono state esaminate da un Tavolo dell'Amministrazione, incaricato di valutarne la sostenibilità e la fattibilità, e infine sottoposti a un'ulteriore votazione online. Nel complesso sono stati raccolti 36.128 voti da parte di circa 17mila votanti; gli interventi ammessi a finanziamento sono stati 65. I numeri della partecipazione parlano dunque di un'esperienza importante e più che soddisfacente, soprattutto se si considerano le caratteristiche ancora sperimentali dello strumento adottato dall'Amministrazione capitolina e di una familiarità con queste pratiche, da parte della comunità locale, non ancora pienamente sviluppata. La chiusura del processo è stata formalizzata con l'inserimento dei fondi nel Bilancio di previsione 2020-2022.

## **4 LE NUOVE SFIDE**

Un progetto come quello descritto non termina con l'identificazione delle opere "vincitrici", ma mira a costruire un percorso di trasparenza e rendicontazione ai cittadini di quanto effettivamente realizzato. Nell'ambito dell'evoluzione dell'area Partecipa è in fase di rilascio un nuovo sistema di tracking dello stato di avanzamento dei lavori di ciascuna opera, nel quale la narrazione viene filtrata attraverso una serie di icone rappresentative delle tappe fondamentali dell'iter (assegnazione del budget, progetto, gara d'appalto, lavori in corso, chiusura lavori). Duplice l'obiettivo: fungere da stimolo interno all'Amministrazione e sostenere ulteriormente la partecipazione, nella consapevolezza che, una volta calato il sipario, quanto progettato insieme non rimarrà lettera morta. Traendo le conclusioni sugli obiettivi raggiunti, questo primo Bilancio ha rappresentato un successo in termini di coinvolgimento di categorie anche molto diversificate di popolazione, in un contesto cittadino di per sé complesso e dispersivo: una riconferma del fatto che le tecnologie della comunicazione possono fornire un più ampio respiro e una maggiore legittimazione ai tradizionali istituti della democrazia rappresentativa. L'evoluzione del percorso dovrà attribuire un ruolo ancora maggiore alle rappresentanze municipali, veri mediatori delle istanze dei diversi territori, e un peso specifico più consistente anche ad altre realtà rappresentative degli interessi delle comunità locali, in modo da consentire forme di co-progettazione più incisive - online e offline - lungo tutte le fasi del processo. Per agevolare questo dialogo costruttivo con il territorio è già allo studio l'organizzazione di moduli formativi rivolti ai dipendenti capitolini - in particolare quelli municipali - sulle tecniche di facilitazione e di gestione dei processi partecipativi. Solo così la partecipazione potrà evolvere da case-history a strumento riconosciuto di governo "condiviso" della città.

## BIBLIOGRAFIA

1. Allegretti, Umberto, a c. di. *Democrazia partecipativa: esperienze e prospettive in Italia e in Europa*. Firenze: Firenze University Press, 2010.
2. Carloni, Enrico. *L'amministrazione aperta: Regole strumenti limiti dell'open government*. Maggioli, 2014.
3. Faini, Fernanda. *Data society. Governo dei dati e tutela dei diritti nell'era digitale*. Giuffrè Francis Lefebvre, 2018.
4. Floridi, Luciano. *Il verde e il blu: Idee ingenue per migliorare la politica*. Raffaello Cortina Editore, 2020.
5. Floridia, Antonio. *Un'idea deliberativa della democrazia. Genealogia e principi*. Bologna: Il Mulino, 2017.
6. Pepe, Gabriele. *Il modello della democrazia rappresentativa tra aspetti teorici e profili applicativi*. Padova: Cedam, 2020.
7. Roma Capitale. «Documento della Partecipazione», 2019.
8. ———. «Regolamento del Bilancio Partecipativo di Roma Capitale», 2019.
9. ———. «Statuto di Roma Capitale», 2018.
10. Sorice, Michele. *Partecipazione democratica. Teorie e problemi*. Milano: Mondadori Università, 2019.
11. Trojani, Fabio, a c. di. *Il nuovo Codice dell'Amministrazione Digitale*. Santarcangelo di Romagna: Maggioli, 2016.
12. Urbinati, Nadia. *Democrazia rappresentativa. Sovranità e controllo dei poteri*. Roma: Donzelli, 2010.
13. Vilella, Giancarlo. *Introduzione alla e-democracy*. Bologna: Pendragon, 2019.



# La partecipazione democratica al tempo delle nuove tecnologie: l'e-Democracy

Luigi Rufo  
Università di Pisa, Italia, luigi.rufo [at] jus.unipi.it

## SINTESI

Il presente contributo mira ad investigare sul potenziale dell'ecosistema delle piattaforme digitali per l'esercizio della democrazia diretta e partecipata.

## PAROLE CHIAVE

e-Democracy, partecipazione democratica, voto elettronico

## INTERVENTO

### 1 INTRODUZIONE: LA PARTECIPAZIONE DEMOCRATICA

In un periodo storico in cui lo sviluppo tecnologico è sempre più predominante, inevitabilmente la società contemporanea sta risentendo di una sostanziale riscrittura di molte dinamiche sociali<sup>1</sup>, una vera e propria rivoluzione digitale<sup>2</sup> che interessa da vicino anche i principi posti a fondamento della democrazia, con una viva ed attenta discussione sugli strumenti da utilizzare per una incisiva manifestazione di "sovranità"<sup>3</sup> popolare.

Si possono immaginare, infatti, sistemi di partecipazione in grado di annullare i limiti spaziali e temporali e di prefigurare scenari nei quali i cittadini possono partecipare alla cosa pubblica ovunque si trovino, procurando così un doppio effetto positivo, ovvero da un lato un maggiore indice di partecipazione popolare da un lato e dall'altro un ingente risparmio, sul piano organizzativo, per la spesa pubblica.

---

1 Per un approfondimento si consiglia: G. Mazzoli, *Profili sociali della comunicazione e nuove tecnologie. Mondi vitali e mondi artificiali: dalla polarizzazione all'interfaccia?*, Franco Angeli, 1992; S. Zuboff, *Il capitalismo della sorveglianza. Il futuro dell'umanità nell'era dei nuovi poteri*. Roma, LUISS University Press, 2019; M. Castells, *The rise of the Network Society*, Blackwell Publishing, Oxford, 1996 (traduzione italiana: *La nascita della società in rete*, Università Bocconi Editore, Milano, 2002).

2 P. Ferri, *La rivoluzione digitale. Comunità, individuo e testo nell'era di Internet*, Eterotopie, 2001.

3 Del resto, già V. Crisafulli, *La sovranità popolare nella Costituzione italiana (Note preliminari)*, in *ID, Stato, popolo, governo: Illusioni e delusioni costituzionali*, Milano, Giuffrè, 1985, p. 122 chiariva come "l'esercizio della potestà sovrana di un popolo non si esaurisce sempre e soltanto in manifestazioni unitarie di volontà, spesso viceversa articolandosi in una serie di atti parziali e particolari che in diversi modi concorrono alla formazione, all'esplicazione ed all'attuazione della potestà medesima, unitariamente considerata".

Il potenziale insito nelle nuove tecnologie appare infatti in grado di rivitalizzare l'interesse di cittadini oramai disillusi e insoddisfatti di una politica sempre più espressione mediata dei propri interessi e preferenze e di procedure democratiche tradizionali, caratterizzate dalla lentezza nelle risposte ai problemi da parte delle assemblee rappresentative<sup>4</sup>. Ad avvalorare ciò, Stefano Rodotà ha addirittura teorizzato l'avvento di una "democrazia immediata"<sup>5</sup> caratterizzata da referendum istantanei, consultazioni permanenti dei cittadini in grado di garantire un rafforzamento e una maggiore legittimazione delle procedure democratiche, assicurando così un aumento della partecipazione dei cittadini alla vita pubblica e alla deliberazione democratica<sup>6</sup>.

Appare tuttavia doveroso segnalare anche la presenza di un atteggiamento molto scettico nei confronti di tali rivoluzionarie virtù "democratizzanti" delle nuove tecnologie dell'informazione<sup>7</sup>. Secondo questo approccio critico, le nuove tecnologie appaiono infatti capaci di determinare sia una spersonificazione dei rapporti sociali che di determinare nuove forme di esclusione legate al *digital divide* destinate ad allontanare molti individui dalle tecnologie alcuni, impedendo così quella prospettiva di agevolazione e di semplificazione la partecipazione democratica dei cittadini. Ciò finirebbe così per creare un paradosso per cui, rimosse, con la *Dichiarazione universale dei diritti dell'uomo* del 1948, le barriere "classiche" che bloccavano l'affermazione dello stato democratico (proclamando l'universalità del voto all'art. 21, comma 3) per ragioni di *razza*, di *colore*, di  *sesso*, di *lingua*, di *religione*, di *opinione politica* o, più in generale, di *pensiero*, di *origine nazionale o sociale*, di *ricchezza*, di *nascita* (art. 2, comma 1), verrebbero a crearsi nuove divisioni basate sull'utilizzo della moderna tecnologia digitale.

Secondo un'altra chiave di lettura, la democrazia elettronica rappresenta un raffinato strumento di sofisticazione della volontà popolare destinato, in assenza di adeguate garanzie tecniche, di alterare le caratteristiche strutturali del voto democratico, ossia *in primis* la libertà e la segretezza<sup>8</sup>. Verrebbe così a materializzarsi il rischio di una democrazia malata, frutto di pensieri e pregiudizi cognitivi.

---

4 Per una maggior approfondimento si veda G. Gometz, *Democrazia elettronica. Teoria e tecniche*, ETS, Pisa, 2017, pag. 38 e ss.

5 Cfr. S. Rodotà, *Tecnopolitica. La democrazia e le nuove tecnologie della comunicazione*, Laterza, Roma-Bari, 2004; S. Rodotà, *Iperdemocrazia. Come cambia la sovranità democratica con il web*, Laterza, Roma-Bari, 2013.

6 In questo senso merita un'approfondita lettura una sentenza della Corte costituzionale nella quale viene sottolineato che il pluralismo delle fonti di informazione, aprendo così la strada all'uso dei nuovi media e strumenti digitali, è necessario al fine di poter suscitare nei cittadini un senso di appartenenza alle questioni della res publica, rappresentando un presupposto indispensabile "per l'attuazione ad ogni livello, centrale o locale, dello Stato democratico", v. Corte Costituzionale, sentenza n. 348 del 1990

7 Sul punto si veda E. Morozov, *L'ingenuità della rete. Il lato oscuro della libertà di internet*, Codice, 2011.

8 Per un approfondimento: A. G. Orofino, *Il voto elettronico*, in G. Cassano (a cura di), *Diritto delle nuove tecnologie informatiche e dell'internet*, Milano, 2002, 247; M. Cuniberti, *Tecnologie digitali e libertà politiche*, in *Dir. inf. infor.*, 2015, 296; S. Elia, *Alcune questioni problematiche circa il voto elettronico*, in *Cyberspazio e diritto*, 2016, 95; P. Carlotto, *Il voto*, cit., 87; M. Folarin Tawa – G. Gometz, *Voto elettronico presidiato e blockchain*, in *Ragion pratica*, 2018, 319 ss.

Alla luce di ciò, decisivo appare l'intervento dello Stato, inteso non solo come "regolatore", ma anche quale "custode" di una vera e propria democrazia elettronica in cui a tutti deve essere assicurato l'accesso sicuro e neutrale agli strumenti digitali<sup>9</sup>, assicurando un uso delle nuove tecnologie volto a rafforzare il valore centrale della sovranità popolare<sup>10</sup>. Peraltro, significativamente, l'art. 9 del Codice dell'Amministrazione digitale, approvato con d.lgs. n. 82/2005, già prevede che debba essere favorita "ogni forma di uso delle nuove tecnologie per promuovere una maggiore partecipazione dei cittadini, anche residenti all'estero, al processo democratico e per facilitare l'esercizio dei diritti politici e civili".

## 2 L'EVOLUZIONE DELLA DEMOCRAZIA: VERSO L'E-DEMOCRACY

Il passaggio dal voto espresso attraverso la tradizionale scheda cartacea a quello per via elettronica si sta diffondendo progressivamente in diversi Paesi e le nuove tecnologie digitali rappresentano sempre più uno "strumento" capace di superare qualsiasi limite spaziale e temporale, permettendo ai cittadini di aggregarsi, creare dibattiti ed esprimersi nelle proprie volontà politiche senza alcun vincolo.

Tuttavia, tra le grandi sfide che dovranno essere affrontate, al fine di poter giungere all'idea avveniristica di un'espressione della volontà dell'elettore veicolata attraverso un terminale elettronico, vi è senz'altro quella di dover trovare il giusto linguaggio comunicativo in grado di valorizzare il tema del divario delle conoscenze digitali tra le varie generazioni. Infatti, si avrà una vera e propria democrazia elettronica solo se la partecipazione digitale dei cittadini inizia ad essere indispensabile per la determinazione del contenuto delle scelte pubbliche<sup>11</sup>, ovvero è possibile creare un "canale informativo" capace di assicurare al cittadino un ruolo pienamente attivo<sup>12</sup>.

Del resto, proprio in questi ultimi anni sono nate nuove forme di movimenti politici e d'iniziativa popolari, quali i partiti "assembleari" (Movimento 5 Stelle, *Podemos*, *France Insoumise* e i partiti pirata) o le liste civiche (nate da piattaforme *online* come *Barcelona en Comú - BComú*), tutti caratterizzate dalla volontà di assicurare una democrazia resa più partecipativa grazie alle nuove tecnologie dell'informazione.

---

9 Sul punto si veda L. Trucco, Il voto elettronico nella prospettiva italiana e comparata, in *Dir. inf. infor.*, 2011.

10 Sul punto si veda P. Costanzo, La democrazia elettronica (note minime sulla cd. e-democracy), in *Dir. inf. infor.*, 2003, 472 ss., il quale afferma che c'è una forte necessità "per le attuali democrazie di continuare a strutturarsi fondamentalmente a livello sociale secondo logiche aggregative e, a livello istituzionale, secondo moduli rappresentativi rispetto ai quali la democrazia elettronica non sembra poter dunque giocare un ruolo completamente sostitutivo".

11 G. Gometz, *Democrazia elettronica*, cit..

12 Per meglio comprendere l'evoluzione dell'attuale contesto della democrazia doveroso è ricordare il passaggio avutosi negli ultimi decenni da una forma di "democrazia diretta", in cui i politici e il corpo elettorale intervenivano in modo collaborativo con strumenti come referendum abrogativo o confermativo alla "democrazia rappresentativa", fondata sull'intervento indiretto dei cittadini, i quali sceglievano coloro che potevano ricoprire le cariche degli organi supremi dello stato e li delegavano per agire in loro rappresentanza, fino ad arrivare all'accrescimento di una democrazia "digitalizzata" avente come principio fondante la volontà dello sviluppo di un processo partecipativo che inverte la tendenza sfiduciaria verso il sistema politico tradizionale e i rispettivi governi. Per un maggior approfondimento si rimanda a N. Urbinati, *Democrazia rappresentativa*, Roma, Donzelli, 2010; U. Allegretti *Democrazia partecipativa*, in *Enciclopedia del diritto*, Milano, Giuffrè.

Al riguardo, non meno importante diventa, in questa fase “evolutiva” della democrazia, quanto previsto al quarto paragrafo dell’articolo 11, ove viene definito il quadro di azione dell’iniziativa dei cittadini europei (*European Citizens’ Initiative* – ECI) e, rinviando a quanto previsto dall’attuale art. 24, par. 1 del Trattato sul Funzionamento dell’UE (TFUE), arriva a disciplinare “*le procedure e le condizioni necessarie*” per rendere possibile a livello comunitario l’iniziativa democratica.

Il rapporto tra democrazia partecipativa e democrazia digitale sembra dunque trovare un punto fondante nel sempre più diffuso accesso alla rete internet. Tutto ciò è riscontrabile anche in una consultazione pubblica lanciata dalla Commissione europea a seguito dell’azione del *Green Paper on a European Citizens’ Initiative* [COM (2009) 622 final], con l’intenzione di “*costruire un reale spazio pubblico europeo*” al fine di addivenire ad “*nuova dimensione alla democrazia europea attraverso l’uso di nuove tecnologie*”.

Infatti, le risposte di circa 160 singoli cittadini, 65 organizzazioni registrate e quasi 40 autorità istituzionali hanno quasi tutte rimarcato la necessaria di aprirsi alle nuove tecnologie come nuova e possibile previsione di raccolta delle firme e delle opinioni espresse, mostrando così il bisogno di adeguare, tramite un’azione concreta e organizzata del Web 2.0, gli strumenti della partecipazione democratica in modo così positivamente ad attivare nuove forme di relazioni sociali in un’epoca storica che ha visto l’affermarsi di una società dell’informazione grazie anche dell’avvento dei c.d. “nativi digitali”<sup>13</sup>.

Posto ciò, i c.d. “partiti digitali”<sup>14</sup> hanno cercato negli ultimi anni, anche se con risultati ancora da verificare, di dare attuazione a forme di *e-Democracy* attraverso non solo il mero utilizzo di piattaforme web partecipative, caratterizzate dal trasferimento del fulcro della loro attività dalla cabina elettorale al *personal computer*, ma anche tramite il ricorso ad un uso bivalente della piattaforma digitale intesa come luogo di espressione della volontà dei cittadini e nel contempo quale spazio comune a tutti i potenziali iscritti in grado di facilitare la pianificazione e la diffusione di materiale informativo

In Italia, il voto di tipo elettronico, seppur nella sua piena ricerca di quella “genuinità”<sup>15</sup> voluta dalla Costituzione, trova riscontro negli ultimi anni anche in alcune leggi regionali finalizzate al raggiungimento di un unico obiettivo, ovvero la mera semplificazione delle operazioni di voto e di scrutinio. Infatti, tali soluzioni legislative non introducono grandi novità rispetto all’impiego riconosciuto alla scheda cartacea visto che la legge regionale friulana del 2007<sup>16</sup>, quanto quella lombarda del 2015<sup>17</sup>, conducono esplicitamente l’adozione del voto elettronico al fine “di *accelerare e semplificare le operazioni di voto e di scrutinio*”, nonché a quello di “*garantire una maggiore trasparenza*”.

---

13 Per un approfondimento sul tema dei nativi digitali, si rimanda a P. Ferri, *Nativi digitali*, Bruno Mondadori, Milano, 2012.

14 Sul punto si veda P. Gerbaudo, *I partiti digitali. L’organizzazione politica nell’era delle piattaforme*, Il Mulino, Bologna, 2020.

15 E. Bettinelli, La lunga marcia del voto elettronico in Italia, in *Quaderni dell’Osservatorio elettorale*, 2002.

16 Legge regionale del Friuli-Venezia Giulia, n. 18 del 27 luglio 2007.

17 Legge regionale della Lombardia, n. 3 del 23 febbraio 2015.

Il tema dell'*e-Democracy*<sup>18</sup> è pertanto uno delle questioni più rilevanti e delicate quando si valuta l'introduzione di un sistema di voto elettronico. Infatti, se da un lato quest'ultimo verrebbe a incidere in modo evidente sul tasso di partecipazione, producendo una trasformazione della nozione stessa di "affluenza alle urne", dall'altro, determinando instaurazione di un regime democratico "immediato", rischierebbe di introdurre nel sistema pericolosi elementi di plebiscitarismo in ragione della perdita dei requisiti di segretezza e di personalità del voto (sanciti dall'art. 48 della Costituzione), dell'attribuzione di un carattere individualistico ed isolato al profilo del cittadino-elettore, nonché della mancanza di un luogo fisico in cui possa manifestarsi l'appartenenza a una comunità politica<sup>19</sup>.

### **3 LE PIATTAFORME DIGITALI: OPPORTUNITÀ E CARATTERISTICHE DELLE NUOVE FORME DI PARTECIPAZIONE DEMOCRATICA**

Con l'avvento delle nuove tecnologie, l'idea alla base dei partiti digitali è la definizione di un nuovo modello di organizzazione della politica e l'introduzione di meccanismi democratici finalizzati a realizzare una partecipazione politica più autentica, così da rivoluzionare tanto le modalità di reclutamento dei propri membri che i processi con cui vengono definite le scelte politiche relative alla gestione della cosa pubblica.

In tale ottica, le piattaforme digitali per la partecipazione democratica<sup>20</sup> hanno iniziato a giocare un ruolo chiave e sono sempre più considerate uno spazio utile a esprimere il proprio parere su un determinato tema di discussione, influenzando così direttamente la linea politico-programmatica di un partito. In altri termini, esse costituiscono un mezzo per neutralizzare le criticità logistiche, temporali, sociali e culturali che solitamente impediscono ai cittadini di partecipare direttamente alle scelte attuate dai propri rappresentanti di partito.

Si sono aperte così nuove e interessanti opportunità per garantire nuove forme di partecipazione democratica, anche se sempre più spesso vengono utilizzati, per il confronto politico, siti web amatoriali, piattaforme di blogging fruibili gratuitamente, ma non in grado di garantire la piena neutralità e trasparenza perché oggetto di possibili manipolazioni.

---

18 In tal senso, in Italia, le primissime proposte volte a introdurre elementi di informatizzazione e di parziale automazione nell'organizzazione elettorale risalgono alla IX Legislatura, sotto l'impulso innovatore dei deputati Adolfo Battaglia e Carlo Fusaro. Precisamente furono formulate le «Norme per l'introduzione dell'elettronica e dell'informatica nelle operazioni di voto», A.C. doc. n. 1895, 11 luglio 1984 e «Modifiche al testo unico delle leggi della Camera dei deputati, ai fini dell'introduzione dello scrutinio elettronico nelle operazioni elettorali», A.C. doc. n. 1896, 11 luglio 1984. Per un approfondimento si rimanda a A. Agosta, F. Lanchester, A. Spreafico, *Elezioni e Automazione. Tutela della regolarità del voto e prospettive di innovazione tecnologica*, Milano, Franco Angeli, 1989.

19 Sul tema si veda: A. Gratteri, *Finalità e problemi del voto elettronico: le prospettive della sua attuazione in Italia*, *Forum di Giustizia Costituzionale*, 2015; G. Sartori, *Democrazia e definizioni*, IV ed., Bologna, Il Mulino, 1976; A.C. Freschi, *E-democracy e politiche per la partecipazione dei cittadini*, in *"Economia e politica industriale"*, n. 121, 2004; F. Marzano, *Siamo maturi per l'eParticipation?*, in *"ASTRID - Rassegna"*, n. 35, 2006.

20 Attualmente le piattaforme di democrazia diretta più conosciute sono "Rousseau" per il Movimento 5 Stelle in Italia e "Participa" del partito spagnolo Podemos.



Al contrario, questi nuovi strumenti dovrebbero essere in grado di garantire l'esercizio autentico del diritto di voto, essendo quest'ultimo, com'è noto, assicurato solo in un contesto di legalità e legittimità *sostanziale* caratterizzato dall'eguale opportunità tra i candidati, dalla mancanza di manipolazione, precedente o successiva alle votazioni, in grado di falsare la volontà popolare e dalla trasparenza nelle operazioni di voto e di scrutinio<sup>21</sup>.

Alla luce di ciò, provando ad immaginare piattaforme digitali che rappresentino per la politica una dimensione partecipativa, ma che allo stesso tempo siano in grado di contemperare la valorizzazione del potenziale dell'innovazione tecnologica, esse dovrebbero avere come caratteristiche tecnico-sociali, la capacità di assicurare l'*accoglienza*, consentendo l'accesso a chi ne faccia richiesta e così dare voce a tutti i cittadini, senza alcuna forma di discriminazione.

In tale ottica, secondo un approccio proattivo, sarebbe possibile addirittura ipotizzare delle possibili strategie per diminuire – quasi fine ad azzerare – molte possibili forme di marginalità sociale, culturale, geografica o fisica. In particolare i non-vedenti, gli affetti da paralisi o da altro impedimento che adesso possono esprimere la loro volontà di elettori nelle forme del voto assistito (art. 55 DPR n. 361/1957), potrebbero invece votare in autonomia, superando lo svilimento tanto della segretezza quanto della personalità del voto. Oltre a ciò, l'uso di piattaforme elettroniche e di strumenti informatici *ad hoc* potrebbe agevolmente consentire anche a chi è fisicamente impedito di esprimere in piena autonomia il proprio voto.

Ovviamente, parlando di "accoglienza" come sinonimo di accesso al voto, un altro importante aspetto su cui soffermarsi è la partecipazione al voto degli elettori fuori sede o all'estero. In Italia il procedimento elettorale vede un rigido e solido vincolo tra luogo di residenza dell'elettore e la sezione elettorale presente nel proprio Comune e per gli elettori fuori sede o all'estero (come, tra gli altri, gli studenti universitari che si trovano per lunghi periodi lontano dal loro luogo di residenza) risulta complesso, se non oggettivamente impossibile, poter votare, mentre attraverso l'uso delle nuove tecnologie si vedrebbero riconosciuta un'effettiva "mobilità" in grado di consentire loro l'esercizio del voto in qualsiasi sezione elettorale posta sul territorio nazionale o in appositi uffici consiliari all'estero.

Le piattaforme, come la maggior parte di *repository web*, costituiscono uno straordinario deposito di informazione e di conoscenza in grado di assicurare, contro ogni rischio di broglio, la *trasparenza* del processo elettorale. La manipolazione del voto, infatti, rappresenta proprio uno dei punti di maggior vulnerabilità dei sistemi di votazione tradizionali, visto che al momento dello scrutinio le schede possono essere annullate con dei segni aggiuntivi oppure ritenute nulle discrezionalmente dallo scrutatore o anche arbitrariamente compilate se lasciate in bianco. Con il voto elettronico, invece, grazie all'eliminazione l'intermediazione umana per la lettura delle preferenze e il conteggio dei voti espressi, si sopprimerebbe alla radice la possibilità che il singolo voto possa essere manipolato o alterato, mentre la centralizzazione e l'automatizzazione degli scrutini permetterebbe anche l'immediatezza della comunicazione dei risultati effettivi.

---

21 Oltre a ciò, valgono ovviamente i requisiti di eguaglianza, personalità, libertà e segretezza stabiliti dall'art. 48 Cost. per assicurare una maggiore efficienza, economicità e trasparenza della partecipazione democratica.

Inoltre, le piattaforme devono essere in grado di garantire la sicurezza dei dati generati degli utenti e, allo stesso tempo, fornire strumenti di partecipazione che tutelino la loro riservatezza. Aspetti, quelli appena richiamati fondamentali dal momento che devono comunque coniugarsi con la trasparenza dei processi di partecipazione democratica. In tale contesto sia la *sicurezza* che la *riservatezza* sono concetti chiave in mancanza dei quali sarebbe possibile attribuire nessuna responsabilità e alcun diritto. Infatti, i problemi legati al voto elettronico e le ragioni per cui è stato abbandonato come progetto in molti Paesi europei, risiedono nella criticità e nell'estrema delicatezza tecnica connesse alla verificabilità dei risultati, alla contestuale tutela della segretezza e alla complessiva sicurezza del sistema di voto.

Analizzando le criticità richiamate non si può che rilevare che il cuore del problema è proprio la sicurezza informatica del procedimento sia con riguardo al *software* che l'*hardware* utilizzati, in relazione a possibili malfunzionamenti o manipolazioni dovute da possibili attacco di *cyber* per realizzare brogli. La necessità della stampa di un talloncino da rilasciare all'interessato affinché sia possibile verificare il voto espresso e provvedere alla contestazione qualora nonostante la sua volontà sia stato "recepito" dal sistema una preferenza diversa determinerebbe la lesione del diritto alla riservatezza e segretezza del voto.

Da sottolineare che tutti i Paesi europei che hanno deciso di abbandonare la prospettiva dell'introduzione del voto elettronico sono stati spinti da duplici ragioni, quali i problemi di sicurezza informatica riscontrati nella fase di certificazione degli apparati di voto o in seguito ad errori o episodi di malfunzionamento degli stessi e la ragionevole certezza della effettiva segretezza del voto espresso.

Alla luce di ciò, la sicurezza va garantita in due direzioni: la *sicurezza privata*, volte a mettere il cittadino nella condizione di esprimere e far valere liberamente la sua genuina volontà politica al riparo da coercizioni, controlli ed altre interferenze esterne, nonché la *sicurezza pubblica* che attiene alla resilienza dei sistemi, a eventuali attacchi informatici, manomissioni occulte e accessi abusivi ai sistemi. Ciò, tuttavia, può essere agevolmente assicurato con l'implementazione di appositi standard tecnici e con l'adozione di apposite misure organizzative prevenendo una c.d. *netiquette* (norme di buon comportamento) sulla gestione e sull'impiego di queste piattaforme da parte dei soggetti utilizzatori, nonché disciplinando anche le modalità relative all'installazione del sistema e alla verifica della regolarità delle operazioni di voto, la formazione del personale addetto, le modalità di accesso ai sistemi, le modalità di conservazione dei supporti elettronici di memoria e ogni ulteriore accorgimento tecnico informatico volto a garantire la segretezza, la genuinità e la neutralità del sistema di voto.

Infine, le piattaforme legate all'*e-democracy* devono consentire anche il dialogo democratico, far proprio il concetto di *centralità dell'utente* caratterizzato, in tale circostanza, da due diversi profili, ovvero per un verso il diritto dell'utente a esercitare il controllo sulle proprie informazioni e sulle attività che vuole svolgere, per l'altro la necessità di modellare il sistema intorno alla figura dell'utente, e quindi alle sue esigenze culturali e sociali alla luce del progresso tecnologico.

## 4 E-DEMOCRACY, ESEMPI DI APPLICAZIONI PRATICHE IN EUROPA

Come già richiamato nei paragrafi precedenti grazie alla diffusione delle nuove tecnologie molti paesi europei (ma anche fuori dall'Europa) sono stati portati a riflettere e, in alcuni casi, ad applicare modalità di voto tecnologicamente più avanzate, già a partire dai primi anni '90<sup>22</sup>.

A questo punto, è importante segnalare la distinzione tra *e-voting* (*electronic voting*) e *i-voting* (*internet voting*) con il primo sistema in grado di consentire il voto sempre presso il seggio, ma attraverso delle macchine digitali, e con il secondo rivolto invece a permettere la partecipazione al processo elettorale da remoto attraverso internet.

In Europa il Paese che può esser considerato all'avanguardia dal punto di vista del voto elettronico è certamente l'Estonia, anche grazie alla completa informatizzazione dei servizi per i cittadini e un'attenta legge elettorale per il Parlamento (il *Riigikogu*) adottata nel 2002. Si deve segnalare come il primo utilizzo del voto elettronico in Estonia sia avvenuto con le elezioni amministrative del 2005, mentre la piena legittimazione del sistema di *e-voting* si è avuta nelle elezioni politiche del 2007.

Un ulteriore esempio di consolidata applicazione del voto elettronico è la Svizzera. Infatti, proprio grazie alla frequente partecipazione alle urne dei cittadini elvetici in ragione dell'impostazione semi-diretta<sup>23</sup> della democrazia svizzera, l'introduzione dell'*e-voting*, avvenuta con la riforma della legge federale del 21 giugno 2002, entrata in vigore il 1° gennaio 2003, è risultata particolarmente agevole.

Tra le altre esperienze europee di *e-Democracy*, rientra anche quella del Belgio, Paese nel quale la democrazia elettronica è utilizzata per le consultazioni elettorali fin dal 1991 sulla base di un sistema di voto di tipo presidiato caratterizzato dall'utilizzo di una carta magnetica anonima per l'espressione della preferenza di voto.

Tuttavia, entrambi i due Paesi hanno affrontato dei problemi di sicurezza, visto che in Svizzera nel 2019 è stata scoperta una falla crittografica del *software* utilizzato per il voto, mentre in Belgio durante le Europee del 2014 si sono manifestati gli effetti di un *bug* presente nel *software* utilizzato sul sistema di voto elettronico.

In questa breve rassegna delle esperienze europee, merita attenzione anche l'utilizzo del voto elettronico in Francia. In particolare, in questo Paese si sono affermate negli anni due tipologie di *e-voting*: l'impiego di macchine elettroniche il giorno stesso delle elezioni (già dal 2002) e il voto *online* concesso ai francesi residenti all'estero in occasione delle elezioni parlamentari del 2012. Si deve tuttavia ricordare che nel 2007, dopo una pronuncia del *Conseil Constitutionnel* il voto elettronico conobbe una battuta di arresto. La Corte ritenne infatti che il voto elettronico, oltre a garantire i requisiti di affidabilità tecnica e di riservatezza, deve anche essere accettato psicologicamente dagli elettori<sup>24</sup>. Nel 2017, l'espe-

---

22 L'introduzione del voto elettronico ha visto molti passi in avanti in democrazie molto ampie come Messico, Brasile, Canada, India, Stati Uniti e democrazie di più modeste dimensioni come Estonia, Svizzera, Spagna, Francia, Belgio, Germania, Paesi Bassi.

23 P. Macchia, La democrazia semidiretta: il caso della Svizzera, in *Diritto pubblico comparato ed europeo*, 2014, fasc. 4, pp. 1654-1675.

24 La Corte costituzionale in merito ha dichiarato: "Leur utilisation, qui rompt le lien symbolique entre le citoyen et l'acte électoral que la pratique manuelle du vote et du dépouillement avait noué, se heurte aussi à une résistance psychologique qu'il convient de prendre en compte".

rienza fu poi definitivamente interrotta quando emersero forti dubbi di sicurezza dichiarati dall’Agenzia Nazionale per la *Cybersecurity*.

In Europa però non mancano anche applicazioni dell’*i-voting*<sup>25</sup>. Al riguardo, l’esempio più lampante è quello della Spagna, in cui, a causa della pandemia da *covid-19*, il 25 marzo 2020, il *Congreso de los Diputados*, fungendo da apripista in Europa nell’utilizzo della tecnologia per le sedute parlamentari, è ricorso per la prima volta in via generale alla votazione elettronica a distanza.

Con riferimento invece all’esperienza italiana, si deve rilevare come nel nostro Paese da tempo sia presente da tempo uno sforzo di elaborazione di un progetto finalizzato all’introduzione di sistemi di voto elettronico, anche se con specifico riferimento a quello reso in luoghi presidiati. Questa scelta è influenzata dal fatto che la riforma, trattandosi infatti di una modalità di consultazione che prevede l’impiego di strumenti digitali al posto della normale scheda cartacea e dell’urna elettorale, non necessiterebbe di riforme normative di dubbia costituzionalità.

Questa iniziativa si inserisce nell’ambito di un più vasto progetto europeo di ricerca e sviluppo, che ha preso avvio nel settembre 2000, denominato E-Poll (*Electronic Poll system for remote voting*), al quale partecipa il Ministero dell’interno, che ha come obiettivo di “stimolare la partecipazione dei cittadini europei alle elezioni mettendo a disposizione un sistema di voto facilitato grazie al quale i votanti non sono più obbligati a recarsi in uno specifico luogo per esercitare il proprio diritto, ma possono farlo da una qualunque postazione di voto elettronico presente sul territorio”<sup>26</sup>.

In particolare, in Italia, le prime prove sono state compiute a San Benedetto del Tronto, nel 2000, ad Avellino nel 2001 (referendum costituzionale), a Campobasso (amministrative di maggio 2002) e a Cremona (simulazione su un campione di 1.000 elettori), mentre in Friuli si è svolta un’ampia sperimentazione di voto elettronico in occasione delle elezioni regionali del giugno 2003 che ha coinvolto 3.800 elettori iscritti in quattro sezioni elettorali poste in altrettanti Comuni.

Da sottolineare come negli ultimi anni la sperimentazione abbia coinvolto anche la fase precedente al voto, ovvero quella relativa al riconoscimento dell’elettore: infatti, in occasione delle elezioni amministrative svoltesi a Campobasso l’elettore è stato identificato utilizzando la carta d’identità elettronica, mentre per l’espressione del voto, è stato adottato il sistema del *touch screen* in cui con il semplice “tocco” del dito su uno dei simboli visualizzati sullo schermo – ove è riprodotta una scheda elettorale “virtuale” corrispondente a quella reale – è stato possibile per l’elettore attribuire il proprio voto alla lista o al candidato prescelto.

---

25 Un importante applicazione del voto elettronico via internet è rappresentata anche dalle elezioni municipali in Ontario (Canada) che già dal 2003 avvengono in tale modalità. Per un approfondimento si veda M. Schirripa, Il ritorno dell’e-voting alla luce della pandemia da Covid-19. Una riflessione sull’esperienza canadese, in *Federalismi.it*, 22 luglio 2020.

26 <http://documenti.camera.it/Leg14/dossier/Testi/MLC110.htm>

## 5 CONCLUSIONI

Appare possibile constatare come le nuove tecnologie siano molto caldegiate per favorire la partecipazione democratica dei cittadini in ragione della loro carica innovativa e della familiarità che riescono a trasmettere, quasi empaticamente. Tuttavia risulta altrettanto evidente il rischio, teorizzato da Rodotà, di addivenire a nuova “*democrazia di massa*” destinata ad assumere i “*caratteri di una living room democracy, la c.d. “democrazia del tinello”*”<sup>27</sup>, in cui la intermediazione del digitale si trasforma in una polarizzata semplificazione strumentale del sì e del no, oscurando scelte che pure sono realisticamente proponibili.

In realtà, il voto elettronico rappresenta indubbiamente un tema dalle molteplici sfaccettature e ogni Paese che conosce la sua sperimentazione finisce per dar vita a un’esperienza assolutamente unica, destinata ad essere attentamente verificata e confrontata con le altre al fine di offrire una panoramica completa dei punti di vantaggio e di svantaggio evidenziatisi.

A ciò si deve aggiungere che recentissimi strumenti tecnologici come la *blockchain* inducono a pensare – al momento solo sulla carta non potendo far affidamento su casi applicativi concreti - come sia possibile prospettare per il futuro il ricorso a soluzioni tecnico-giuridiche in grado di garantire, anche per il voto elettronico, i fondamentali principi costituzionali dell’anonimato, della segretezza, della libertà, dell’autenticità e della sicurezza<sup>28</sup>.

## BIBLIOGRAFIA

1. Agosta, Antonio, Fulco Lanchester, Alberto Spreafico, A. Agosta, G. Ciaurro, S. Costa, M. D’Antonio, V. Di Ciolo, S. Furlani, e C. Fusaro. *Elezioni e automazione: tutela della regolarità del voto e prospettive di innovazione tecnologica*. F. Angeli, 1989.
2. Allegretti, U. «Democrazia partecipativa». *Enciclopedia del diritto*. s.d.
3. Allegri, Giuseppe. «Il diritto di iniziativa dei cittadini europei (ECI) e la democrazia partecipativa nell’unione europea: prime riflessioni». *federalismo.it* 23 (2010).
4. Bettinelli, E. «La lunga marcia del voto elettronico in Italia». *Quaderni dell’osservatorio elettorale* 46 (2002).
5. Carlotto, P. «il voto», 2018.
6. Castells, Manuel. *The rise of the network society*. Vol. 12. Oxford: Blackwell Publishing, 1996.
7. Costanzo, Pasquale. «La democrazia elettronica (note minime sulla cd. e-democracy)». *Dir. inf. infor*, 2003.
8. Crisafulli, V. «La sovranità popolare nella Costituzione italiana (Note preliminari)». In *Stato, popolo, governo: Illusioni e delusioni costituzionali*. Milano: Giuffrè, 1985.
9. Cuniberti, M. «Tecnologie digitali e libertà politiche», 2015.
10. Elia, Silvia. «Alcune questioni problematiche circa il voto elettronico». In *Cyberspazio e diritto*, 2016.

---

27 S. Rodotà, Tecnopolitica. *La democrazia e le nuove tecnologie della comunicazione*, Laterza, Roma-Bari, 2004, p.37.

28 G. Gometz, M. T. Folarin, Voto elettronico presidiato e blockchain, in *Ragion pratica*, il Mulino, Fascicolo 2 – 2018.

11. Ferri, P. *La rivoluzione digitale. Comunità, individuo e testo nell'era di Internet*. Eteropie, 2001.
12. Freschi, Anna Carola. «E-democracy e politiche per la partecipazione dei cittadini». *Economia e politica industriale*, n. 121 (2004).
13. Gerbaudo, P. *I partiti digitali. L'organizzazione politica nell'era delle piattaforme*. Bologna: Il Mulino, 2020.
14. Gometz, Gianmarco. *Democrazia elettronica. Teoria e tecniche*. Pisa: Edizioni ETS, 2017.
15. Gometz, Gianmarco, e Monica Tawa Folarin. «Voto elettronico presidiato e blockchain». *Ragion pratica*, n. 2 (2018): 317–328.
16. Gratteri, A. «Finalità e problemi del voto elettronico: le prospettive della sua attuazione in Italia». *Forum di Giustizia Costituzionale*, 2015.
17. Macchia, Patrizia. «La democrazia semidiretta: il caso della Svizzera», 2014.
18. Marzano, F. «Siamo maturi per l'eParticipation?» *ASTRID - Rassegna*, n. 35 (2006).
19. Mazzoli, Graziella. *Profili sociali della comunicazione e nuove tecnologie. Mondi vitali e mondi artificiali: dalla polarizzazione all'interfaccia?* Vol. 8. FrancoAngeli, 1996.
20. Orofino, A. G. «Il voto elettronico». In *Diritto delle nuove tecnologie informatiche e dell'internet*, a cura di G. Cassano. Milano, 2002.
21. Rodotà, Stefano. *Iperdemocrazia: Come cambia la sovranità democratica con il web*. Gius. Laterza & Figli Spa, 2013.
22. ———. *Tecnopolitica. La democrazia e le nuove tecnologie della comunicazione*. Roma-Bari: Laterza, 2004.
23. Sartori, Giovanni. *Democrazia cosa è*. Bur, 2012.
24. Urbinati, Nadia. *Democrazia in diretta: le nuove sfide alla rappresentanza*. Feltrinelli Editore, 2013.
25. Zuboff, Shoshana. *Il capitalismo della sorveglianza. Il futuro dell'umanità nell'era dei nuovi poteri*. Roma: Luiss University Press, 2019.



# Reflecting on case-law databases and publicity of judgments in the light of the ECHR

Gianluca Borgia<sup>1</sup>, Benedetta Galgani<sup>2</sup>

<sup>1</sup> Università di Pisa – gianluca.borgia [at] jus.unipi.it

<sup>2</sup> Università di Pisa - benedetta.galgani [at] unipi.it

## ABSTRACT

The research aims to examine the impact of private and public case-law databases and algorithms on which the related consultation software are based on the publicity of judgments in the framework of the European Convention of Human Rights.

## KEYWORDS

case-law database; publicity of judgments; public trial; fair trial; AI; algorithms.

## TALK

Public and private case-law databases are an increasingly indispensable tool also to judges: the AI huge computational capabilities have made it possible to carry out complex research in even shorter timeframe with a growing accuracy rate.

However, these instruments hide many pitfalls which can affect the courts in the exercise of their judicial functions. For instance, there is the danger that under the weight of such a huge amount of data the judge abandons himself to a sort of “decisional conformism”. Even if this evokes the value of consistency of case-law, as held by the European Court of Human Rights, the case-law development is essential since a failure to maintain a dynamic and evolutive approach would risk hindering reform or improvement<sup>1</sup>.

The issue of the right to access to the databases and to the algorithms on which the related consultation software are based has only recently been addressed and solely under the point of view of the rights of defence, without considering the position of citizens.

The greater attention on the aspect concerning the rights of defence derives from the fact that the accused person and the prosecutor in criminal proceedings or the parties in civil trials are the most interested in knowing and understanding all the elements that have contributed to the grounds for the decision, in view of the effective and efficient exercise of the right of appeal.

---

<sup>1</sup> See ECtHR, *Lupeni Greek Catholic Parish and Others v. Romania*, 29 November 2016, application no. 76943/11, § 116

<http://hudoc.echr.coe.int/eng?i=001-169469>



Nevertheless, this does not justify an attitude of complete disregard to the second question. Article 6 § 1 of the European Convention on Human Rights (“Right to a fair trial”) states that «[j]udgment shall be pronounced publicly».

This is, therefore, a fundamental principle, which protects citizens from a “secret justice” that escapes the control of the collectivity. Moreover, it contributes to preserving confidence in courts<sup>2</sup>.

However, it is clear that in order to achieve these results, it’s necessary that not only the decision taken must be made public, but also that citizens should be able to comprehend the judge’s reasoning. In other words, to ensure the effectiveness of the guarantee, it is not sufficient only the publicity of the final result: what is also indispensable is get to know the factors that influenced such a result.

This implies that also individuals inevitably have a legitimate interest in accessing the same databases used by the court as well as in knowing how the related consultation software work.

From a normative point of view, the above-mentioned Article 6 § 1 of the European Convention on Human Rights provides for exceptions at the publicity of the “trial” exclusively based on the necessity to safeguard morals, public order, national security, interests of juvenile or the private life of the parties.

Nonetheless, the free access to databases and to algorithms on which the related consultation software are based has to deal in practice with the economic interest of the service provider, particularly when it is a private body.

The provision of entire algorithms or the underlying software code to the public – as has been said by the Council of Europe’s Committee of Experts on Internet Intermediaries (MSI-NET) study on “Algorithms and Human Rights” – «is an unlikely solution in this context, as private companies regard their algorithm as key proprietary software that is protected»<sup>3</sup>.

This does not mean, however, that is not possible to identify a compromise solution. Notably, the “European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment” drafted by the European Commission for the Efficiency of Justice (CEPEJ) in 2018 states that, in order to ensure the principle of transparency, the system «could be explained in clear and familiar language (to describe how results are produced) by communicating, for example, the nature of the services offered, the tools that have been developed, performance and the risks of error»<sup>4</sup>.

As it is not possible to completely exclude any control, the Charter also affirms that «Independent authorities or experts could be tasked with certifying and auditing processing methods or providing advice beforehand. Public authorities could grant certification, to be regularly reviewed»<sup>5</sup>.

---

2 See ECtHR, *Serre v. France*, 29 September 1999, application no. 29718/96, § 21: <http://hudoc.echr.coe.int/eng?i=001-62936>; ECtHR, *Gautrin and Others v. France*, 20 May 1998, application no. 38/1997/822/1025-1028, § 42

<http://hudoc.echr.coe.int/eng?i=001-62726>

3 MSI-NET, *Algorithms and Human Rights*. Study on the human rights dimensions of automated data processing techniques (in particular algorithms) and possible regulatory implications, Council of Europe, 2018, p. 38

<https://rm.coe.int/algorithms-and-human-rights-en-rev/16807956b5>

4 CEPEJ, *European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment*, Council of Europe, 2019, p. 11

<https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c>

5 *Ibidem*.

The aforesaid remarks on the existence of an economic interest in limiting access to the case-law databases cannot apply to the public service provider.

From this perspective, the solution adopted by the Council of Europe with regard to the decisions of the European Court of Human Rights is to be welcomed. And in fact, the HUDOC database, which is the official case-law database of the Strasbourg Court has always been open source.

The implementation of the “ItalgireWeb”, which is a database developed by the “Centro Elettronico di Documentazione” of the Italian Supreme Court of Cassation that has always been available to the judiciary, instead, has been very different. In fact, it was only last February that the Italian Minister of Justice concluded an agreement which allows not the individuals, but solely lawyers to access such a database freely and at no cost<sup>6</sup>.

In conclusion, the hope, in the light of the above-mentioned considerations, is that, with regard to the private databases, the solution proposed by the CEPEJ will be accepted and, with respect to the public ones, the model developed in the context of the Council of Europe will become the starting point, but not the milestone.

To ensure the effectiveness of the publicity principle is in fact necessary that even the terms used for the single research as well as the algorithms behind the databases’ consultation software are publicly available.

Only in this way can the individuals and *a fortiori* the defence reach a sufficient degree of understanding of the decision.

Even if it can appear far from the current reality, it would be essential in any case that the judge’s reasoning shows the ways through which the research in the database has been conducted.

## REFERENCES

1. Buffa, F. «Le risorse informatiche per la conoscenza della giurisprudenza della Corte e la banca dati HUDOC». *Questione giustizia – Speciale “La Corte di Strasburgo”*, 2019, 136–38. <https://www.questionegiustizia.it/speciale/2019-1>.
2. Di Cerbo. «Banche dati di giurisprudenza, nomofilachia e trasparenza dell’attività giurisdizionale. L’esperienza del Ced della Corte di cassazione». *Questione giustizia* 3, 2017. [https://www.questionegiustizia.it/data/rivista/articoli/447/qg\\_2017-3\\_14.pdf](https://www.questionegiustizia.it/data/rivista/articoli/447/qg_2017-3_14.pdf).
3. Di Chiara, G. «“ Against the administration of justice in secret”: la pubblicità delle procedure giudiziarie tra Corte europea e assetti del sistema italiano», 2008.
4. Fiandese, F. «Il Centro Elettronico di Documentazione della Corte Suprema di Cassazione». In *L’informatica giuridica in Italia. Cinquant’anni di studi, ricerche ed esperienze*, 151–68. Napoli: ESI, 2014. <http://www.ittig.cnr.it/EditoriaServizi/AttivitaEditoriale/CollanaSeD/sed-12/Fiandese.pdf>.
5. Gaito, A., e S. Furfaro. «Consensi e dissensi sul ruolo e sulla funzione della pubblicità delle udienze penali». *Giurisprudenza costituzionale* 2 (2010): 1065–76.

---

6 See the information on ItalgireWeb data-base

<http://www.italgiure.giustizia.it/>

6. Galgani, B. «Considerazioni sui “precedenti” dell’imputato e del giudice al cospetto dell’IA nel processo penale». *Sistema penale*, 2020.
7. ———. «Giudizio penale, habeas data e garanzie fondamentali». *Archivio penale* 1, 2019.
8. Gambino, A.M. «Informatica giuridica e diritto dell’informatica». In *Enciclopedia Treccani – Diritto on line*, 2013.
9. Harris, D.J., M. O’Boyle, E.P. Bates, e M. Buckley. «Law of the European Convention on Human Rights». Oxford University Press, 2018.
10. Hildebrandt, Marielle. «Profile transparency by design?: Re-enabling double contingency». *Privacy, due process and the computational turn: The philosophy of law meets the philosophy of technology*, 2013, 221–46.
11. Peluso, F. *L’informatica per il giurista*. A cura di G. Cassano. Santarcangelo di Romagna: Maggioli, 2019.
12. Vincenti, E. «Massimazione e conoscenza della giurisprudenza nell’era digitale», 2018. [https://www.questionegiustizia.it/rivista/articolo/massimazione-e-conoscenza-della-giurisprudenza-nell-era-digitale\\_588.php](https://www.questionegiustizia.it/rivista/articolo/massimazione-e-conoscenza-della-giurisprudenza-nell-era-digitale_588.php).

**Sessione XII**  
Henri de Saint-Simon



# Tell us what you think: home and destination attachment for migrants on Twitter

Jisu Kim<sup>1</sup>, Alina Sîrbu<sup>2</sup>, Giulio Rossetti<sup>3</sup>, Fosca Giannotti<sup>3</sup>, Hillel Rapoport<sup>4</sup>

<sup>1</sup> Scuola Normale Superiore di Pisa - jisu.kim [at] sns.it

<sup>2</sup> Università di Pisa - alina.sirbu [at] unipi.it

<sup>3</sup> ISTI-CNR - fosca.giannotti, giulio.rossetti [at] isti.cnr.it

<sup>4</sup> Université Paris 1 Panthéon-Sorbonne, CEPPII – hillel.rapoport [at] psemail.eu

## ABSTRACT

Integration of migrants touches upon multidimensional aspects of a host country. Successful integration of a migrant could bring new opportunities, and hence an overall improvement of their living conditions and well-being. Integration can be described using two dimensions: the preservation of links to the home country and culture, i.e., home attachment, and creation of new links and adoption of cultural traits from the new residence country, i.e., destination attachment. In this talk we introduce a means to quantify these two aspects based on Twitter data. Our new indicators reflect well the behaviour of migrants across different countries. The home and destination attachment indices are compared with various elements such as language proximity, distance between countries and also with Hofstede's cultural dimension scores. The results highlight the link between host-country language proficiency and destination attachment. Moreover, we observed that destination attachment levels correlate to cultural aspects of the destination country, while home attachment levels seem to be more related to the country of origin.

## KEYWORDS

international migration, integration, assimilation, segregation, Shannon entropy, big data, Twitter

## TALK

### 1 INTRODUCTION

Integration of migrants touches upon multidimensional aspects of the host country. Successful integration of a migrant could bring new opportunities, and hence an overall improvement of their living conditions and well-being. Failure to integrate migrants in the host country's society may result in social conflict, creation of ghetto or/and possible economic losses. The studies of cultural integration have been mainly done by sociologists, by employing survey data such as World Values Survey, Eurobarometer, and European Social Survey. The main elements used in the studies are often inter-marriage, religion and language on cultural aspects [1, 5, 6, 7].

## 2 METHODOLOGY AND RESULTS

Over the years, several integration theories have been introduced. While a complete agreement of what integration means and how it should be measured does not exist, there have been efforts to identify different types of integration patterns. The patterns can be described using two dimensions: the preservation of links to the home country and culture, i.e. home attachment, and creation of new links and adoption of cultural traits from the new residence country, i.e. destination attachment. In this paper, we introduce a methodology to quantify these two aspects based on Twitter data. We first identify international migrants using a technique developed by [4]. Subsequently, we compute home and destination attachment based on topics discussed on Twitter. Our new indicators reflect well the behaviour of migrants across different countries. For instance, figure 1 displays the home and destination attachment indices at a country level for Italian emigrants across different countries of residence. In general, we observe that Italians are more attached (on average 0.21) to home than attached to their destination country (on average 0.073). However, Italians tend to integrate well in English speaking countries as the destination attachment levels are at highest in the United States, Australia, and the United Kingdom. Although the destination attachment level is not as high as in English speaking countries, they also tend to do relatively well in neighbouring countries like Spain, France, Belgium and the United Kingdom. Spain, for instance, also has a high destination attachment level due to the language similarity between the two countries. In contrast, Italian emigrants residing in countries such as China, Hungary, and Thailand that are culturally far from Italy have the lowest destination attachment index levels. To understand the link to language proficiency, we computed box plots for two groups in figure 2; a group that often speaks the language of the host country and a group that very rarely speaks the language of the host country. Here, we are looking at all the migrants we have in the dataset regardless of the country of origin or the country of residence. We observe that the group that speaks the language has a higher destination attachment level. For this group, the destination attachment level on average is 0.1 and 0.09 for the home attachment level. On the other hand, the group that does not speak the language show higher home attachment level. For this group, the home attachment level on average is 0.14 and 0.04 for the destination attachment level. Our indexes confirm the link between language and destination attachment of immigrants in the host country. In addition, it shows the relationship between the language and attachment to the home country, which is not evident in the literature. We also compared the home and destination attachment indices with Hofstede’s cultural dimension scores [2, 3]. We observed that destination attachment levels correlate to cultural aspects of the destination country, while home attachment levels seem to be more related to the country of origin.

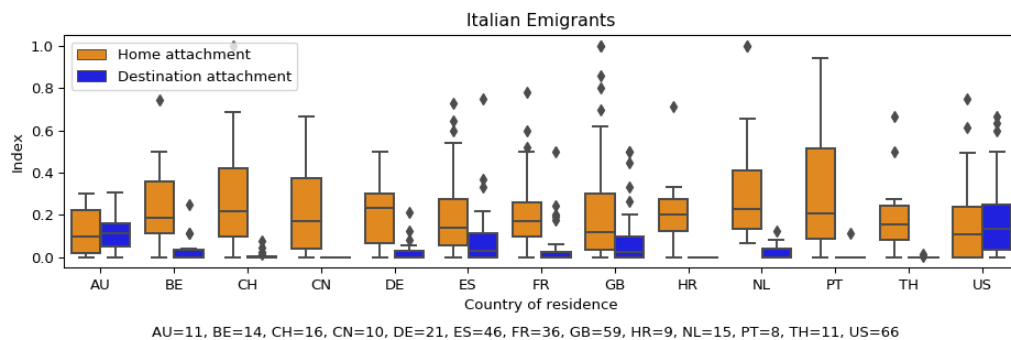


Figure 1: box plots for home and destination attachment indices grouped by destination countries for Italian nationals living abroad. The x-axis displays countries of residence of Italians. The numeric labels correspond to the number of Italian nationals in each group.

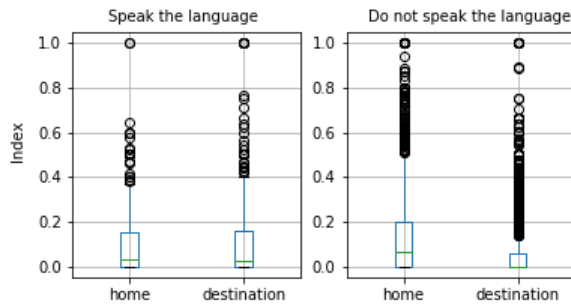


Figure 2: box plots of home and destination attachment indices for the subgroup of migrants who speak the language of the host country on the left and the subgroup of migrants who do not speak the language of the host country on the right.

## REFERENCES

1. Esser, Hartmut. *Migration, Language and Integration*. Berlin: WZB, 2006.
2. Hofstede, Geert. *Culture's Consequences: International Differences in Work-Related Values*. Vol. 5. Sage, 1984.
3. ———. 'Dimensionalizing Cultures: The Hofstede Model in Context'. *Online Readings in Psychology and Culture* 2, no. 1 (2011): 2307–0919.
4. Kim, Jisu, Alina Sîrbu, Fosca Giannotti, and Lorenzo Gabrielli. 'Digital Footprints of International Migration on Twitter'. In *International Symposium on Intelligent Data Analysis*, 274–286. Cham: Springer, 2020.
5. Lochmann, Alexia, Hillel Rapoport, and Biagio Speciale. 'The Effect of Language Training on Immigrants' Economic Integration: Empirical Evidence from France'. *European Economic Review* 113 (2019): 265–296.
6. Sîrbu, Alina, Gennady Andrienko, Natalia Andrienko, Chiara Boldrini, Marco Conti, Fosca Giannotti, Riccardo Guidotti, Simone Bertoli, Jisu Kim, and Cristina Ioana Muntean. 'Human Migration: The Big Data Perspective'. *International Journal of Data Science and Analytics*, 2020, 1–20.
7. Vigdor, Jacob L. *Measuring Immigrant Assimilation in the United States*. Civic Report No. 53., 2008.





# Famiglie, minori e cyberbullismo: la rete familiare e sociale alla prova

Caterina Murgò

Università degli Studi di Pisa, Italia, caterina.murgo [at] unipi.it

## PAROLE CHIAVE

famiglia, scuola, cyberbullismo

## KEYWORDS

family, school, cyberbullying

## INTERVENTO

### 1 OBIETTIVI

Il contributo si propone di verificare l'efficacia delle misure d'intervento volte a contrastare il fenomeno del bullismo in generale, con speciale riferimento alla sua versione *cyber* o informatica.

Si vuole, in particolare, procedere all'esame delle migliori sinergie tra famiglia e scuola in grado di prevenire, se possibile, la formazione del contesto ambientale utile allo sviluppo delle condotte dei bulli.

L'elaborato prende avvio dalla normativa interna, che di recente ha assistito all'intervento del legislatore sul tema e che consente d'inquadrare il bullismo come vera e propria piaga sociale; spunti di rilievo sono offerti anche dalle corti, chiamate di frequente a intervenire sulle modalità di esercizio della responsabilità genitoriale e a ridefinire i confini degli obblighi spettanti ai genitori verso i figli, specialmente se minori d'età, proprio alla luce dell'evoluzione dei mezzi di comunicazione.

L'obiettivo è anche quello d'indagare le misure mediante le quali l'ordinamento garantisce l'esplicazione della personalità dei minori e disciplina gli strumenti finalizzati alla progressiva autonomia delle persone minori, vietando al contempo che nella sfera privata di tali soggetti si producano intrusioni illegittime, incrementate dalle moderne tecnologie informatiche.

## 2 STATO DELL'ARTE

Il bullismo e il cyberbullismo, quali condotte vessatorie e mortificanti *off* e *on line*, si caratterizzano per vicende che coinvolgono non soltanto l'ambiente scolastico, ma in misura primaria il contesto familiare e sociale nel quale i minori vivono e operano: la conferma si trae non soltanto da recenti interventi normativi<sup>1</sup>, finalizzati ad arginare il fenomeno e all'adozione di misure di prevenzione, ma anche da importanti provvedimenti della Corte di Cassazione che, decidendo sui fatti sottoposti all'attenzione, hanno operato un'attenta ricostruzione dei comportamenti dei bulli, soffermandosi sulla formazione familiare di provenienza e inquadrando il bullismo come fenomeno sociale che interroga la collettività<sup>2</sup>.

L'area del diritto della famiglia è stata interessata di recente da interventi normativi rilevanti, che hanno consentito di ridefinire la sfera delle relazioni familiari, unificando la posizione dei figli all'interno della famiglia e rendendola autonoma dallo stato giuridico dei genitori.

Al tempo stesso, la riforma della filiazione, intervenuta con la l. n. 219/12, ha configurato in capo agli ascendenti un complesso di obblighi che ruotano intorno ai figli, specie se minori d'età, riformulando quelle obbligazioni in veri e propri diritti dei figli, di cui all'art. 315 *bis* c.c.<sup>3</sup>

Dall'esame della legislazione speciale sulla posizione dei minori nell'ordinamento possono evidenziarsi taluni aspetti: per un verso, i soggetti minori d'età sono destinatari di posizioni di diritto, quali il diritto all'ascolto, e appaiono sulla scena giuridica in qualità di titolari di posizioni attivabili in via autonoma.

Tra gli interventi di maggior rilievo può ricordarsi il d. lgs. n. 101/18, che ha consentito l'adeguamento della normativa interna al Regolamento europeo 2016/679 sulla protezione delle persone fisiche rispetto al trattamento dei dati personali (GDPR): il provvedimento, introducendo nel Codice della *privacy*, d. lgs. 196/03, l'art. 2 *quinqüies*, ha previsto la facoltà per i minori di prestare direttamente il consenso per l'acquisto di beni e servizi della società dell'informazione al compimento del quattordicesimo anno d'età, una soglia che ha trovato spazio anche nell'ambito della normativa sul cyberbullismo, l. n. 71/17<sup>4</sup>.

Ancora, i minori risultano certamente titolari di posizioni meritevoli di tutela: l'interesse del minore, quale clausola generale del sistema giuridico, rappresenta il controlimite necessario nell'adozione di tutte le decisioni che lo riguardano<sup>5</sup>.

---

1 Bocchini-Montanari, Le nuove disposizioni a tutela dei minori per la prevenzione ed il contrasto del fenomeno del cyberbullismo, in *Nuove leggi civ. comm.* 2018, 340 ss.

2 Cass., 10 settembre 2019, n. 22541, massima in *Resp. civ. prev.* 2020, 505 ss.

3 Ambrosini, Dalla "potestà" alla "responsabilità": la rinnovata valenza dell'impegno genitoriale, in *Dir. fam. per.* 2015, 687 ss.

4 A. Astone, L'accesso dei minori d'età ai servizi della c.d. Società dell'informazione: l'art. 8 del Reg. (UE) 2016/679 e i suoi riflessi sul Codice per la protezione dei dati personali, in *Contr. impr.* 2019, 614 ss.

5 Di Landro, Best interest of the child e tutela dei minori nel dialogo tra legislazione e giurisprudenza, in *Nuove leggi civ. comm.* 2020, 460 ss.

D'altra parte, l'ordinamento individua nei soggetti minori d'età una condizione di sostanziale fragilità o di vera e propria debolezza, che li renderebbe bisognosi di protezione e di tutela: è quanto emerge dalle recenti modifiche al codice civile (art. 315 *bis* c.c., sul diritto all'assistenza morale in favore dei figli), ma anche dalla disciplina sulla responsabilità per l'illecito dei figli minori, che chiama gli ascendenti a rispondere per il fatto dannoso della prole in età minore, salva la concorrente responsabilità dei figli in caso di capacità d'intendere e di volere, *ex art.* 2048 c.c.

### **3 METODOLOGIA ADOTTATA**

Il fenomeno del bullismo in tutte le sue manifestazioni ha spesso lambito la disciplina giuridica della responsabilità civile per il fatto illecito del terzo, sia in occasione di vicende gravissime e irreparabili<sup>6</sup> sia in situazioni ove le corti giudicanti hanno evidenziato la complessità delle responsabilità educative dei genitori con riguardo alle condotte pregiudizievoli dei figli minori e le criticità legate ai doveri di vigilanza e di controllo da parte delle strutture scolastiche, ove spesso le vicende persecutorie e dannose hanno trovato svolgimento.

Sul versante della responsabilità degli ascendenti per il fatto illecito della prole minore d'età, gli organi giudiziari hanno spesso evidenziato un sostanziale appiattimento delle funzioni genitoriali sul dovere di educare, considerato preminente tra gli obblighi genitoriali e necessario nel prevenire la formazione dei bulli.

A ciò si aggiunga la misura in cui le corti ritengono essenziale, a carico dei genitori, l'esercizio di un dovere di vigilanza sull'assimilazione da parte dei figli delle misure educative e sul corretto inserimento dei figli nel contesto socio-ambientale.

Con riguardo al ruolo delle autorità scolastiche, le stesse sono chiamate a un generale dovere di protezione dell'integrità psico-fisica degli studenti, in particolare se minori, per tutto il tempo in cui essi sono affidati alla responsabilità della scuola e fino a quando gli allievi non siano presi in carico da altri soggetti responsabili; è quanto evidenziato anche dalla legge sul cyberbullismo, l. n. 71/17, che ha posto in capo alle strutture scolastiche l'attivazione di misure volte a prevenire e contrastare il fenomeno in tutte le sue manifestazioni.

L'indagine si svolgerà mediante un approfondimento delle funzioni della responsabilità, in quanto effettivamente rispondenti al perseguimento degli interessi tutelati: l'interesse dei genitori a un'adeguata impostazione della relazione con i figli e con i terzi estranei al nucleo familiare; l'interesse della vittima della condotta dei cyberbulli alla prevenzione degli atti dannosi e alla riparazione dei danni; l'interesse delle strutture scolastiche e delle famiglie a una corretta esecuzione dei doveri di protezione verso gli allievi minori d'età.

### **4 SOLUZIONI**

Quali soluzioni possono formularsi?

Deve ritenersi certamente positivo il tentativo in corso della legislazione speciale di ridefinire l'inquadramento giuridico del minore, scardinando la cristallizzazione della sua posizione operata dal codice civile e qualificando il minore come soggetto effettivo dell'ordinamento.

---

6 Trib. Sulmona, 9 aprile 2018, in *Fam. dir.* 2018, 763 ss.

Nella lotta al bullismo, quale fenomeno che coinvolge profili diversi, da quello familiare a quello risarcitorio, come emerge dai lavori citati nella bibliografia, essenziali sono le misure d'intervento in funzione preventiva e riparatoria: 1) i contributi del contesto familiare e scolastico nella misura in cui siano antecedenti alle condotte prevaricatrici, ma anche successivi, consentendo la riparazione del pregiudizio patito; 2) la conoscenza delle modalità di comunicazione tra giovanissimi, come emerso anche dalle previsioni della normativa sul cyberbullismo (art. 4 l. n. 71/17).

## **BIBLIOGRAFIA**

Busnelli. «La famiglia e l'arcipelago familiare», 2002.

Cassano. «La responsabilità genitoriale nell'uso dell'odierna tecnologia telematica», 2020.

Cassano-Marvasi. «La responsabilità educativa dei genitori per minori cyberbulli», 2018.

Mantovani. «Commento sub art. 2048 c.c.» Commentario al codice civile, 2011.

Manzella. «Are States winning the fight? Evidence of the impact of state law on bullying in schools». *Economics of education Review*, n. 61 (2018): 264.

Messinetti. «Diritti della famiglia e identità della persona», 2005.

Ricci. «Sulla “funzione sociale” del diritto alla protezione dei dati personali,» 2017.

Sonelli. «L'interesse superiore del minore. Ulteriori “tessere” per la ricostruzione di una nozione poliedrica,» 2018.

Thiene. «Ragazzi perduti on line: illeciti dei minori e responsabilità dei genitori», 2018.

# Digital Commons as new Infrastructure Towards an Industrial Policy for the Digital Age

Marco Berlinguer  
URGOCIS – IGOP – UAB (Spain) - marco.berlinguer [at] gmail.com

## ABSTRACT

Free and Open Source Software (FOSS) has come to largely dominate software production. This means that the main technology and industry of the digital revolution has integrated digital commons among its main institutional regulations.

However, to understand the successful trajectory of FOSS, it is necessary to revisit the initial approaches to the production logic behind FOSS and this new generation of commons.

This is particularly necessary with regard to public policy, which have so far failed to participate productively in these new productive arrangements.

A review of the notion of infrastructure and an analysis of the principles emerging in the architecture of the latest generation of digital infrastructures provide a perspective from which to identify new possible approaches both to the construction and governance of these shared goods and resources and to a renewed and extended scope of public policies.

Moreover, software trajectory provides a meaningful model for the entire information paradigm. Its relevance is high and immediate, potentially, for many other sectors, including cultural production and creation.

## KEYWORDS

free and open source software (FOSS), digital commons, infrastructure, public policy, innovation, standards, modularity

## TALK

Free and Open-source software (FOSS) has come to largely dominate software production. This means that the main technology and industry of the digital revolution has adopted the commons [2; 28] among its main institutional regulations. And with that, that digital commons represent an organizational and economic innovation that has much to say about the organizational models of production in the digital age.

Despite the importance of this evolution, the phenomenon is still little known, and even among scholars, its understanding is in the making [5].

Commons, for example, are often celebrated because they are open access resources and potentially democratize productive and cultural environments. But the open-source has not prevented the formation of new forms of concentration of power and wealth. In its current form, it could even encourage its expansion. This is visible by observing the leading role that the giants of the web (Google, Facebook, Amazon, and ultimately Microsoft itself)<sup>1</sup> - the same ones that monopolize much of the digital economy - have had and are having in the growth of open-source [4; 5].

Another necessary adjustment concerns the relationship between commons and markets. Common goods and markets have idiosyncratic institutional forms: if a good is open access, it cannot be appropriated exclusively, nor can it be sold. However, if open source has been absorbed by markets and capitalist competition, it is because commons and markets can not only coexist but can grow in synergy. Successful inclusion of a commons in a production ecosystem eliminates the market in its domain but can create, reconfigure or grow adjacent and complementary markets [6].

For this reason, digital ecosystems typically operate on a multi-layered scale, exploiting the existence of “multi-sided markets”, but they have also increasingly learned to operate through different regimes of value creation and appropriation, adopting and/or orchestrating in some levels or layers, regimes of shared value and collaborative production [25; 21; 6].

Practically, all the most innovative and successful companies of the digital age have learned to use these hybrid strategies. The case of Google-Android is a clear example. Google's condemnation for abuse of dominant position obtained through the open-source operating system Android also indicates how the Antitrust Authority must revisit its doctrine [18].

The ability to move in production environments made of multi-level integrated systems operating through different regimes of ownership, governance and value generation and appropriation [17; 15], is, in any case, one of the important lessons that we can draw from the trajectory of open source and its absorption in the market.

This innovation also indicates a possible entry point for an approach to public intervention in digital production environments, radically different from the previous industrial-Fordist paradigm.

In fact, the formation of the above sketched productive ecosystems outlines the contours of a new type of mixed economy and can provide many insights for an industrial policy more apt to the digital age.

So far, public policies have lagged behind in their ability to participate virtuously in these new environments. There has been no lack of attempts since there have been hundreds of public administrations that have announced and implemented policies to adopt or promote FOSS [22]. But, somehow, it is fair to say that the first versions of public policies applied to FOSS have not worked [4].

Looking at the evolution of the FOSS ecosystem, a new approach can be glimpsed.

It goes through a re-examination of the notion of infrastructure in the information society and a new understanding about the extended role of this type of shared goods and resources in digital production [11; 13; 3; 12, 19; 9; 24].

A further look at the latest generation of digital infrastructures (Cloud computing, Internet of Things, Artificial Intelligence, etc.) also highlights an innovative matrix, which is increasingly shaping their architecture and design, that is organized around three principles: FOSS, standardization and modularity [26; 1; 7; 8].

---

1 A similar approach is spreading more recently among China's top tech companies, like Huawei, Tencent, Alibaba.

This institutional and architectural matrix responds to unprecedented rhythms of evolution, complexity and interdependence, and aims to combine the advantages of specialization, stabilization and scalability with the need to maintain maximum flexibility to allow agile experimentation, innovation, adoption and integration, but also greater resilience to disruptive events and changes. These same design principles, however, are emerging along with a new hierarchy of layers, as more stable blocks are being created that constitute new common infrastructures (e.g. platforms) that largely establish the technological trajectories and economic regimes that can be built on them [10; 14; 9; 6].

The increasing adoption of FOSS, standardization and modularity is giving rise to a new understanding of the very "fabric" of innovation [3; 16]. But it also outlines new possible approaches to the construction and governance of this new generation of infrastructures and to a renewed and extended scope of public sector action, both regulatory and productive, to foster, monitor, orchestrate, exploit a new generation of commons or global and trans-national public goods, co-generated and co-managed with a multiplicity of other agents and stakeholders, and according to a new productive logic. This requires new approaches however, with particular attention to the combination of open-source, modularity and standardization in the design of the architecture of these infrastructures and to the orchestration of the different value creation logics and models of governance implied by these new production environments.

Furthermore, there are good reasons to consider this new production logic emerged in the software and this trajectory as a model or "exemplar" [20] for the informational techno-economic paradigm [23]. Its relevance is high and immediate, potentially, for many other areas, starting from the increasingly central and controversial issues related to data governance, property and economic exploitation. But the same cultural production and creation, which are undergoing a radical transformation induced by the digital revolution, have much to learn from a better understanding of the trajectory of software.

## REFERENCES

1. Baldwin, Carliss Young, Kim B. Clark, and Kim B. Clark. *Design Rules: The Power of Modularity*. Vol. 1. MIT press, 2000.
2. Benkler, Yochai. 'Coase's Penguin, or, Linux and "The Nature of the Firm"'. *Yale Law Journal*, 2002, 369–446.
3. ———. 'Commons and Growth: The Essential Role of Open Commons in Market Economies'. *The University of Chicago Law Review* 80, no. 3 (2013).
4. Berlinguer, Marco. 'Commons, Markets and Public Policy', 2020.
5. ———. 'New Commons: Towards a Necessary Reappraisal'. *Popular Communication* 18, no. 3 (2020): 201–215.
6. ———. 'The Value of Sharing. How Commons Have Become Part of Informational Capitalism and What We Can Learn from It. The Case of FOSS'. *Rassegna Italiana Di Sociologia* 59, no. 2 (2018): 263–288.
7. Blind, Knut. 'The Impact of Standardisation and Standards on Innovation'. In *Handbook of Innovation Policy Impact*. Edward Elgar Publishing, 2016.



8. Blind, Knut, Mirko Böhm, and Nikolaus Thumm. 'The Relationship Between Open Source Software and Standard Setting', 2019.
9. Constantinides, Panos, Ola Henfridsson, and Geoffrey G. Parker. *Introduction—Platforms and Infrastructures in the Digital Age*. INFORMS, 2018.
10. Dosi, Giovanni. 'Technological Paradigms and Technological Trajectories: A Suggested Interpretation of the Determinants and Directions of Technical Change'. *Research Policy* 11, no. 3 (1982): 147–162.
11. Edwards, Paul N. 'Infrastructure and Modernity: Force, Time, and Social Organization in the History of Sociotechnical Systems'. *Modernity and Technology* 1 (2003): 185–226.
12. Eghbal, Nadia. 'Roads and Bridges'. *The Unseen Labor behind Our Digital Infrastructure*, 2016.
13. Frischmann, Brett M. *Infrastructure: The Social Value of Shared Resources*. Oxford University Press, 2012.
14. Gawer, Annabelle, and Michael A. Cusumano. 'Industry Platforms and Ecosystem Innovation'. *Journal of Product Innovation Management* 31, no. 3 (2014): 417–433.
15. Geels, Frank W. 'Technological Transitions as Evolutionary Reconfiguration Processes: A Multi-Level Perspective and a Case-Study'. *Research Policy* 31, no. 8–9 (2002): 1257–1274.
16. Hawkins, Richard, Knut Blind, and Robert Page. *Handbook of Innovation and Standards*. Edward Elgar Publishing, 2017.
17. Jessop, Bob. 'Regulationist and Autopoieticist Reflections on Polanyi's Account of Market Economies and the Market Society'. *New Political Economy* 6, no. 2 (2001): 213–232.
18. Katz, Michael, and Jonathan Sallet. 'Multisided Platforms and Antitrust Enforcement'. *Yale LJ* 127 (2017): 2142.
19. Kawalek, Peter, and Ali Bayat. 'Data as Infrastructure', 2017.
20. Kuhn, and Thomas S. 'The Structure of Scientific Revolutions'. *Journal of Chemical Education*, 2012. <https://doi.org/https://doi.org/10.1021/ed041pA206>.
21. Lepak, David P., Ken G. Smith, and M. Susan Taylor. 'Value Creation and Value Capture: A Multilevel Perspective'. *Academy of Management Review* 32, no. 1 (2007): 180–194.
22. Lewis, James A. 'Government Open Source Policies'. *Center for Strategic and International Studies*, 2010.
23. Perez, Carlota. 'Technological Revolutions and Techno-Economic Paradigms'. *Cambridge Journal of Economics* 34, no. 1 (2010): 185–202.
24. Rahman, K. Sabeel. 'Infrastructural Regulation and the New Utilities'. *Yale J. on Reg.* 35 (2018): 911.
25. Rochet, Jean-Charles, and Jean Tirole. 'Platform Competition in Two-Sided Markets'. *Journal of the European Economic Association* 1, no. 4 (2003): 990–1029.
26. Shapiro, Carl, Shapiro Carl, and Hal R. Varian. *Information Rules: A Strategic Guide to the Network Economy*. Harvard Business Press, 1998.
27. Tonn, Bruce. *The Nature of Technology: What It Is and How It Evolves*, WB Arthur, Free Press, New York (2009), 246 Pp. Pergamon, 2010.
28. Weber, Steve. *The Success of Open Source*. Harvard University Press, 2004.

# Contro l'odio online

Giorgio Mennella<sup>1</sup>, Giulia Pirozzi<sup>2</sup>  
1 Ciaopeople – mennella.gio [at] gmail.com  
2 FutureVox – prz.giulia [at] gmail.com

## SINTESI

Il digitale ha radicalmente rivoluzionato il modo di intendere, produrre e distribuire informazioni. Siamo passati da un sostanziale oligopolio produttivo e distributivo ad un mondo nel quale chiunque può diventare un polo informativo. I social network hanno contribuito all'aumento sia di lettori sia del contenuto prodotto, come mai visto nella storia dell'umanità. Purtroppo, questa rivoluzione strutturale non ha avuto un corrispettivo in termini educativi. Fake news ed hate speech sono due facce della stessa medaglia, figlie di un ecosistema informativo ad altissimo potenziale ma che ha ancora una lunga strada da fare tra algoritmi, dati e vuoti normativi. Per contrastare le moderne forme di discriminazione online ed interrompere la catena dell'odio, è necessario che avvenga una rivoluzione che metta al centro le parole e porti allo sviluppo costante e reciproco di tre grandi dimensioni: normativa, digitale e culturale. Da un punto di vista giuridico, riconoscere ed utilizzare una terminologia condivisa in materia di hate speech rappresenta il primo passo verso la definizione di metodologie di contrasto trasversali e multi-livello, permettendo di avviare politiche di contrasto sia a livello locale sia interstatale in un'ottica di prevenzione di un'escalation di violenze. In termini di ecosistema strutturale digitale è necessario porre l'accento sul ruolo delle piattaforme sociali e distributive. Ben consci del dilemma che verrebbe a crearsi in merito alle libertà di espressione e di opinione, l'attuale de-responsabilizzazione pressoché totale delle piattaforme, quale Facebook, Twitter e YouTube non permette un'azione concreta di contrasto all'odio e alla disinformazione. Infine, da un punto di vista culturale è necessario sviluppare un'educazione che porti le persone a riflettere nuovamente sull'importanza delle parole: un uso consapevole del proprio linguaggio è determinante per un uso corretto della rete, che rappresenta oggi lo strumento con il quale possiamo garantire lo sviluppo di una vita interconnessa. Se ci poniamo nei confronti della rete in modo consapevole, sviluppando un'educazione digitale che ci permetta di conoscere i meccanismi e gli algoritmi che regolano le piattaforme online, riusciremo a vivere in armonia tra mondo online e mondo offline. Solo attraverso il raggiungimento di obiettivi comuni a queste tre dimensioni sarà possibile sviluppare un efficace contrasto all'odio online.

## PAROLE CHIAVE

internet, informazione, fake news, filter bubble, odio in rete, linguaggi, social media, società civile

# INTERVENTO

## 1 INTRODUZIONE

Secondo l'ultimo Global Digital Report sulle tendenze del mondo digitale, nel 2019 Internet ha registrato un aumento del 9% degli utenti ed ha raggiunto, con un totale di 4,39 miliardi, oltre la metà della popolazione globale [5]. Questa espansione costante a partire dagli anni '60 del secolo scorso ha reso Internet lo strumento più potente del 21<sup>esimo</sup> secolo e, nel raggiungere un'ampia fetta della società civile, ha accompagnato e al contempo generato profonde trasformazioni sociali. L'avvento di Internet e dei social ha infatti portato ad un cambiamento del paradigma dell'informazione che da verticale è divenuto pressoché orizzontale, passando da una logica top-down ad una bottom-up, dove potenzialmente ogni utente è capace di produrre e distribuire contenuti. Negli ultimi 15 anni si è assistito ad una sostanziale modifica dell'ecosistema informativo mondiale, con un flusso informativo non più nelle mani dei pochi gruppi media. Se gli inizi del terzo millennio hanno visto nascere i citizen journal, con l'arrivo dei social network (Facebook e Twitter su tutti), l'informazione ha subito una rivoluzione copernicana. I social network diventano piattaforme distributive, potenzialmente universali, a disposizione di chiunque decida di produrre un contenuto: sia esso un'immagine, un testo o un video. L'informazione online in sé ed i social network come piattaforme distributive hanno portato ad un aumento esponenziale dei lettori di news. Il paragone tra il numero di copie vendute al giorno dei quotidiani [2], e il numero di utenti unici dei quotidiani online (fonte Audiweb) è esemplificativo: 6.000.000 di copie vendute al giorno nel 2000 contro i 32.000.000 di utenti unici al giorno nel 2020<sup>1</sup>.

Tav. n. 12

EVOLUZIONE TIRATURE E VENDITE GIORNALI QUOTIDIANI  
(medie giornaliere - 1990/2001)

| anno     | tirature medie | % variaz. | vendite   | % variaz. medie |
|----------|----------------|-----------|-----------|-----------------|
| 1990     | 9.763.197      | 1,2       | 6.808.501 | 0,6             |
| 1991     | 9.492.087      | -2,8      | 6.505.426 | -4,5            |
| 1992     | 9.429.250      | -0,7      | 6.525.529 | 0,3             |
| 1993     | 9.245.797      | -1,9      | 6.358.997 | -2,6            |
| 1994     | 9.030.007      | -2,3      | 6.208.188 | -2,4            |
| 1995     | 8.599.394      | -4,8      | 5.976.847 | -3,7            |
| 1996     | 8.503.177      | -1,1      | 5.881.350 | -1,6            |
| 1997     | 8.143.897      | -4,2      | 5.869.602 | -0,2            |
| 1998     | 8.156.405      | 0,2       | 5.881.421 | 0,2             |
| 1999     | 8.204.477      | 0,6       | 5.913.514 | 0,5             |
| 2000     | 8.469.856      | 3,2       | 6.073.158 | 2,7             |
| 2001 (*) | 8.639.253      | 2,0       | 6.098.058 | +0,4            |

(\*) Proiezione in base ai dati forniti da 49 testate quotidiane.

Fig. 1 - Tirature medie e vendite dei quotidiani cartacei dal 1990 al 2001 secondo FIEG

1 Sono disponibili online i dati della total digital audience del mese di luglio 2020 (Audiweb).  
- <http://www.audiweb.it/news/comunicati-stampa/total-digital-audience-Luglio-2020.html>

| INTERNET AUDIENCE<br>(browser + app)  |                                   |                     |                      |                          |
|---|-----------------------------------|---------------------|----------------------|--------------------------|
| Fonte: Audiweb Database. DATI GIORNO MEDIO DEL MESE DI LUGLIO 2020 - Audiweb powered by Nielsen |                                   |                     |                      |                          |
|   | TOTAL DIGITAL AUDIENCE<br>2+ anni | COMPUTER<br>2+ anni | TABLET<br>18-74 anni | SMARTPHONE<br>18-74 anni |
| Utenti unici giorno medio (.000)  | 32.409                            | 9.233               | 4.134                | 28.214                   |
| Pop. di riferimento giorno medio (%)  | 54,3                              | 15,5                | 9,2                  | 62,7                     |
| Tempo per persona giorno medio (hh:mm:ss)   | 2:30:14.000                       | 1:19:12             | 1:24:11.000          | 2:14:19.000              |
| Utenti unici mese (.000)  | 42.062                            | 25.010              | 7.292                | 37.610                   |
| Pop. di riferimento mese (%)  | 70,5                              | 41,9                | 16,2                 | 83,5                     |
| Tempo per persona mese (hh:mm:ss)   | 2:11:48:33.000                    | 15:06:23.000        | 1:03:39:16.000       | 2:40:34:000              |

Fig. 2 - Utenti unici digitali nel mese di Luglio 2020 secondo Audiweb Database

Figura 2 – Utenti unici digitali nel mese di luglio 2020 secondo Audiweb Database

A tale aumento di lettori digitali non è però corrisposto un incremento della digital literacy. L'assenza di un'adeguata educazione all'utilizzo dei nuovi strumenti e delle nuove piattaforme di informazione ha portato ad una serie di distorsioni, dall'incapacità di verifica delle fonti che porta alla diffusione di fake news, all'acuirsi di fenomeni di polarizzazione informativa già presenti nella società [4]. Quelle che comunemente vengono definite filter bubble sono diventate delle sacche di utenti sempre più polarizzati ed integralisti, poco permeabili a fonti informative non validate dalla loro ristretta cerchia. Gli algoritmi distributivi delle piattaforme social, favorendo l'esposizione ad un dato utente di contenuti che hanno ottenuto maggiori interazioni dalla rete sociale dell'utente stesso, hanno concorso all'isolamento di queste sacche. Tali sacche si riducono ad effettive casse di risonanza, o echo chambers, che spingono l'utente verso un'ulteriore polarizzazione.

Tale utente esposto a contenuti sempre più simili e sempre più condivisi dalla propria cerchia di contatti social, tenderà ad auto-convincersi di un dato argomento e ad avere la percezione che ci sia ancora più odio in rete, ad osservare il mondo attraverso un filtro basato sul compiacimento delle sue inclinazioni e, a livello di linguaggio, a portare gli hater a diventare ancora più hater, poiché radicalizzati all'interno del sistema polarizzato delle echo chamber [3].

## 2 IL DISCORSO D'ODIO ONLINE

I discorsi d'odio, che rappresentano una peculiarità e una criticità della realtà contemporanea, sono fortemente legati con il concetto della discriminazione all'interno di una cornice giuridica ancora troppo frammentata nell'ambito della normativa sull'hate speech, incapace di adottare una linea comune per la difesa del principio della libertà di espressione e della tutela della dignità della persona. Molti discorsi d'odio trovano infatti la propria giustificazione nella libertà di manifestazione del pensiero in un processo che ha portato alla legittimazione di discorsi offensivi e discriminatori. Parole come "negro" e "zingaro" sono tornate nel lessico comune attraverso la divulgazione di linguaggi ed espressioni utilizzate in particolar modo da figure pubbliche, come leader o personalità politiche [1]. Tali discorsi vengono facilmente ripresi negli articoli dei quotidiani e delle testate giornalistiche all'interno delle piattaforme online, dove i discorsi d'odio possono trovare la propria cassa di risonanza, nelle interazioni e nelle conversazioni degli utenti.

La mancanza di filtri porta le persone a riversare impulsivamente in rete sensazioni comuni della natura umana, come frustrazione, indignazione, rabbia o sconforto. Quando queste reazioni si sviluppano a margine degli articoli o dei fatti di cronaca dei quotidiani e delle testate giornalistiche online, l'uso improprio di certi termini, la lettura parziale dell'articolo o l'incapacità di discernere un'informazione reale da una fake news portano le persone a scrivere commenti e ad interagire con toni non pacati all'interno delle conversazioni online. Attivando questo processo, l'hate speech può acquisire un ruolo protagonista nelle piattaforme online, poiché i post che tendono a generare il maggior numero di interazioni sono proprio quelli caratterizzati da opinioni contrastanti e conversazioni infuocate, dando la percezione a molti utenti che il web sia sempre più colmo di odio. All'interno delle pagine online, infatti, vengono a generarsi le cosiddette echo chamber, dove le persone con opinioni simili si scambiano contenuti ed idee che si confermano a vicenda, facendosi eco l'un l'altra [3]. La caratteristica delle echo chambers è la presenza di un'unica voce, l'assenza del dissenso e l'allontanamento degli utenti con idee diverse da quelle del gruppo di maggioranza, portando così alla polarizzazione del discorso e all'appiattimento del dialogo.

Vengono quindi a crearsi delle situazioni di immobilità mentale, ma anche comunicativa, che portano a dare ascolto soltanto a quei contenuti che assecondano i propri interessi e le proprie aspettative, rifiutando ogni tipo di alternativa alla propria visione. La legittimazione di certi discorsi ha portato, nel presente come nel passato, alla giustificazione di atti di violenza nei confronti delle persone maggiormente discriminate, trasformando l'odio online in violenza offline. All'interno di questa cornice, in che modo è possibile strutturare ed attivare metodologie di prevenzione alla diffusione dell'odio, in un'ottica di lungo termine? che modo è possibile servirsi della tecnologia per attivare metodologie di contrasto all'hate speech? Ma soprattutto, qual è il ruolo delle persone, in questo scenario? Per rispondere a queste domande è prima di tutto necessario individuare convergenze tra i diversi metodi di contrasto all'odio, per arrivare alla definizione di linee guida e di parametri condivisi da adottare nelle diverse tipologie di intervento.

Inoltre, per il raggiungimento di un obiettivo comune (il contrasto e la prevenzione dei discorsi d'odio online) è fondamentale cominciare da una base di partenza condivisa: prima di tutto, bisogna ripartire dalle parole. Stare sui social network non è tanto una questione di acquisire o esercitare particolari capacità tecniche, non si esaurisce nell'uso di un computer o di un dispositivo: è essenzialmente comunicare e al centro della comunicazione ci sono proprio le parole. Anche in una società sovraccarica di immagini, schermi e contenuti che stimolano e coinvolgono i cinque sensi nel loro complesso, sono ancora le parole e sempre loro a dare significato a tutto.

### **3 METODOLOGIA**

Ripartire dalle parole significa definire e condividere i significati e i valori dei termini che si utilizzano. Da un punto di vista giuridico, riconoscere ed utilizzare una terminologia condivisa in materia di hate speech rappresenta il primo passo verso la definizione di metodologie di contrasto trasversali e a multi-livello. È proprio attraverso la definizione univoca del termine hate speech, infatti, che sarà possibile adottare ed avviare politiche di contrasto comuni a più Stati, rendendo efficace una battaglia ai discorsi d'odio a livello internazionale. All'interno di una cornice giuridica più chiara e definita, le persone possono sentirsi maggiormente tutelate in merito alle forme di discriminazione, senza che siano private della libertà di espressione e di pensiero.

Una normativa trasparente che disciplini i discorsi d'odio porta alla definizione dei linguaggi offensivi che sono tornati velocemente in uso nel lessico comune, in un'ottica di prevenzione di un'escalation di violenze, online e offline. Allo stesso modo, una maggiore responsabilizzazione delle piattaforme social, per quanto sensibile e pericolosa sia la tematica, diventa di vitale importanza. Sebbene demandare il fact-checking alle piattaforme, come avvenuto durante le elezioni americane del 2020 in merito alle affermazioni del Presidente Donald Trump<sup>2</sup>, può risultare controverso, un maggior intervento di Facebook e Twitter nei discorsi d'odio, attraverso strumenti di verifica automatica, è fortemente auspicabile. Parallelamente ad un'attività di informazione ed educazione civica circa le tematiche dei discorsi d'odio, è necessario sviluppare un'educazione al linguaggio che porti le persone a riflettere nuovamente sull'importanza delle parole. La diversificazione del registro linguistico sulla base del contesto, dunque la separazione tra la sfera pubblica e privata è un primo importante accorgimento alla portata di tutti per riconsiderare la propria esposizione nella rete e prevenire la diffusione di commenti che possono provocare discorsi discriminatori.

## 4 CONCLUSIONI

Una corretta espressione può cominciare da una corretta attività di informazione. La continua ricerca e lo sviluppo di un pensiero critico, attraverso la lettura e l'individuazione delle fonti, senza limitarsi ad un'unica versione dei fatti, richiama la possibilità di riconoscere che possano esistere molteplici punti di vista.

Questo comportamento è l'antidoto per prevenire reazioni impulsive ed emotive che si possono riversare nello spazio dedicato ai commenti delle pagine, portando al cosiddetto "effetto tinello" [3], quindi la modalità di comportarsi in uno spazio pubblico (quale è il social network) come se si fosse all'interno del salotto della propria casa. Partendo da una differenziazione del mezzo è invece promuovere un dialogo civile e moderato. Un uso consapevole del proprio linguaggio è determinante per un uso corretto della rete, che rappresenta al giorno d'oggi lo strumento con il quale possiamo garantirci lo sviluppo della vita interconnessa attuale. Se ci poniamo nei confronti della rete in modo consapevole, sviluppando un'educazione digitale che ci permetta di conoscere i meccanismi e gli algoritmi che regolano le piattaforme online, riusciremo forse a vivere in armonia tra mondo online e mondo offline. Nello scenario digitale così come in quello giuridico, la definizione di norme chiare e condivise che regolino queste realtà lascia a ciascuno di noi un'ampia possibilità di movimento ed espressione. In conclusione, affinché possano essere messe in atto misure di prevenzione e di contrasto dei discorsi d'odio online, è necessario che ci sia uno sviluppo, costante e condiviso, di tutte le tre dimensioni analizzate in questo lavoro: giuridica, culturale e digitale.

---

2 [https://www.repubblica.it/esteri/2020/11/03/news/trump\\_l'estensione\\_dei\\_voti\\_postali\\_e\\_fisicamente\\_pericolosa\\_e\\_twitter\\_censura\\_il\\_presidente-272848745/](https://www.repubblica.it/esteri/2020/11/03/news/trump_l'estensione_dei_voti_postali_e_fisicamente_pericolosa_e_twitter_censura_il_presidente-272848745/)

## BIBLIOGRAFIA

1. Faloppa, Federico. *Razzisti a Parole (per tacere dei fatti)*. Bari: Laterza, 2011.
2. FIEG. *La stampa in Italia 1998-2001*. Roma: FIEG, 2001. [http://www.fieg.it/upload/studi\\_allegati/studio19982001.pdf](http://www.fieg.it/upload/studi_allegati/studio19982001.pdf).
3. Gheno, Vera, e Bruno Mastroianni. *Tienilo Acceso*. Milano: Longanesi, 2018.
4. Sustein, Cass R. *Voci, gossip e false dicerie. Come si diffondono, perché ci crediamo, come possiamo difenderci*. Milano: Feltrinelli, 2010.
5. We are social. *Global Digital Report 2019*. New York: We are social, 2019. <https://wearesocial.com/global-digital-report-2019>.

# I servizi sociali e socio-educativi in tempi di Covid-19: strategie di digitalizzazione in due programmi nazionali di contrasto alla povertà e alla vulnerabilità familiare

Andrea Petrella<sup>1</sup>, Marco Ius<sup>2</sup>, Matteo Tracchi<sup>3</sup>, Paola Milani<sup>4</sup>

<sup>1</sup> Università di Padova – andrea.petrella [at] unipd.it

<sup>2</sup> Università di Padova – marco.ius [at] unipd.it

<sup>3</sup> Università di Padova – matteo.tracchi [at] unipd.it

<sup>4</sup> Università di Padova – paola.milani [at] unipd.it

## SINTESI

Nello scenario creatosi in seguito all'emergenza Coronavirus le esigenze socio-assistenziali ed educative delle fasce più fragili della società si sono ulteriormente ampliate e diversificate, ponendo enormi sfide al sistema dei servizi e, parallelamente, alle loro infrastrutture informatiche. Si è pertanto assistito a una riorganizzazione delle modalità di offerta ed erogazione di servizi e interventi professionali (a distanza, o in presenza ma con i dovuti dispositivi di protezione) e la nascita o il consolidamento di nuove connessioni tra enti pubblici e privati, professionisti e società civile, gruppi formali e informali, al fine di garantire continuità ai percorsi di accompagnamento e di diversificare la programmazione del welfare locale. Pertanto, nel campo dei servizi socio-educativi e sociosanitari, è osservabile una riconfigurazione in prospettiva smart potenzialmente in grado di fornire risposte via via più coerenti con i bisogni della cittadinanza. Si prefigura così il consolidamento del paradigma dello smart welfare, caratterizzato dall'utilizzo delle Digital Humanities (DH), dalle Information and Communication Technologies (ICT) e da una reinterpretazione delle prassi operative sia tra operatori e famiglie che all'interno dei servizi stessi.

## PAROLE CHIAVE

smart working; smart welfare; vulnerabilità; povertà; Coronavirus; servizi sociali



# INTERVENTO

## 1 CONTESTO

Il Laboratorio di ricerca e intervento in educazione familiare (LabRIEF) dell'Università di Padova, integrando saperi e pratiche in una prospettiva interdisciplinare e in stretta connessione con i professionisti, coordina scientificamente il programma nazionale P.I.P.P.I. (Programma di Intervento Per la Prevenzione dell'Istituzionalizzazione) rivolto a nuclei familiari in situazione di vulnerabilità [19] e l'accompagnamento formativo dei *case manager* all'interno della misura di contrasto alla povertà Reddito di Cittadinanza (RdC). Entrambi i programmi sono finanziati dal Ministero del lavoro e delle politiche sociali.

P.I.P.P.I. prevede che famiglie e bambini partecipino a tutte le fasi del progetto educativo attraverso il metodo della valutazione partecipativa e trasformativa [26], concordando obiettivi e azioni e sostenendo l'attivazione di dispositivi d'intervento (servizio di educativa domiciliare, gruppi genitori, partenariato scuola-famiglia-servizi, vicinanza solidale) e di equipe multidisciplinari costituite dalle famiglie e dai professionisti di differenti servizi. Da dicembre 2017 l'esperienza maturata in P.I.P.P.I. è confluita, grazie al contributo di Università di Padova, Regioni, Ambiti territoriali e Ministero del lavoro e delle politiche sociali, nelle Linee di indirizzo nazionali "L'intervento con bambini e famiglie in situazione di vulnerabilità. Promozione della genitorialità positiva", documento che ambisce a essere lo sfondo comune per tutti i territori e i servizi che, a vario titolo, lavorano nei percorsi di accompagnamento con bambini e genitori [20]. Nell'ambito del Reddito di Cittadinanza, in caso di bisogni complessi e di assenza dal mercato del lavoro da più di due anni, i beneficiari e i loro nuclei familiari sono indirizzati ai servizi sociali del territorio dove avviano insieme ai *case manager* e, in alcuni casi, a un'equipe multidisciplinare, un patto di inclusione sociale che prevede diverse fasi (dall'analisi preliminare alla progettazione personalizzata).

Entrambi i programmi presuppongono una costante interazione non solo tra servizi e famiglie incluse nei percorsi di accompagnamento, ma anche tra professionisti. Per fronteggiare le diverse forme di negligenza e povertà educativa, economica e sociale P.I.P.P.I. e RdC lavorano, pur con strumenti leggermente differenti, rafforzando le risposte genitoriali ai bisogni di sviluppo dei bambini [18], valorizzando le risorse relazionali e innescando dinamiche di *empowerment* capaci di mettere in circolo i punti di forza di ogni famiglia [21].

Fra le possibili risposte alla situazione di emergenza che ha investito anche i servizi, il Laboratorio, attraverso l'attivazione di una "Comunità di Pratiche" *online* ha invitato gli operatori (insegnanti, assistenti sociali, educatori, psicologi) a condividere pratiche professionali per loro significative in tempi di Coronavirus. Tale "Comunità di Pratiche" è stata messa a disposizione nel portale Moodle (*Modular Object-Oriented Dynamic Learning Environment*), solitamente utilizzato per la formazione *online* e per la gestione delle implementazioni dei programmi in corso.

Inoltre, essa è stata resa accessibile anche agli operatori esterni ai programmi con la duplice finalità di raccogliere e documentare le pratiche per offrire strumenti di promozione della resilienza delle famiglie, degli operatori e dell'intero sistema dei servizi, anche attraverso l'utilizzo di strumenti informatici [7; 29; 30]. Lo scambio nella "Comunità di Pratiche" è stato pensato come una modalità per garantire una memoria del lavoro svolto e per stimolare la riflessione affinché quanto sperimentato possa trasformarsi in apprendimento durevole, accrescere i saperi professionali della comunità stessa e, conseguentemente, offrire alle fasce più vulnerabili della società servizi differenziati e rispondenti a specifiche esigenze.

## 2 OBIETTIVI

Partendo dalle trasformazioni avvenute nei servizi sociali e socio-educativi durante e dopo il periodo di *lockdown della primavera 2020*, il presente *paper* intende analizzare i limiti e le potenzialità delle nascenti e diversificate pratiche digitali nei percorsi di accompagnamento che vedono coinvolte le famiglie in situazione di vulnerabilità. Basandosi su alcune esperienze condotte nei servizi sociali italiani a partire dall'inizio delle misure restrittive per limitare i contagi da Coronavirus (marzo 2020) il *paper* esplora sia gli aspetti del lavoro sociale che è stato possibile convertire con maggiore immediatezza ed efficacia sia quelli che hanno rivelato alcuni rischi connessi a questo approccio: come si organizzano i professionisti per implementare interventi e strategie di azione con le famiglie a fronte delle misure restrittive? Come cambia il ruolo dei professionisti e dei servizi socio-educativi (nuovi spazi virtuali di incontro con cui misurarsi, nuove fragilità da affrontare, nuove interazioni da gestire mediate da strumenti telematici, ecc.)?

Infine, il *paper* si interroga su quali siano le sfide organizzative e digitali che l'emergenza Covid-19 ha reso più evidenti per i servizi alla persona. Nonostante alcuni interventi siano stati immediatamente convertiti attraverso modalità online, restano molte questioni aperte relative all'utilizzo e all'accessibilità delle tecnologie, alla possibilità di garantire adeguati percorsi di accompagnamento e alla frammentarietà delle risposte organizzative e gestionali riscontrabili nei singoli ambiti territoriali.

## 3 METODO

Attingendo dalle esperienze di P.I.P.P.I. e del Reddito di Cittadinanza, condivise dagli operatori sociali attraverso i forum *online* della "Comunità di Pratiche", le pratiche raccolte riguardano sia aspetti micro (strategie per convertire in modalità telematica interventi quali il servizio di educativa domiciliare o i colloqui con i professionisti) che aspetti macro (iniziative di riorganizzazione dei servizi, nascita di nuovi servizi, pratiche di integrazione sociosanitaria). Nei mesi di aprile e maggio 2020 sono state raccolte 26 pratiche, provenienti da altrettanti ambiti territoriali italiani di 11 regioni (Calabria, Campania, Emilia-Romagna, Lazio, Liguria, Lombardia, Provincia Autonoma di Trento, Puglia, Sicilia, Toscana, Veneto). Per ciascuna pratica è stato richiesto di descrivere il contesto, il numero di partecipanti coinvolti, di delineare in sintesi l'esperienza, includendo, in base alla tipologia della stessa e alle informazioni a disposizione, gli obiettivi, la durata, gli spazi e gli strumenti utilizzati, come è stata attuata nelle diverse fasi, e gli eventuali esiti raggiunti.

I testi raccolti, riferiti a entrambi i programmi nazionali, sono stati sistematizzati e analizzati individuando e selezionando temi ricorrenti (la continuità dei dispositivi di intervento, la gestione a distanza degli incontri di equipe, la riorganizzazione delle diverse fasi degli interventi) e mettendone in risalto criticità e potenzialità.

Il gruppo di ricerca LabRIEF sta continuando a raccogliere queste esperienze e, pur nella consapevolezza di riferirsi a pratiche e considerazioni maturate all'interno di due specifici programmi, i dati raccolti, proprio per la loro scala nazionale, permettono di rilevare alcune tendenze rintracciabili nell'intero sistema dei servizi socioeducativi rivolti a famiglie e bambini in situazione di vulnerabilità.

## 4 RISULTATI

Per i servizi sociali ed educativi rivolti alle famiglie, l'impossibilità di incontri in presenza ha rappresentato un'enorme sfida, di cui ancora si sta comprendendo la portata, tanto per i cittadini che per il sistema dei servizi stessi. I cambiamenti nelle configurazioni organizzative e nelle modalità di lavoro più micro offrono l'occasione per comprendere come il *welfare* dei bambini e delle famiglie vulnerabili abbia recepito la sfida trasformativa che la pandemia ha costretto ad affrontare. Se in alcuni servizi lo *smart working* e il ricorso – repentino, disordinato, parziale – a strumenti tecnologici e telematici è stato accolto con scetticismo e difficoltà, in altri è stato una opportunità per misurarsi con nuovi strumenti, nuove modalità di lavoro e nuovi modi di interagire con famiglie e bambini, inaugurando, forse, un *welfare* in chiave *smart* [17; 23; 24] capace non di sostituire i percorsi di accompagnamento in presenza, ma di completarli, di facilitarne alcune aspetti e assicurare anche in condizioni critiche la risposta ai bisogni educativi, di accompagnamento e di inclusione sociale delle fasce vulnerabili della nostra società.

Dai contributi raccolti emergono con chiarezza alcune indicazioni che si collocano a scale diverse all'interno del lavoro sociale e del sistema dei servizi:

- una prima indicazione riguarda le diverse “velocità digitali” che tuttora il Paese esprime;
- una seconda fa riferimento ai livelli di preparazione, aggiornamento e formazione degli operatori sociali rispetto all'utilizzo delle *Digital Humanities*;
- una terza è relativa a come gli strumenti telematici impattano sull'integrazione tra i professionisti;
- infine, la quarta indicazione riguarda le differenti strategie messe in atto dagli operatori sociali per includere le famiglie attraverso l'utilizzo della tecnologia.

1) Le diverse “velocità digitali”. Una variabile fondamentale per permettere ai servizi di rispondere ai bisogni di bambini e famiglie in situazione di vulnerabilità è la dotazione di infrastrutture digitali e strumenti informatici per convertire e facilitare il lavoro sociale. Le tuttora persistenti differenze territoriali riguardanti la connettività digitale determinano livelli disomogenei tra centri e periferie, aree urbane ed extra urbane. I territori più discriminati sono quelli delle aree cosiddette interne [5] o montane con problemi di connettività digitale [13; 14] e caratterizzati da “disuguaglianze economiche e sociali, esclusione da fondamentali servizi di ‘cittadinanza’ quali trasporti, istruzione e servizi, senza scordare le più impalpabili disuguaglianze di riconoscimento” [1]. L'attenzione è quindi da porre su quelle aree del Paese in cui a criticità di carattere economico, lavorativo, sociale si sommano difficoltà a dotare le pubbliche amministrazioni di adeguati strumenti per il lavoro agile, con ulteriori impatti negativi per genitori e bambini in situazione di vulnerabilità.

L'emergenza sanitaria e il *lockdown* hanno reso evidente una volta di più la necessità per il Paese di dotarsi di un piano capillare e solido per l'estensione della banda larga, al fine di mettere tutti i territori nelle medesime condizioni.

2) Preparazione, aggiornamento e formazione degli operatori. Inoltre, è opportuno riflettere sulla dotazione di strumenti informatici e la diffusione di competenze informatiche tanto nei servizi quanto tra le famiglie. La sospensione, tra le altre, delle attività scolastiche, educative, assistenziali e la loro parziale riconversione in modalità *online* ha fatto emergere situazioni altamente differenziate. Se per alcune famiglie connettersi a un dispositivo per permettere ai bambini di seguire lezioni, laboratori e attività extrascolastiche o per proseguire il percorso di accompagnamento intrapreso con i servizi sociali o specialistici è stata un'operazione semplice e immediata, per una buona parte ha comportato enormi difficoltà, sia di natura materiale (possessione/non possessione di computer, *smartphone*, *tablet*, rete internet) che di natura immateriale. Il *digital divide* è una criticità ancora molto presente in Italia e il rischio, a fronte di un repentino passaggio a modalità *smart* e di un massiccio ricorso ai mezzi telematici, è di aumentare le disuguaglianze sociali escludendo di fatto intere fasce della popolazione dalla fruizione di servizi essenziali come, appunto, quelli scolastici, educativi o sociali.

Un altro tema, connesso a questo, è relativo alle culture professionali degli operatori, che sono strettamente connesse con l'acquisizione di nuove competenze specifiche rispetto all'utilizzo di tecnologie considerate nuove nel lavoro dei servizi sociali. Ad esempio, si nota che nella formazione di base dei professionisti non vengono ancora sistematicamente trattati aspetti relativi all'adozione di tecnologie. Ciò determina una situazione in cui i professionisti conoscono solo parzialmente le possibilità offerte dall'utilizzo delle ICT con i loro pro e contro, sviluppando idee fuorvianti su di esse a causa di una conoscenza non strutturata perché basata solo sull'esperienza personale o in mancanza di un quadro teorico e professionale [11] debolmente sostenute dal confronto con i dati di ricerca e dall'evidenza non solo sui rischi, ma anche sui possibili benefici dell'utilizzo delle ICT [31].

A tal proposito sono due i principali elementi evidenziabili. Il primo fa riferimento al lavoro diretto con le persone, per il quale si riscontrano tra gli operatori "credenze" che portano a ritenere che solo la relazione in presenza possa essere significativa ed efficace, pena il rischio di spersonalizzare gli interventi [9]. Lo stesso tema si ritrova già in Parton [22] secondo cui le sfide più importanti da cogliere fanno riferimento alla cultura dei professionisti più che agli strumenti in sé. Il secondo riguarda il lavoro inter-professionale, rivolto sia agli aspetti organizzativi delle equipe, sia all'organizzazione del lavoro all'interno del singolo servizio e tra i diversi servizi coinvolti nell'accompagnamento alle famiglie in situazione di vulnerabilità. La differente combinazione tra competenze digitali, dotazioni tecnologiche e culture professionali contribuisce a determinare esiti diversi per le famiglie che accedono ai servizi, come rilevato anche da una recente indagine condotta da IRS [2].

3) L'integrazione tra i professionisti. Come brevemente anticipato nel secondo punto, al focus sulla relazione tra professionisti e famiglie va aggiunto quello sul rapporto tra i professionisti stessi e l'organizzazione del lavoro all'interno dei loro servizi. Coerentemente con l'approccio della valutazione partecipativa e trasformativa, la risposta all'emergenza sanitaria ha portato molti Ambiti territoriali a rimodulare e riprogettare le loro modalità di incontro e scambio aprendo inedite occasioni di apprendimento. Infatti, se da un lato si riscontrano le difficoltà di alcuni territori che non hanno avuto le risorse personali, culturali e materiali per attivare un cambiamento nelle pratiche professionali, dall'altro si evidenziano risposte resilienti [28] in cui l'innovazione è stata innescata spesso per iniziativa di singoli operatori che hanno fatto da "pionieri" più che da azioni di sistema.

Ad esempio sono state utilizzate piattaforme (*Zoom*, *Meet*, *Skype*, *GoToMeeting*, *Teams*, *WebEx*) per gli incontri di equipe multiprofessionali e per il coordinamento organizzativo all'interno del servizio, modalità che prima dell'emergenza sanitaria sarebbero parse inopportune se non addirittura impossibili.

Tra le potenzialità che gli strumenti telematici hanno reso evidenti menzioniamo le opportunità rese disponibili per le equipe di professionisti per riunirsi, confrontarsi e lavorare insieme. Il distanziamento fisico e l'impossibilità di riunioni in presenza ha determinato la trasposizione di questi momenti in modalità telematiche. Da un lato ciò, insieme all'iniziale maggiore disponibilità di tempo, ha moltiplicato le possibilità di incontro tra i professionisti, venendo meno – almeno apparentemente – le difficoltà legate alla distanza, agli spostamenti, alle sovrapposizioni di impegni e mansioni. Questi elementi sembrano andare nella direzione di rendere finalmente più praticabile l'integrazione sociosanitaria prospettata da decenni e disattesa in molti territori [15], facilitando scambi e travalicando confini disciplinari. La condivisione delle stesse modalità organizzative, l'utilizzo dei medesimi strumenti (informatici e non solo), l'investimento nella trasparenza delle procedure, l'ottimizzazione dei tempi e delle agende di ciascun professionista si sono rivelate pertanto delle condizioni di partenza utili per la diffusione del lavoro agile in settori e servizi che non si considerino più isolati e separati gli uni dagli altri. Tuttavia, questa presunta maggiore facilità nel garantire raccordi interistituzionali e continuità alle riunioni di equipe non si è manifestata ovunque: per alcuni territori il lavoro da casa dei propri operatori ha di fatto rappresentato una brusca interruzione delle loro interazioni e una sospensione delle occasioni di confronto reciproco e collegiale tra servizi diversi. In altri, invece, gli incontri di equipe sono avvenuti in maniera frammentata e con colloqui individuali gestiti dall'assistente sociale con i singoli professionisti, ma senza realizzare incontri collettivi.

4) Strategie e tecnologie per includere le famiglie. Una ulteriore indicazione positiva riguarda le tante possibilità offerte dalle nuove tecnologie e dal loro utilizzo creativo e flessibile, capace di aprire scenari interessanti e generativi [10]. I servizi di messaggistica, *in primis WhatsApp*, le video-chiamate attraverso *tablet* o *smartphone* hanno permesso di ricreare un clima inclusivo per le famiglie e garantire continuità a interventi educativi domiciliari, colloqui specialistici o gruppi di confronto tra genitori o tra bambini. La scelta dello strumento, della piattaforma o della modalità da adottare è essa stessa un aspetto sul quale investire spazi di condivisione e confronto per rendere genitori e bambini coinvolti nelle decisioni e per individuare congiuntamente le migliori opzioni per comunicare e interagire (Ius, 2020a). L'innovazione richiesta dal lavoro a distanza rende necessario adottare soluzioni sia organizzative che tecnologiche adeguate alla sfida, per gli operatori e per le famiglie. Si richiama qui il paradigma delle tecnologie di comunità [25], secondo il quale *media* e *social media* possono essere fondamentali per attivare e mantenere connessioni e ricostituire il capitale sociale delle comunità attraverso, anche, azioni di *capacity building* [3; 12].

Affrontare questi strumenti e intraprenderne un utilizzo efficace e realmente rispondente alle esigenze delle famiglie non è un passaggio scontato. Uno dei rischi più evidenti è quello di standardizzare le procedure e gli strumenti telematici individuati come funzionali ai servizi e che, a prime e superficiali analisi, appaiono essere replicabili per tutte le famiglie. La riflessione che alcuni operatori condividono, invece, pone al centro la necessità di differenziare e personalizzare i percorsi e le modalità *smart*, senza cioè trascurare l'importanza di un approccio *tailor-made* [27; 4] capace di plasmarsi sulle esigenze e le caratteristiche di ciascuna famiglia.

Per molti versi le modalità di lavoro telematiche risultano purtroppo inedite nel settore dei servizi socioeducativi (Marzo, 2020). Sembrano pertanto rendersi necessarie azioni formative e di aggiornamento che possano fornire ai professionisti e ai servizi le più aggiornate competenze informatiche, tecnologiche e comunicative per continuare quel lavoro al fianco delle famiglie in situazione di vulnerabilità che P.I.P.P.I. e Reddito di Cittadinanza promuovono.

Rispetto, invece, alle criticità raccolte, ricorre il tema della difficoltà a ricreare, pur con strumenti telematici, video-chiamate, messaggistica, ecc., la vicinanza relazionale, vero valore aggiunto dei percorsi di accompagnamento con bambini e famiglie in situazione di vulnerabilità (Formenti, 2014). La dinamica relazionale è la dimensione probabilmente più complessa, ma non impossibile, da riprodurre a distanza, da reinventare attraverso modalità non convenzionali, piuttosto che ridimensionare alla luce delle limitazioni. In particolare, alcune delle testimonianze raccolte sottolineano le difficoltà nell'assicurare adeguate relazioni d'aiuto e di supporto necessarie per affrontare la fase di analisi e valutazione multidimensionale [26; 18] delle singole situazioni, funzionale a sua volta a co-costruire progetti condivisi realmente rispondenti ai bisogni di sviluppo dei bambini e benessere del nucleo familiare. Occorre tuttavia notare che, in questa fase, alcune famiglie, e in certi casi soprattutto i bambini, sembrano essere state particolarmente recettive e propositive rispetto alle modalità di interazione e di intervento a distanza, favorendo la continuità relazionale tra percorsi pre e post *lockdown*. Molti operatori hanno anche riferito di aver fatto una nuova, inattesa, esperienza, ossia che gli "assi di risonanza" [8], ossia quel qualcosa che ci tocca profondamente e emotivamente in una relazione, sono possibili anche online, come dimostra la situazione dei tifosi che piangono, urlano, parlano ai giocatori che sono nel campo della partita che stanno vedendo dallo schermo.

## BIBLIOGRAFIA

1. Barbera, Filippo, e Tania Parisi. «Gli innovatori sociali e le aree del margine». In *Riabitare l'Italia. Le aree interne tra abbandoni e riconquiste*, a cura di Antonio De Rossi. Roma: Donzelli, 2018.
2. Casartelli, Ariela, e Francesco Di Ciò. «Il Servizio sociale nell'emergenza Coronavirus». *Welforum.it*, 2020. <https://welforum.it/il-servizio-sociale-nellemergenza-coronavirus/>.
3. Chaskin, Robert J. «Building community capacity: A definitional framework and case studies from a comprehensive community initiative». *Urban affairs review* 36, n. 3 (2001): 291–323.
4. De Mönink, Herman. *The Social Workers' Toolbox: Sustainable Multimethod Social Work*. Routledge, 2017.
5. De Rossi, Antonio. *Riabitare l'Italia: le aree interne tra abbandoni e riconquiste*. Donzelli Editore, 2019.
6. Formenti, L. «Sguardi di famiglia. Tra ricerca pedagogica e pratiche educative.», 2014.
7. Haenens, Leen d', Sofie Vandoninck, e Verónica Donoso. «How to cope and build online resilience?», 2013.
8. Hartmut, Rosa. «Pedagogia della risonanza». Scholé, 2020.
9. Hill, Andrew, e Ian Shaw. *Social Work and ICT*. London: Thousand Oaks: SAGE Publications, 2011.
10. Ius, Marco. «L'accompagnamento come cura generativa. Una lettura del Programma Nazionale P.I.P.P.I.» *Pedagogia e Vita* 3 (2019): 35–45.
11. ———. «Progettare resiliente con bambini e famiglie in situazione di vulnerabilità». *Padova University Press*, 2020.
12. ———. «Social services cooperating with schools to enable community: The case of two territories of Southern Italy». *Child Care in Practice* 26, n. 4 (2020): 416–430.
13. Lo Presti, Veronica. «Positive thinking e sviluppo locale: quali approcci per la promozione dell'innovazione nelle aree interne.» *Sociologia e ricerca sociale*, n. 112 (2017).

14. Mariotti, I., e D. Di Matteo. «Coworking in emergenza Covid-19: quali effetti per le aree periferiche?», 2020.
15. Martelli, Alessandro. «Verso una nuova governance locale delle politiche sociali?» *Autonomie locali e servizi sociali* 1 (2007): 97–108. <https://www.rivisteweb.it/doi/10.1447/24563>.
16. Marzo, Patrizia. «Una circolare può cambiare il servizio sociale?» *Animazione sociale* 334 (2020).
17. Milani, Paola. «È nato un nuovo smart welfare?» *Animazione Sociale* 334 (2020): 31–34.
18. ———. *Educazione e famiglie. Ricerche e nuove pratiche per la genitorialità*. Roma: Carocci, 2018.
19. Milani, Paola, Marco Ius, Sara Serbati, Ombretta Zanon, Diego Di Masi, e Marco Tuggia. *Il Quaderno di P.I.P.P.I. Teorie, Metodi e strumenti per l'implementazione del programma*. Padova: Becco Giallo, 2015.
20. Ministero del lavoro e delle politiche sociali. *Linee di Indirizzo nazionali "L'intervento con bambini e famiglie in situazione di vulnerabilità"*. Roma: Ministero del lavoro e delle politiche sociali, 2017.
21. ———. «Linee Guida. I quaderni dei Patti per l'inclusione sociale». Ministero del lavoro, 2019.
22. Parton, Nigel. «Challenges to practice and knowledge in child welfare social work: From the 'social' to the 'informational'?» *Children and youth services review* 31, n. 7 (2009): 715–721.
23. Petrella, Andrea. «Distanti ma connessi? Lo smart welfare nei servizi socio-educativi ai tempi del Coronavirus». *Encyclopaideia* 24, n. 57 (2020).
24. Petrella, Andrea, Marco Ius, e Paola Milani. «Smart Welfare». *Studium Educationis* XXI, n. 2 (2020).
25. Rivoltella, Pier Cesare. *Tecnologie di comunità*. ELS La Scuola, 2017.
26. Serbati, Sara, e Paola Milani. «La tutela dei bambini. Teorie e strumenti di intervento con le famiglie vulnerabili». 2013, Carocci edizione.
27. Thompson, Neil. *Understanding social work: Preparing for practice*. Red Globe Press, 2020.
28. Ungar, Michael. «Systemic resilience». *Ecology and Society* 23, n. 4 (2018).
29. Vaquero, Eduard, Marco Ius, Sara Serbati, Paola Milani, e Maria Angels Balsells Bailón. «El uso de las TIC en el ámbito de la intervención socio-familiar: revisión de la literatura y experiencias teórico-prácticas». *EDUCACIÓN y TECNOLOGÍA. Propuestas desde la investigación y la innovación educativa*. 2016.
30. Vaquero, Eduard, Aida Urrea, e Anna Mundet. «Promoting resilience through technology, art and a child rights-based approach». *Revista de cercetare si interventie sociala* 45 (2014): 144.
31. Zenarolla, Anna. *Dalla cartella dell'operatore al fascicolo sociale elettronico*. Milano: Franco Angeli, 2013.

# Poster





# Inclusività per differenti disabilità e DSA: il caso del pacchetto LATEX Aaccessibility

Dragan Ahmetovic<sup>1</sup>, Tiziana Armano<sup>2</sup>, Cristian Bernareggi<sup>3</sup>, Anna Capietto<sup>4</sup>,  
Chiara Magosso<sup>5</sup>, Adriano Sofia<sup>6</sup>, Luisa Testa<sup>7</sup>

<sup>1</sup> Università di Milano Statale – dragan.ahmetovic [at] unimi.it

<sup>2</sup> Università di Torino – tiziana.armano [at] unito.it

<sup>3</sup> Università di Milano Statale – cristian.bernareggi [at] unimi.it

<sup>4</sup> Università di Torino – anna.capietto [at] unito.it

<sup>5</sup> Università di Torino – chiara.magosso [at] edu.unito.it

<sup>6</sup> Università di Torino – adriano.sofia [at] unito.it

<sup>7</sup> IIS “Bodoni - Paravia” – luisa.testa [at] bodoniparavia.it

## SINTESI

Presentiamo l'utilizzo, all'interno di software compensativi con sintesi vocale adoperati da persone con DSA, del pacchetto LATEX Aaccessibility, inizialmente sviluppato dal Laboratorio Polin per l'accesso a testi contenenti formule da parte di persone con disabilità visive.

## PAROLE CHIAVE

sintesi vocale, disabilità visive, DSA

## POSTER

I documenti PDF contenenti formule generati da LATEX (ma non solo) non sono solitamente accessibili mediante tecnologie assistive utilizzate da persone con disabilità visive (i.e., screen reader e display Braille). Il pacchetto LATEX Aaccessibility [3,5], sviluppato dal Laboratorio Polin, risolve questo problema permettendo di creare documenti PDF in cui le formule vengono lette da tali tecnologie assistive. Infatti, vengono generati automaticamente dei commenti nascosti nel documento PDF (mediante gli attributi /Alt e /ActualText) in corrispondenza di ogni formula. Tale testo alternativo risulta nascosto nel documento PDF, ma i principali screen reader (Jaws, NVDA, VoiceOver) vi accedono correttamente. Inoltre, sono stati creati dizionari<sup>1</sup> (in inglese e italiano) per NVDA e Jaws che forniscono la lettura delle formule in linguaggio naturale nel caso in cui l'utente non conosca i comandi LATEX. Nel 2020, con la collaborazione di ricercatori esperti di standard PDF, il pacchetto Aaccessibility è stato aggiornato alla v. 3.0 [5] e permette di ottenere un PDF con tag di struttura del testo. Il pacchetto Aaccessibility è gratuito e disponibile sul repository CTAN. È incluso nelle principali distribuzioni standard di LaTeX. Il pacchetto non genera al momento PDF/UA.

---

1 <https://github.com/integr-abile/axessdicts>

D'altro canto, esistono software compensativi dotati di sintesi vocale che supportano persone con DSA (Disturbi Specifici dell'Apprendimento) nella lettura e fruibilità di documenti digitali. Le prestazioni di tali strumenti tuttavia non sono al momento soddisfacenti nel caso di testi scientifici contenenti formule. Il Laboratorio Polin ha iniziato a studiare alcuni di questi software valutando nel dettaglio la possibilità di migliorarne le prestazioni in presenza di testi scientifici.

Si è osservato infatti che testi digitali in pdf (anche contenenti formule) redatti utilizzando il pacchetto LATEX Axessibility risultano accessibili con la sintesi vocale dei software compensativi usati da persone con DSA. Anche in questo caso per avvicinare l'esperienza di lettura a quella del linguaggio naturale (nelle lingue italiano e inglese) è stato creato un dizionario<sup>2</sup> per il software compensativo ePico!<sup>3</sup>.

I primi risultati ottenuti, ed attualmente soggetti ad una sperimentazione che coinvolge oltre al gruppo di ricerca del Laboratorio molti studenti dell'università con diagnosi di DSA, sono molto buoni.

I risultati definitivi della sperimentazione saranno presentati alla conferenza WEB4ALL 2021 [2].

In considerazione della difficoltà da parte degli studenti con disabilità e DSA di ottenere testi scientifici digitali in formato accessibile il Laboratorio ha avviato il progetto Biblioteca digitale accessibile [1]. L'obiettivo è di rendere disponibili testi scientifici in formato digitale accessibile per corsi universitari a indirizzo scientifico (matematica, fisica, ingegneria, economia, ecc...), e libri di matematica di base e statistica per corsi universitari anche non a indirizzo scientifico. Essi saranno disponibili in formato PDF e, se possibile, HTML nella sezione apposita del sito web del laboratorio<sup>4</sup>. Attualmente sono già disponibili alcuni libri di testo e dispense universitarie. I documenti in formato PDF sono stati realizzati appunto grazie al pacchetto Axessibility e all'applicazione Audiofunctions.web [4]. I documenti in formato HTML sono stati convertiti dal LaTeX usando il software LaTeXML.

---

2 [https://github.com/integr-abile/axessdicts/blob/master/ePico/latex\\_ePico\\_it\\_en.ldm](https://github.com/integr-abile/axessdicts/blob/master/ePico/latex_ePico_it_en.ldm)

3 <https://www.anastasis.it/catalogo-generale/epico/>

4 <http://www.integr-abile.unito.it/knowledge-transfer/accessible-library-2/>

## BIBLIOGRAFIA

1. Abdeljawad, Ahmed, Dragan Ahmetovic, Tiziana Armano, Cristian Bernareggi, Anna Capietto, Sandro Coriasco, Laura Garbolino, e Nadir Murru. «Integrabile: “biblioteca digitale” accessibile per le STEM». In *X Convegno Nilde Biblioteche virtuali per utenti Reali*, 2019. <https://iris.unito.it/retrieve/handle/2318/1703979/512383/poster27maggio2019.pdf>.
2. Ahmetovic, Dragan, Tiziana Armano, Cristian Bernareggi, Anna Capietto, Sandro Coriasco, Sara Magosso, Nadir Murru, Adriano Sofia, e Luisa Testa. «On the use of speech synthesis for the access of scientific documents by students with learning disabilities». In *International Cross-Disciplinary Conference on Web Accessibility*, 2021.
3. Ahmetovic, Dragan, Tiziana Armano, Cristian Bernareggi, Anna Capietto, Sandro Coriasco, e Nadir Murru. «Accessibility 2.0: creating tagged PDF documents with accessible formulae» 27/28 (2019): 138–45.
4. Ahmetovic, Dragan, Cristian Bernareggi, João Guerreiro, Sergio Mascetti, e Anna Capietto. «AudioFunctions. web: Multimodal Exploration of Mathematical Function Graphs». In *Proceedings of the 16th Web For All 2019 Personalization-Personalizing the Web*, 1–10. San Francisco, 2019.
5. Armano, Tiziana, Anna Capietto, Sandro Coriasco, Nadir Murru, Alice Ruighi, e Eugenia Taranto. «An automatized method based on LaTeX for the realization of accessible PDF documents containing formulae». *Lecture Notes in Computer Science* 10896 (2018): 583–589.

# Sharing knowledge digitally, the Muruca case study

Chiara Aiola<sup>1</sup>, Giulio Andreini<sup>2</sup>, Francesca Di Donato<sup>3</sup>, Tiziana Lombardo<sup>4</sup>

<sup>1</sup> Net7, Italy - aiola [at] netseven.it

<sup>2</sup> Net7, Italy - andreini [at] netseven.it

<sup>3</sup> ILC-CNR, Italy - francesca.didonato [at] ilc.cnr.it

<sup>4</sup> Net7, Italy - lombardo [at] netseven.it

## ABSTRACT

In this poster we present the evolution of Muruca<sup>1</sup>, a platform that has been conceived as a framework to allow Digital Humanities researchers and research teams to create, curate and share their own Digital Editions. The poster describes the main features of the framework, its evolution over the years in synergy with the researchers' needs and with the progress of technologies and policies, and the open challenges and opportunities of the framework in a European context.

## KEYWORDS

open culture, digital humanities, digital libraries, digital editions, electronic objects

## POSTER

### 1 INTRODUCTION

In the Digital Humanities, there is a growing demand for software suites that allow users to manage the Digital scholarly Edition<sup>2</sup> of a corpus of digital documents (texts, images, metadata), from the management phase (entering texts and metadata, semantic enrichment, management of editorial workflows), to the publication on the Web.

This demand is indicative of an expanding market, characterized by a high number of small individual projects, normally with very high scientific content but often with an insufficient level of optimization and pooling of software resources and technical and editorial procedures [5]. In this context, the collaboration between academic and technical teams is at the centre of the Digital Edition publication process.

---

1 <http://www.muruca.org/>

2 While providing a complete definition of the expression Digital scholarly Edition is out of the scope of this presentation, we define it here as a critical representation of texts and historic documents, created, shared, and published on the Web [see 1, 6].

## 2 MURUCA'S WORLD

Muruca is a framework composed of different components. In its macroscopic structure we can divide it into 2 parts:

- The backend. It's the platform used to insert and manage data. It allows to create “records” for each work with its metadata, like in physical libraries, but also to associate transcriptions and images or other multimedia. The type of data relay on each project's object of study.
- The frontend. It's the interface used by the users for consultations. It gives tools to consult the materials, such as textual search, search of metadata, image viewer and text reader. It should be a tool for researchers, but also a way to disseminate the results of the research itself.

Its evolution took place in line with its research ecosystem and rests upon these three main pillars:

- Easiness of use. Technology shouldn't be an obstacle for users, but a useful work tool
- Standardization of the framework. It is a choice of maintainability and sustainability
- Split between content and its design. The contents of a publication are the real value of the project. The results of the researches had to be maintained regardless of the evolution of the digital platform.

### Data model and data preservation

After an analysis of the whole scenario of available content management systems, it has been decided to build the backend on the popular CMS Wordpress<sup>1</sup> by developing a group of plugins, one for each functionality, which allows the creation of as many content types as we need with all their related metadata. This choice responds to the need for a flexible data model definition coupled with the need of having a usable interface to manage all these entities. Furthermore, attention to preservation issues has been given by adding functionality that allows exporting data into files that can be stored everywhere, such as in a Zenodo<sup>3</sup> record or into thematic databases' aggregators.

### Data visualisation

The Muruca's frontend is the result of many years of experience and iterations with Digital Humanities research teams, its design and development took into account also case studies belonging to different domains like data visualisation and digital archiving to offer the possibility of organising all the elements to obtain a scientifically valid content visualisation and a usable interface for selected typologies of users.

---

3 “ZENODO is a CERN Data Centre-backed research data repository for the long-tail of science, enabling researchers to preserve and share their research output from any science, regardless of the size and format.” (from <https://zenodo.org/record/51902#.X1czg3kzZEY>).

The platform also gives a Rest API Service to publish content used by Muruca.

Documentation on <https://developers.zenodo.org>

## Sustainability and reusability in the long term

Muruca is designed to be a modular platform, in which there are a number of components available to build each project, but each component is the same (with a different configuration) for all Muruca installations. This solution makes it easy to create a new instance of Muruca or update all its installations.

The frontend uses external tools and libraries to implement some functionalities, the choice of which of them have to be included in the framework is based on those that are widely used and accepted as a standard or developed by a large community. The choice to adopt TEI Publisher<sup>4</sup> as the text reader in the new version of Muruca is a step forward in this direction.

This way, the platform is conceived as a modular object where the sustainability of every single object is meticulously planned.

## 3 PERSPECTIVES AND FUTURE EVOLUTIONS

Although we present Muruca as a virtuous example, the fast-changing scenario in the Digital Humanities opens to new challenges and issues to deal with.

The first tension is between the so-called pret-a-porter and tailor-made solutions [5]. Digital scholarly Editions are offered from many different platforms, which are very often designed and developed for one specific project and are mostly produced for it by research teams and specialized developers. The consequences of this situation are manifold: firstly, the costs of the development are very high and are not scalable, secondly, despite the fact that most editions use the TEI format for their data model, software standardization is far from being reached. The fragmentation and (unpredictable) discontinuity of funding was the main limitation we had to face when building a full "pret-a-porter" solution which, on the contrary, needs a planned long-term strategy and financial support.

Moreover, there is a lack of harmonization in the specifications of users' needs - which depends on the fact that there's not a full agreement on common standards within the community.

In order to address these challenges, we are planning a new design phase in Muruca, centred around the development of open science methodologies and standards. Muruca's evolution will be focused on FAIR data, reproducible processes, open access and long-term sustainability, in order to conceive a way to measure the scholarly edition impact and outreach, taking into account multiple metrics and indicators and, at the same time, considering standardization and preservation. All these aspects will be investigated thanks to a collaboration between the authors of this paper and other potential European partners experts on this domain.

---

4 <https://teipublisher.com/index.html> is a text viewer for XML-TEI encoded text developed by the community and licensed under the GPLv3.

## REFERENCES

1. Gabler, Hans Walter. 'Theorizing the Digital Scholarly Edition'. *Literature Compass* 7, no. 2 (2010): 43–56. 10.1111/j.1741-4113.2009.00675.x.
2. Gomes, Daniel, and Mário J. Silva. 'Modelling Information Persistence on the Web'. In *Proceedings of the 6th International Conference on Web Engineering*, 193–200, 2006.
3. Griffin, Gabriele, and Matt Steven Hayler. 'Collaboration in Digital Humanities Research—Persisting Silences'. *Digital Humanities Quarterly* 12, no. 1 (2018).
4. Mancinelli, Tiziana, and Elena Pierazzo. *Che cos'è un'edizione scientifica digitale*, 2020.
5. Pierazzo, Elena. 'What Future for Digital Scholarly Editions? From Haute Couture to Prêt-à-Porter'. *International Journal of Digital Humanities* 1, no. 2 (1 July 2019): 209–20. <https://doi.org/10.1007/s42803-019-00019-3>.
6. Sahle, Patrick. '2. What Is a Scholarly Digital Edition?' In *Digital Scholarly Editing: Theories and Practices*, edited by Matthew James Driscoll and Elena Pierazzo, 19–39. Digital Humanities Series. Cambridge: Open Book Publishers, 2017. <http://books.openedition.org/obp/3397>.



# L'Urban Digital Twin per cittadini consapevoli

Fulvio Ananasso<sup>1</sup>, Andrea Cruciani<sup>1</sup>, Bernardo De Bernardinis<sup>1</sup>,  
Sergio Farruggia<sup>1</sup>, Sandro Fontana<sup>1</sup>, Flavia Marzano<sup>1</sup>, Sandra Troia<sup>1</sup>  
Stati Generali dell'Innovazione – info [at] statigeneralinnovazione.it

## SINTESI

Il concetto *Digital Twin* applicato al contesto delle città/comunità intelligenti è ancora poco dibattuto in Italia. **L'inserimento del *Digital Twin* incrementa la complessità di una città/territorio**, anziché rappresentare uno strumento reso disponibile dal progresso tecnico-scientifico di cui sia indispensabile dotarsi per semplificare lo sviluppo della città. Occorrerà quindi **porre attenzione anche alle forme di comunicazione adottate per introdurre il *Digital Twin* nel contesto delle comunità urbane**.

Il *Digital Twin* quale “ponte” tra mondo fisico e mondo virtuale impone l'approfondimento del **bisogno di competenze umanistiche indotto dalla Trasformazione Digitale**. Un contesto, questo, in cui **si auspicano confronti e collaborazioni con gli esperti delle *Digital Humanities***: per esplorare le condizioni e le modalità con cui queste potranno favorire la partecipazione ai processi decisionali, facilitando la comunicazione tra le Amministrazioni e i cittadini.

## PAROLE CHIAVE

Digital Twin, smart city, city model, digital divide

## POSTER

### 1 IL DIGITAL TWIN DELLE CITTÀ

Il Digital Twin [4, 5] di una città/comunità [1, 6] è un modello 3D intelligente della città (City Model). Rappresenta accuratamente tutte le infrastrutture presenti della realtà fisica, dagli edifici, alle strade, alle reti tecnologiche. Tale modello è alimentato da una grande quantità di dati (Big Data) generati da sistemi sensoriali e di acquisizione dati in tempo reale in varie parti dell'infrastruttura urbana (Internet of Things, IoT).

Affinché una città possa essere modellata, è quindi implicita la sua ridefinizione in Smart City, aggregando e integrando Digital Twin specifici di sistemi quali gli smart building, le infrastrutture stradali, le reti elettriche, la gestione idrica e, in generale, quanto disponibile dalle infrastrutture critiche [2].

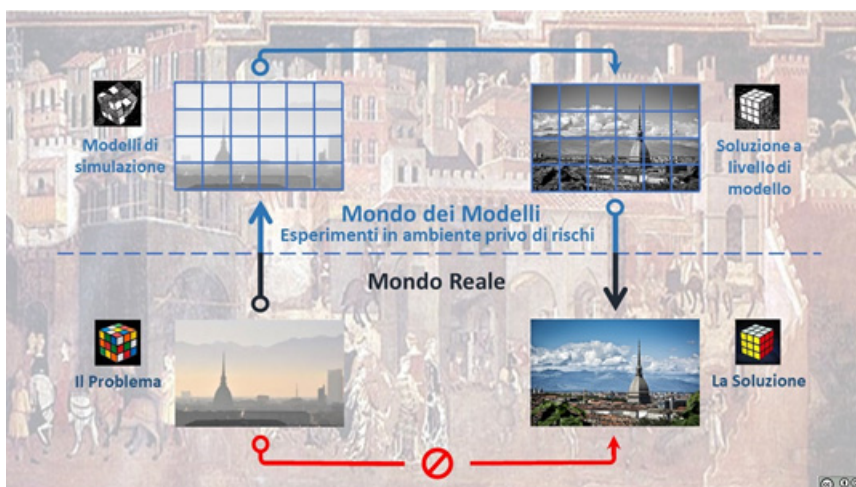


Figura 1: schema concettuale del Digital Twin per la Smart City

Tramite l'**Intelligenza Artificiale**, è consentita la navigazione e l'**analisi in tempo reale** in questo universo di informazioni, rendendo il *Digital Twin* di una città un modello affidabile.

Questa tecnologia declinata per la Città avrebbe potuto dare un grande contributo per una migliore gestione della pandemia COVID-19. Ma occorre uno dei componenti principali: **i dati provenienti dai suoi abitanti per modellare e poterne prevedere anche i comportamenti**. Andranno quindi acquisiti e valutati i **dati provenienti dai social network e i metadati dalle reti di comunicazione cellulare e dai sistemi di domotica**. Ciò permetterà di anticipare, ad esempio, le soluzioni per i rischi ambientali, pianificare le evacuazioni di emergenza, monitorare grandi eventi pubblici, prevedere situazioni critiche,...

## 2 COMPETENZE TECNICHE E UMANISTICHE

*“Cerchiamo di non guardarci indietro con rabbia o in avanti con paura, ma intorno con consapevolezza”* (J. Thurber).

Questa frase riassume quanto oggi, anche nel mondo delle tecnologie, la pandemia ci ha insegnato e non dobbiamo dimenticare. L'eliminazione del divario digitale è la prima delle priorità da considerare: divario non solo infrastrutturale, ma anche sociologico, economico, culturale, generazionale e di genere.

Tale sfida potrà essere affrontata solo con interventi coordinati di tutti gli *stakeholder*: politici, istituzionali, imprese, università e centri di formazione, nonché, ovviamente, i cittadini.

I cambiamenti culturali necessari per garantire la consapevolezza che richiamava *James Thurber* passano dalla coscienza dei limiti, dal riesame degli obiettivi e dall'analisi dei dati che possono indicare con maggior chiarezza gli interventi necessari per raggiungere gli obiettivi. Obiettivi che devono essere concreti e misurabili evitando pregiudizi e percezioni, quindi, ancora una volta, basati sui dati.

La costruzione delle città/comunità intelligenti ruota attorno ai dati. Tale prassi, inserita nel contesto di *Public-Private Partnership*, genera preoccupazioni presso la cittadinanza rispetto ai temi della *privacy*, a causa di ambiguità sulla proprietà dei dati e sulla liceità del loro utilizzo. Emblematico il caso del progetto riguardante la riqualificazione del lungomare *Quayside* di Toronto (circa cinque ettari) intrapresa tramite una *partnership* tra l'Agenzia governativa *Waterfront Toronto* e *Sidewalk Labs*, società di *Google Alphabet* (attratta da una possibile estensione del progetto su un'area di oltre 300 ettari). La raccolta dei dati per questa finalità, iniziata nel 2018, ha innescato una opposizione da parte della popolazione locale, accentuatasi con l'avvio del progetto nel 2019. *Sidewalk Labs* ha annunciato il 7 maggio 2020 l'abbandono del progetto. Attualmente la riqualificazione del quartiere è sospesa<sup>1</sup>. La raccolta dell'enorme quantità di dati riguardanti l'ambiente urbano, la vita cittadina e i risvolti relativi alla libertà dei cittadini sono assurti agli onori della cronaca in questo periodo di emergenza dovuta alla pandemia COVID-19. Tuttavia, la tematica, come dimostra il caso segnalato sopra, precede la crisi sanitaria in corso, e resterà inevitabilmente sempre aperta. **L'adozione di *Digital Twin* per le *Smart City/Community* ne amplifica i confini di approfondimento, anche grazie ai contributi delle *Digital Humanities*.**

### 3 URBAN DIGITAL TWIN E DIGITAL HUMANITIES

Le potenzialità della **Trasformazione Digitale** sono sovente riferite ad un processo di digitalizzazione concepito come una sorta di “digi-copiatura” di oggetti e processi materiali in termini di informazioni gestibili al *computer*. Diretta conseguenza di questo approccio mentale è la tendenza a fare riferimento a due mondi separati, quello reale, fisico, a cui viene affiancato quello digitale come “virtuale”, nell'accezione di mondo “non-reale”, “non-materiale”.

Il tema della natura materiale del digitale, in quanto energia resa manipolabile dalla tecnologia, porta a sostituire il concetto di “digitalizzazione” con quello assai più ampio di “virtualizzazione”, riprendendone il significato originario latino “vis”, ovvero “essenza vitale” da cui “virtus”, la capacità di una persona di agire in maniera coerente con la sua natura di essere umano. Virtualizzare, quindi, come individuazione e descrizione algoritmica della “vis”, ossia dell'essenza dei processi che osserviamo intorno a noi: “Il virtuale è il mondo dell'informazione, non dell'astrazione. L'informazione è già presente nella realtà, l'informazione è la realtà.” [3].

**La capacità di rappresentare informazioni attraverso l'energia invece che tramite la massa ha aperto all'essere umano un mondo di opportunità completamente nuove, grazie alla diversa essenza del mezzo di trasmissione dell'informazione.**

In tale quadro, Stati Generali dell'Innovazione considera lo sviluppo e la diffusione degli *Urban Digital Twin* un ambito applicativo di rilievo delle *Digital Humanities*, ad iniziare dall'approfondimento delle condizioni e delle modalità con cui l'informatica umanistica potrà favorire la partecipazione ai processi decisionali, facilitando la comunicazione tra amministrazioni e i cittadini e, in generale, contrastando e mitigando l'acuirsi del divario digitale, indotto dall'impiego delle tecnologie.

---

1 Sito web di Sidewalk Toronto e della campagna #blocksidewalk

<https://www.sidewalktoronto.ca/plans/quayside/>

<https://www.blocksidewalk.ca/>

Vuoi, se non misurarla esattamente, **per contribuire a percepire e possibilmente stimare l'incertezza di cui è permeato il *Digital Twin* della città**, pur nella certezza dei suoi *bit*. Vuoi, ancora, per raccogliere e analizzare **“l'impronta digitale”** dei cittadini che apprendono, intendendo **l'ambiente di apprendimento digitale come “gemello”** dell'organizzazione formativa analogica.

## BIBLIOGRAFIA

1. ABI Research. 'Transformative Urban Digital Twin and City Modeling Deployments to Exceed 500 by 2025', 2019. <https://www.abiresearch.com/press/transformative-urban-digital-twin-and-city-modeling-deployments-exceed-500-2025/>.
2. ARUP. 'Digital Twin: Towards a Meaningful Framework', 2019. <https://www.arup.com/perspectives/publications/research/section/digital-twin-towards-a-meaningful-framework>.
3. Canali, Mario. 'Quando l'informazione Si Fa Mondo', 1993. <http://www.arcnaut.it/installazioni/satori/contenutiframe/skvir.html#indice>.
4. Miskinis, Carlos. 'The Mysterious History of Digital Twin Technology and Who Created It'. Challenge Advisory, 2019. <https://www.challenge.org/insights/digital-twin-history/>.
5. ———. 'What Is Digital Twin Technology and What This Concept Means'. Challenge Advisory, 2018. <https://www.challenge.org/insights/what-is-digital-twin/>.
6. Weekes, Sue. 'The Rise of Digital Twins in Smart Cities'. Smart Cities World, 2019. <https://www.smartcitiesworld.net/special-reports/special-reports/the-rise-of-digital-twins-in-smart-cities>.

# Fostering the collaborative creation of Linguistic Linked Open Data with LexO, an open source editor of multilingual lexicons and terminologies

Andrea Bellandi, Emiliano Giovannetti, Simone Marchi, Silvia Piccini, Flavia Sciolette  
ILC-CNR – name.surname [at] ilc.cnr

## ABSTRACT

In the context of the Semantic Web and the Linguistic Linked Open Data paradigm, we propose LexO, a web editor of computational lexicons and terminologies based on the OntoLex-Lemon model.

## KEYWORDS

LexO, Linguistic Linked Open Data, Semantic Web, Computational Lexicography, Terminology, CLARIN

## POSTER

### 1 INTRODUCTION

The growing availability of digitized data has led to the increased diffusion of linguistic resources compliant to the Semantic Web (SW) [2]. In a framework oriented to new paradigms, like Linked Open Data (LOD), the research perspective has increasingly focussed on sharing standards and methods, connecting and reusing resources, encouraging collaborative approaches [3]. In this context, we conceived LexO [1], an open-source editor that, in line with the open science philosophy and the FAIR principles [5], allows lexicographers and terminologists to collaboratively create computational lexicons and terminologies according to the OntoLex-Lemon model<sup>1</sup> (lemon). This latter is, indeed, considered the de facto standard in the context of the SW and has been recently introduced in VocBench [4], born to manage OWL ontologies and SKOS(/XL) thesauri. However, unlike VocBench, LexO is tailored to exclusively manage lemon resources and is designed for users who are not familiar with the formalisms underlying the model, and, in general, the SW and LOD related technologies.

---

1 <https://www.w3.org/2016/05/ontolex/>

## 2 LEXO

LexO makes the creation of resources easy, by abstracting from the verbosity of the lemon model that is based on complex ontological design patterns<sup>2</sup>. In the LexO interface, distinct views are devoted to the linguistic and the conceptual dimensions, thus allowing a twofold approach in the construction of resources. On the one hand, in accordance with the onomasiological approach traditionally adopted in terminology, it is possible to import a domain ontology and then to lexicalize the concepts, on the other hand, users can adopt the semasiological perspective typical of lexicography, focussing their work on the definition of lexical senses. Furthermore, LexO includes functions to improve collaborative work. A team of users can work simultaneously on the same resource: the system keeps it consistent by locking the entries in the works, and users can add notes about a specific entry compiled by other users. LexO provides three types of user-profiles: viewer (can only access the dictionary view), editor (can create/modify entries), administrator (can add/remove users with specific profiles, validate entries, import the ontology, access the statistics of the lexicon).

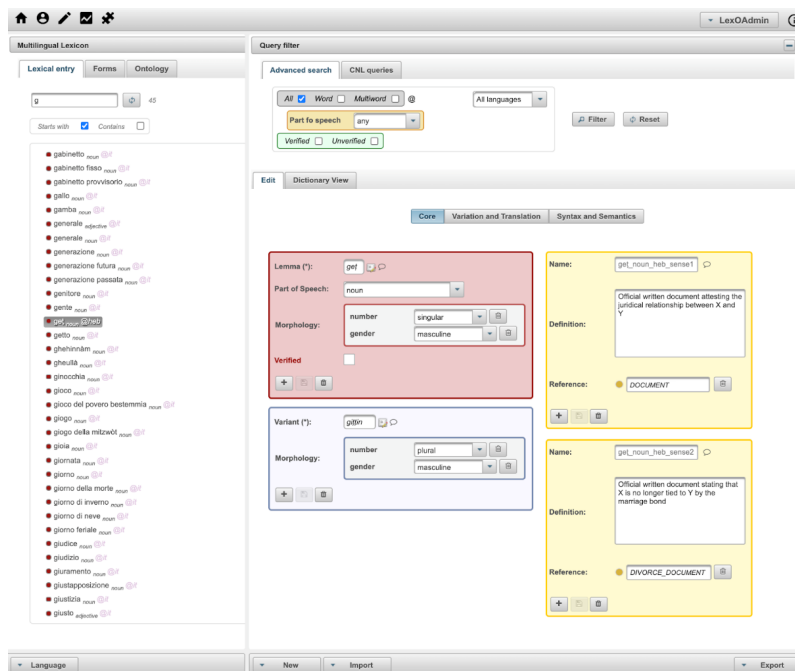


Figure 1: lexical entries are listed on the left part of the GUI. Red, blue, and yellow boxes represent respectively the lemma, the forms, and the lexical senses related to the selected entry

LexO GUI is depicted in figure 1. Through the select button at the centre of the interface, a lexicographer can switch to the “variation and translation” or to the “syntax and semantics” part, according to the lemon model.

2 [http://ontologydesignpatterns.org/wiki/Main\\_Page](http://ontologydesignpatterns.org/wiki/Main_Page)

The former module consists of two kinds of relations: i) the semantic relations holding between senses and including terminological relations (dialectal, register, chronological, discursive, and dimensional variation) and the translation relation, ii) the relations linking lexical entries and/or forms, which describe, e.g. the morphological and orthographic variations of a word. On the other hand, the “syntax and semantics” module allows describing the syntactic behaviour of a word and its government pattern. LexO allows mapping the argument of a predicate defined in an ontology and the syntactic argument introduced in a given syntactic frame. LexO has been already applied and tested in several projects, among which we cite DiTMAO (Dictionary of Old Occitan medico-botanical terminology), FdS (Diachronic multilingual lexicon of Ferdinand De Saussure’s terminology), Totus Mundus (Bilingual Chinese-Italian resource of Matteo Ricci’s map), and the Project for the Translation of the Babylonian Talmud into Italian. Furthermore, LexO was recently presented in the context of the CLARIN-IT Tour<sup>3</sup>. The tool can be freely downloaded from GitHub<sup>4</sup> and installed following the instructions included therein.

### 3 ACKNOWLEDGEMENT

This work was supported by the TALMUD project and carried out within the scientific collaboration between S.c.a r.l. PTTB and ILC-CNR (09/11/2017).

---

3 <https://www.clarin.eu/blog/clarin-it-presents-lexo-where-lexicography-meets-semantic-web>

4 The suffix “lite” refers to the current limitation to the management of large-sized resources. A full version of LexO overcoming this restriction is under development.

<https://github.com/andreabellandi/LexO-lite>

## REFERENCES

1. Bellandi, Andrea, Emiliano Giovannetti, and Anja Weingart. 'Multilingual and Multiword Phenomena in a Lemon Old Occitan Medico-Botanical Lexicon'. *Information* 9, no. 3 (2018): 52. <https://doi.org/https://doi.org/10.3390/info9030052>.
2. Cimiano, Philipp, Christian Chiarcos, John P. McCrae, and Jorge Gracia. *Linguistic Linked Data*. Springer, 2020.
3. McCrae, John, Guadalupe Aguado-de-Cea, Paul Buitelaar, Philipp Cimiano, Thierry Declerck, Asunción Gómez-Pérez, Jorge Gracia, Laura Hollink, Elena Montiel-Ponsoda, and Dennis Spohr. 'Interchanging Lexical Resources on the Semantic Web'. *Language Resources and Evaluation* 46, no. 4 (2012): 701–719. <https://doi.org/https://doi.org/10.1007/s10579-012-9182-3>.
4. Stellato, Armando, Manuel Fiorelli, Andrea Turbati, Tiziano Lorenzetti, Willem van Gemert, Denis Dechandon, Christine Laaboudi-Spoiden, Anikó Gerencsér, Anne Waniart, and Eugeniu Costetchi. 'VocBench 3: A Collaborative Semantic Web Editor for Ontologies, Thesauri and Lexicons'. *Semantic Web* 11, no. 5 (2020): 1–27. <https://doi.org/10.3233/SW-200370>.
5. Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, and Philip E. Bourne. 'The FAIR Guiding Principles for Scientific Data Management and Stewardship'. *Scientific Data* 3, no. 1 (2016): 1–9. <https://doi.org/https://doi.org/10.1038/sdata.2016.18>.



# Translation Studies Assistant for Master Students with Different Abilities

Irina Belyakova<sup>1</sup>, Nataliia Drozhashchikh<sup>2</sup>

<sup>1</sup> University of Tyumen – i.e.belyakova [at] utmn.ru

<sup>2</sup> University of Tyumen – n.v.drozhashhikh [at] utmn.ru

## ABSTRACT

Learners with special needs are in the focus of modern digital pedagogy. They are part of a wide field of inclusive education that covers a whole range of physical, cognitive, sensory, behavioural, and emotional disorders. Such students enter higher educational institutions to study foreign languages and theoretical disciplines in the Applied Linguistics programme. To date, in Russia, there are few educational platforms for students with special needs. The methodology for teaching such students to work with children in schools or with the computer software for translation is in the process of its development. Our project aims to identify the existing problems in the system of higher inclusive education in Russia and suggest ways to address these problems. Our special attention is focused on human-computer assistive technologies enabling to teach foreign languages, linguistic disciplines, and translation. The poster reveals the main principles and methodology of working with the source and target texts in Translation Studies class using an interactive application that allows students with learning disabilities to take their first steps in translation. The application is planned to be implemented and tested at the University of Tyumen, where the master students study the methodology of teaching foreign languages and explore the theory and practice of translation.

## KEYWORDS

human-computer interaction, assistive technologies, computer-aided tools, translation students, learning disabilities, translator assistant

# POSTER

## 1 INTRODUCTION

In the modern educational process, a number of students with diverse learning disabilities is steadily growing. To fully integrate them into the educational context, it is necessary to create conditions for the successful implementation of their educational route. This can be accomplished using assistive technologies (AT) and computer-aided tools (CAT) [2, 3, 6, 7] that support the students majoring in Applied Linguistics and Translation Studies. CAT tools include e-dictionaries, concordances, databases, parallel corpora, and translation memory tools which help translators to handle the source language (SL) and target language (TL) texts. They improve the translator's work and make it easy to work with the texts and reference literature. However, there are few applications for the students with learning disabilities aimed at working with the texts [4, 5]. The proposed translator assistant is an interactive technology, a fully functional tool for the English-Russian language pair intended to facilitate the translation workflow for such students. Creating and implementing the tool is a part of a pilot project at University of Tyumen «Assistive Technologies of Digital Tutors for Students with Learning Disabilities» aimed at master students. The assistant is addressed to creatively inclined students interested in literature, English as a foreign language, creative writing and translation. The main goal of the training assistant is not so much to teach professional translation skills to the students with learning disabilities but to provide them with an opportunity to express themselves creatively through English in the comforting atmosphere of their home, receive positive feedback from their tutor personified in the form of a digital chat-bot assistant with their teacher's facial avatar, pump up their self-esteem, enhance self-confidence, avoid negative emotions due to bullying, intentional or unintentional insulting behaviour in class. The chat-bot will not only test the students and provide them with the knowledge ontologies, but also offer them individual modes of work - as a teacher, a coach in foreign languages, and a translator. The poster aims at highlighting the translator mode.

## 2 METHODOLOGY AND DATA

The digital tutor is a program for interaction between a virtual teacher and a real student. The architecture of the prototype includes two groups of technologies: educational technologies (methods and techniques) and prototype technologies (algorithms and parameters). The prototype contains an interface with the dialogue control modules. The operation of the product is carried out in a dialogue mode with the user (messaging). The dialogue management module «prescribes» a scenario that defines the beginning and end of the dialogue, marks the types of speech moves, sets the inclusive modes that determine the content of the dialogue.

The prototype will use free tools such as the DeepPavlov library<sup>1</sup> (a set of programs for creating dialogue systems), a corpus of educational texts and encyclopedias, and the DBpedia ontology.<sup>2</sup>

---

1 <https://deeppavlov.ai/agent>

2 <https://wiki.dbpedia.org/services-resources/ontology>

The translator assistant takes an SL text as an input and produces an output as a marked-up interactive TL text with the highlighted noun and verb phrases and as a list of SL-TL units found in the input texts. The tool can be used as an e-dictionary. The tool interface will be easy to use. The process of translation will not be necessarily challenging. It is designed to be entertaining and interactive. Users will get an opportunity to acquire basic translation skills through the following stages of translation: linguistic, communicative, functional and ethical/aesthetic [1]. For the input, the English texts are submitted to the assistant. To date, the dataset includes the Aesop's tales<sup>3</sup>. The choice of the texts explains the concept of creating a comforting, entertaining and thought-provoking environment for the students. Animal characters are easy to animate, thus making the learning process more engaging and motivating. They teach empathy and universal life wisdom. In the future, the project will be supplemented with other texts and expanded to include texts of different registers and topics.

### **3 THE TRANSLATOR ASSISTANT: CHALLENGES AND PERSPECTIVES**

To create a prototype of a digital assistant, we identified the main criteria for building a prototype, and at present, we are at the stage of reviewing the literature on assistive technologies and developing the interface and content of the programme. How will the programme be executed? The human-computer interaction will be carried out in the following mode: the assistant will suggest the students guidance with their translation skills. It will find encouraging statements to assist them in the translation workflow and will choose the appropriate texts. During the process of interaction, only the positive emotive vocabulary will be used. The students press the button and listen to the story, or they can read along if they want to. Then they will look at the words and phrases they might not know. The students click on any word they want to check and see what it means. Then the assistant suggests checking grammar and working with different translation exercises (finding names, noun and verb phrases, translating the sentences into Russian, etc.). The assistant will offer different translation options. The students will practice and acquire new English language knowledge (grammar, vocabulary, style), culture-specific linguistic and extra-linguistic information (facts, terms, names of cultural objects, notions, etc.), and polish their style of writing in Russian. Eventually, besides language skills, through dialogues with the assistant, students will practice attention to detail, being focused on the task and curious, perform fully accurate and well-worded translations, self-checking and review steps. The proposed tool will make the translation process more accessible for students with cognitive, emotional, sensory or developmental disabilities.

### **4 ACKNOWLEDGEMENT**

This research is part of the «Assistive Technologies of Digital Tutors for Students with Learning Disabilities» project supported by the 2019/2020 Vladimir Potanin Foundation grant for the Master's programme teachers awarded to the authors<sup>4</sup>.

---

3 <http://read.gov/aesop/>

4 <https://www.fondpotanin.ru/competitions/projects/tekhnologii-inklyuzivnogo-obucheniya-na-baze-tsifrovyykh-tyutorov/>

## REFERENCES

1. Abdelaal, Nouredin. ‘Translation Theory’. In *Translation between English and Arabic*, 9–67. Cham: Palgrave Macmillan, 2020. [https://doi.org/10.1007/978-3-030-34332-3\\_2](https://doi.org/10.1007/978-3-030-34332-3_2).
2. Burstein, Jill, and Daniel Marcu. ‘Translation Exercise Assistant: Automated Generation of Translation Exercises for Native-Arabic Speakers Learning English’. In *Proceedings of HLT/EMNLP on Interactive Demonstrations*, 16–17. Vancouver: Association for Computational Linguistics, 2005.
3. González, Luis D., and Glenda M. Mejía. ‘Information-Based Translation Training Helps Undergraduates Carry out Translation Tasks in Actual Settings’. *Translation Journal*, October 2014. <https://translationjournal.net/October-2014/information-based-translation-training-helps-undergraduates-carry-out-translation-tasks-in-actual-settings.html>.
4. Goyal, Lalit, and Vishal Goyal. ‘Tutorial for Deaf–Teaching Punjabi Alphabet Using Synthetic Animations’. In *Proceedings of the 14th International Conference on Natural Language Processing (ICON-2017)*, 172–177. Colcata, India, 2017.
5. Pompili, Anna, Cristiana Amorim, Alberto Abad, and Isabel Trancoso. ‘Speech and Language Technologies for the Automatic Monitoring and Training of Cognitive Functions’. In *Proceedings of SLPAT 2015: 6th Workshop on Speech and Language Processing for Assistive Technologies*, 103–109. Dresden, Germany, 2015.
6. Schneider, Dominik, Marcos Zampieri, and Josef van Genabith. ‘Translation Memories and the Translator: A Report on a User Survey’. *Babel* 64, no. 5–6 (2018): 734–762.
7. Vela, Mihaela, Santanu Pal, Marcos Zampieri, Sudip Kumar Naskar, and Josef van Genabith. ‘Improving CAT Tools in the Translation Workflow: New Approaches and Evaluation’. In *Proceedings of MT Summit XVII*, 8–15. Dublino, 2019.

# Per un catalogo annotato della letteratura greca antica

Monica Berti  
Universität Leipzig – monica.berti [at] uni-leipzig.de

## SINTESI

Questa proposta di poster illustra un progetto di estrazione, analisi e annotazione semi-automatica delle forme linguistiche pertinenti alle citazioni bibliografiche nelle fonti greche antiche. Scopo del progetto è la realizzazione di un catalogo annotato della letteratura greca basato sul linguaggio dei testi antichi e sullo sviluppo di una metodologia aperta e condivisa per la conservazione del patrimonio testuale storico. Il progetto offre dati aperti e accessibili che possono essere utilizzati e arricchiti in ambienti collaborativi per la ricerca e la didattica.

## PAROLE CHIAVE

annotazione linguistica, CITE architecture, digital libraries, greco antico, named entities, NLP

## POSTER

### 1 ANNOTAZIONI LINGUISTICHE E CITAZIONI ANTICHE

Cataloghi e indici cartacei degli autori e delle opere della letteratura greca antica sono il risultato di una lunga tradizione filologica ed editoriale che ha permesso – e ancora permette – di raccogliere informazioni preziose sul patrimonio testuale dell'antichità classica. Il limite di queste risorse è rappresentato dal fatto che esse generalmente comprendono soltanto forme lemmatizzate in latino o in lingue moderne, richiedendo dunque una lettura separata e analitica dei passi da esse raccolti al fine di comprendere la varietà e la complessità del linguaggio utilizzato dagli autori antichi per esprimere le citazioni bibliografiche (come, per esempio, l'onomastica degli autori e le descrizioni e i titoli delle opere). Anche se in ambiente digitale le annotazioni morfo-sintattiche del greco e del latino stanno crescendo [4, 3], mancano ancora collezioni complete e annotazioni arricchite delle cosiddette «named entities» (NEs) riguardanti le citazioni bibliografiche antiche di autori e opere estratte dalle fonti testuali [1].

Θεόπομπος [Theopompos] (lemma)

Lemma: **Θεόπομπος** [Theopompos]  
lemma in LOGEION

Inflected form: **Θεόπομπον** [Theopompon] - [search in Kaibel Q](#)  
inflected form in TLG  
Named Entity Class: PER

10 . 8 [ και αὐτὸς ἦν πολυφάγος προειρήκαμεν, παραθέμενοι **Θεόπομπον** ιστοροῦντα ἐν τῇ πέμπτῃ καὶ τριακοστῇ ]  
11 . 63 [ [Θυατειρηγὸς](#) τὸ χωρὶς ὠτίων ποτήριον ὀνομακέναι **Θεόπομπον** ἐν [Μήδῳ Φιλήμων Φάσματι](#); ἐπει δ' ]

Inflected form: **Θεόπομπος** [Theopompos] - [search in Kaibel Q](#)  
inflected form in TLG  
Named Entity Class: PER

1 . 42 [ γυναικῶν ἀνακείμενοι ὑπὸ τῶ αὐτῶ ἱματίῳ. **Θεόπομπος**; ἐπίνομεν μετὰ ταῦτα ..... κατακείμενοι μαλακώτατ' ]  
1 . 47 [ ταῖς ἔξεσι τῶν πινόντων πλεῖστον χρόνον. **Θεόπομπος** δέ φησι παρὰ [Χίους](#) πρώτοις γενέσθαι ]  
1 . 61 [ ὅτι ὁ **Θεόπομπος** ὁ [Χίος](#) τὴν ἀμπελον ἱστορεῖ εὐρεθῆναι ]  
2 . 18 [ ἐκδιδούσης εἰς αὐτὸν κρήνης τινὸς πικρᾶς. **Θεόπομπος** δέ φησι περὶ τὸν [Ἐριγῶνα](#) ποταμὸν ]  
2 . 23 [ καθάπερ καὶ τὸν [Φρύγα Μίδα](#) φησὶ **Θεόπομπος**, ὅτε ἔλεῖν τὸν [Σιληνόν](#) ὑπὸ μέθης ]  
2 . 35 [ δ' αὐτομάτοισιν τὰ μιμαίκυλ' ἐφύετο πολλὰ. **Θεόπομπος**; τρώγουσι μύρτα καὶ πέποινα μιμαίκυλα. [Κράτης](#); ]  
2 . 62 [ διὰ τοῦ φ ἀσφάραγον ὀνομάζει. καὶ **Θεόπομπος**; κάπειτ' ἰδὼν ἀσφάραγον ἐν θάμνῳ τινί. ]  
2 . 77 [ καὶ παρ' [Αἰσχύλῳ](#); διαβρέχεις τάρτυματα. καὶ **Θεόπομπος** δέ φησι· πολλοὶ μὲν ἀρτυμάτων μέδιμοι, ]

Figura 1. Digital Athenaeus: Named Entities Concordance

Figura 1 mostra l'esempio della concordanza dei *Deipnosofisti* di Ateneo di Naucrati, la quale è liberamente consultabile online e offre diversi livelli di annotazione delle NEs (colori rosso e blu) pertinenti anche alle citazioni di autori e opere conservate nel testo greco (ed. Kaibel) per un totale di circa 23.000 token estratti, lemmatizzati e annotati. Questa risorsa si accompagna ad altri strumenti del progetto *Digital Athenaeus*<sup>1</sup>, i quali permettono una lettura ragionata del testo dei *Deipnosofisti* in quanto miniera preziosa di riferimenti bibliografici ad autori e opere della letteratura greca che risultano oggi per la maggior parte irrimediabilmente perdute [2]. Queste risorse possono essere utilizzate non soltanto per la ricerca filologica e storiografica, ma anche per la didattica delle lingue classiche e dell'informatica umanistica ad esse applicata.

## 2 CATALOGAZIONE APERTA E CONDIVISA DELLA LETTERATURA GRECA ANTICA

Obiettivo del poster è la presentazione del lavoro di estrazione e annotazione semi-automatica di NEs al fine di creare un catalogo annotato della letteratura greca antica. Il testo di riferimento è rappresentato dai *Deipnosofisti* di Ateneo, ma lo scopo finale comprende l'elaborazione di un modello di annotazione e la produzione di dati che saranno utilizzati per l'annotazione di altri testi della letteratura greca, a partire dal lessico *Suda* e dal lessico di Arpocrazione.

1 <http://www.digitalathenaeus.org>

Mediante la presentazione di un esempio concreto, le fasi illustrate nel poster sono le seguenti: 1) estrazione e lemmatizzazione delle NEs, 2) annotazione delle singole NEs in classi generiche sulla base dell'interrogazione di «authority lists» esterne, 3) disambiguazione delle NEs pertinenti ad autori e opere antiche con sottoclassi semantiche e creazione di citazioni canoniche previste dalla CITE Architecture [1, 2], 4) utilizzo della piattaforma di annotazione INCEpTION (<https://inception-project.github.io/>) per «relation extraction», «NE linking» e «co-reference resolution», 5) esportazione dei dati in formati standard per la condivisione e il riuso (CSV, TEI XML, CTS e CITE urns, PAULA XML, CoNLL, TCF, WebAnno TSV 3, XMI). Figura 2 mostra un esempio di diversi livelli di annotazione testuale all'interno di INCEpTION [2].

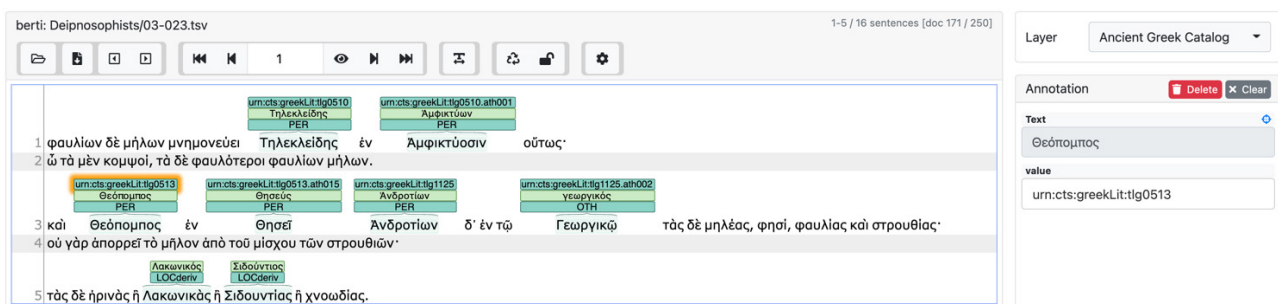


Figura 2. Livelli di annotazione linguistica in INCEpTION (Ath. Deipn. 3.23)

Il progetto si inserisce nelle linee guida del tema «Open Culture» proposto dagli organizzatori di AIUCD2021 per le seguenti ragioni: 1) generazione di un corpus testuale annotato per il greco antico liberamente accessibile e riutilizzabile, 2) produzione di dati e risorse linguistiche (forme flesse, lemmi e NEs) concernenti nomi propri in greco antico che ancora mancano per la conservazione e la rappresentazione digitale del patrimonio testuale della civiltà classica, 3) utilizzo di una piattaforma open source sviluppata per la linguistica e liberamente utilizzabile da studiosi e studenti per il lavoro collaborativo e la condivisione dei risultati, 4) creazione di una risorsa didattica e scientifica che permette l'analisi contestuale del greco antico, la creazione di dati, lo sviluppo di basi di conoscenza («knowledge bases») e l'apprendimento delle metodologie di annotazione ed estrazione delle lingue storiche, 5) la promozione e l'implementazione di metodologie, dati, risorse linguistiche e ambienti collaborativi che offrono giuste opportunità nel mondo della ricerca, della formazione e della condivisione del patrimonio culturale.

### 3 RINGRAZIAMENTI

Si ringrazia la Deutsche Forschungsgemeinschaft (progetto nr. 434173983) per il sostegno finanziario al progetto.

## BIBLIOGRAFIA

1. Berti, Monica. «Named Entity Annotation for Ancient Greek with INCEpTION». In *Proceedings of CLARIN Annual Conference 2019*. Leipzig: CLARIN, 2019.
2. Berti, Monica, Christopher Blackwell, Mary Daniels, Samantha Strickland, e Kimbell Vincent-Dobbins. «Documenting Homeric Text-Reuse in the Deipnosophistae of Athenaeus of Naucratis». *Bulletin of the Institute of Classical Studies* 59, n. 2 (2016): 121–139.
3. Celano, Giuseppe. «The Dependency Treebanks for Ancient Greek and Latin». In *Digital Classical Philology. Ancient Greek and Latin in the Digital Revolution*. Berlin: De Gruyter, 2019.
4. Vatri, Alessandro, e Barbara McGillivray. «The Diorisis Ancient Greek Corpus». *Research Data Journal for the Humanities and Social Sciences* 3, n. 1 (2018).

# Il ruolo del Terzo Settore nella lotta al divario digitale culturale

Pietro Biase  
pietro [at] pietrobiase.it

## SINTESI

Negli ultimi anni le discussioni e le iniziative per rendere le nostre città più *smart* si sono moltiplicate, con risultati altalenanti e dipendenti più dalla caparbietà dei singoli che da una capacità diffusa a livello di sistema. Molto spesso ci si è concentrati sull'aspetto tecnico delle smart cities, tralasciando quasi ovunque il ruolo centrale delle persone in questa trasformazione, ma un luogo, una città, una comunità quanto possono davvero definirsi *intelligenti* se i cittadini non riescono a *superare il digital divide di tipo culturale* e a diventare parte integrante di essa? Per aiutare l'Italia ad uscire dal grave ritardo rispetto alla maggior parte delle nazioni europee, si stanno muovendo istituzioni, pubblica amministrazione centrale e locale, aziende pubbliche e private, mondo dell'istruzione a tutti i livelli, gruppi di interesse ed entità del Terzo Settore.

## PAROLE CHIAVE

alfabetizzazione digitale, associazioni, cittadini, collaborazione, co-progettazione, digital divide, istituzioni, pubblica amministrazione, terzo settore



# POSTER

## 1 TITOLETTO

L'espressione *divario digitale* (traduzione letterale del termine anglofono *digital divide*), è nata in seno all'amministrazione statunitense all'epoca della presidenza Clinton (1993-2001) ed indica la disparità nelle possibilità di accesso ai servizi telematici tra la popolazione. L'uso dell'espressione è oggi diffuso a livello mondiale ed, in generale, viene usato per indicare la consapevolezza globale di una problematica di accesso ai mezzi di informazione e di comunicazione digitali, da parte di determinate aree geografiche o fasce di popolazione. In maniera più semplice, oggi il *digital divide* può essere definito come il punto di separazione tra chi ha accesso alla rete Internet ed è in grado di utilizzare tecnologie digitali e chi invece ne rimane escluso o fortemente penalizzato.

I motivi dell'esclusione possono essere molteplici: condizioni economiche, livello di istruzione (*digital divide culturale*), presenza e qualità dell'infrastruttura di connessione alla rete mobile o fissa (*digital divide infrastrutturale*), differenza di età (*digital divide intergenerazionale*), di sesso (*digital divide di genere*) e di etnia (*digital divide linguistico-culturale*).

## 2 IL DIGITAL DIVIDE CULTURALE IN ITALIA

Per quanto riguarda l'Italia, la situazione nella lotta al *divario digitale* non è rosea: l'edizione 2020 dell'indice DESI (Digital Economy and Society Index) [1] certifica ancora una volta l'arretratezza del nostro paese nella classifica che ha l'obiettivo di mappare il grado di digitalizzazione dell'economia e della società delle nazioni europee. Dopo aver guadagnato alcune posizioni nel 2019, quest'anno l'Italia torna in quartultima posizione tra gli stati dell'Unione Europea (UE). Tra le tante criticità, spicca il grave ritardo (ultima posizione in UE) del c.d. "*capitale umano*" in termini di competenze digitali di base e avanzate, a questo si aggiunge che il numero di specialisti e laureati nel settore ICT è al di sotto della media UE.

Il fenomeno ha quindi ampie ripercussioni negative sull'uso effettivo delle tecnologie digitali da parte dei cittadini italiani. L'incremento quantitativo e qualitativo dell'offerta di servizi di e-government<sup>1</sup> e business (si pensi ad esempio all'aumento del numero di piattaforme per gli *acquisti online* o della disponibilità di strumenti per il *lavoro remoto* o per la *didattica a distanza* durante le misure di confinamento dovute all'epidemia COVID-19) sembra quindi un'azione di puro adeguamento verso l'auspicata digitalizzazione dei servizi, ma privo di reale utilità pratica a causa del mancato utilizzo strutturale, diffuso e generalizzato da parte della collettività.

---

1 Termine derivato per analogia da e-business, che identifica l'ambito disciplinare relativo alle attività delle pubbliche amministrazioni realizzate grazie all'ausilio delle reti telematiche e della rete Internet.

### 3 IL RUOLO DI GOVERNANCE NEL SETTORE PUBBLICO

Dai risultati emersi con l'indice DESI 2020 risulta evidente come il ruolo di guida e di governance della pubblica amministrazione centrale e locale abbia di fatto fallito qualsiasi tipo di politica e di azione focalizzata sul *superamento del divario digitale* tra i cittadini italiani. Sono ancora pochissime le istituzioni locali con un assessorato dedicato alla *trasformazione digitale* o alla *smart city*, ed un po' ovunque – da nord a sud – si può toccare con mano lo scadente livello di trasformazione digitale della macchina comunale e la lentezza con cui i comuni e le regioni si stanno dedicando alla conversione tecnologica della PA locale.

In questo scenario, il Terzo Settore può giocare un ruolo importante nella collaborazione con tutte quelle pubbliche amministrazioni che vogliono riconoscere un ruolo attivo al mondo dell'associazionismo. Questo ruolo può portare ad una stretta collaborazione tra istituzioni pubbliche e Terzo Settore con l'obiettivo di ottenere la massima inclusione nelle politiche di *alfabetizzazione digitale* che vanno portate avanti in maniera capillare e trasversale, anche nei confronti delle fasce più deboli della società (giovani a bassa scolarizzazione, NEET - Not in Education, Employment or Training, immigrati, persone con disabilità, adulti over 50 esclusi dal mondo del lavoro, donne sole non occupate o con figli, anziani, persone detenute).

Le associazioni del territorio, da sempre schiacciate dalla capacità economica di pubblico e privato, hanno quindi l'occasione per dimostrare di poter dare un enorme contributo alla causa della *lotta all'esclusione digitale* di singoli individui o di gruppi sociali, grazie ad un approccio snello dal basso verso l'alto e sfruttando al meglio ciò che è nella loro natura: essere persone che fanno cose per le persone.

Le relazioni fra PA ed Enti di Terzo Settore dovranno essere ispirate ai principi di perseguimento dell'interesse pubblico, di efficacia, di trasparenza e, più in generale, di buon andamento della PA. La pubblica amministrazione, quindi, dovrebbe agire per riconoscere e coordinare le entità di volontariato presenti sul territorio utili ad affrontare la *lotta al digital divide*, potenziando e amplificando l'azione allocando a tale scopo le risorse utili a raggiungere gli obiettivi condivisi e prefissati.

### 4 AFFRONTARE IL DIVARIO DIGITALE

Affrontare il *divario digitale culturale* non è cosa semplice, e la costruzione del *ciudadino digitale*<sup>2</sup> del futuro va affrontata con serietà e determinazione.

Secondo recenti stime dell'Ocse, in Italia circa il 26% della popolazione tra i 16 e i 74 anni non ha mai navigato in rete, al fronte di una media del 14% negli altri Paesi dell'Organizzazione. Si tratta di 10 milioni di cittadini che non utilizzano Internet, e questo risulta essere uno degli elementi bloccanti verso l'uso diffuso di *servizi mediati dalla rete*. Non può essere considerato un caso se, già dal 2012, il Consiglio sui diritti umani delle Nazioni Unite ha approvato la risoluzione A/HCR/20/L.13, considerando espressamente l'accesso e l'utilizzo della rete Internet come uno strumento di circolazione delle idee e delle informazioni ed un insostituibile mezzo di attuazione dei diritti fondamentali dell'uomo e del cittadino.

---

2 Per cittadinanza digitale si intende l'insieme di diritti/doveri che, grazie al supporto di una serie di strumenti e servizi (identità, domicilio, firma digitale), ha l'obiettivo di semplificare il rapporto tra i cittadini, le imprese e la pubblica amministrazione tramite le tecnologie digitali.

Riuscire a progredire nell'*alfabetizzazione digitale di base* può consentire alle persone un accesso ed un utilizzo consapevole dei servizi online sia pubblici sia privati e, cosa ancor più importante in questo momento storico post COVID-19, riuscire a dare un minimo set di competenze ed informazioni per poter poi procedere con l'*auto-formazione online* attraverso i molteplici canali messi a disposizione dalla rete internet (blog, siti di informazione, corsi online, webinar, ecc.).

Lo strumento principale che il Terzo Settore può mettere a disposizione della collettività è l'opera di *tutoraggio* diffuso verso le persone, sia all'interno delle associazioni stesse sia presso i cittadini più vicini sul territorio. Questo genere di tutoraggio si può declinare in molti modi: nella formazione digitale generica o specialistica, nell'organizzazione di corsi o conferenze, fino ad arrivare ad attività più mirate come l'affiancamento e la consulenza one-to-one. Affrontare questa sfida partendo dal territorio nel quale è radicato il mondo dell'associazionismo non profit rende immediata la conoscenza e la percezione di necessità e bisogni, riuscendo ad evidenziare azioni macro e micro, frange deboli e/o emarginate nella popolazione. L'associazione o il soggetto non profit diventano quindi un *cuscinetto sociale* che aiuta a semplificare il rapporto tra cittadini ed istituzioni e a mettere in campo azioni concrete per il superamento del divario digitale in ambito sociale, generazionale, economico, culturale e di genere.

In questa sfida il Terzo Settore non può essere lasciato solo: il ruolo delle istituzioni con una visione strategica a medio-lungo termine è basilare e deve andare nella direzione di mantenere una funzione di governance dell'interoperabilità del sistema pubblico-privato-Terzo Settore, fungendo anche da garante per una collaborazione e co-progettazione scevra da interessi di parte. La *definizione delle strategie d'azione* dev'essere implementata da soggetti istituzionali che possano definire i bisogni, controllare la corretta messa in campo di tutte le azioni necessarie e innescare processi di miglioramento nel modo con cui si fanno le cose. Fare rete diventa dunque la chiave per mettere a fattore comune dati, informazioni, buone pratiche ed esperienze.

## 5 ESEMPI CONCRETI DI APPLICABILITÀ DEL MODELLO

In Italia ci sono già state un buon numero di applicazioni pratiche relativamente a questo paradigma di collaborazione tra associazioni e pubblica amministrazione locale, con l'obiettivo comune di ridurre il *divario digitale culturale* tra i cittadini del territorio.

Come non citare le esperienze pionieristiche dei *Punti Roma Facile*<sup>3</sup>, dei *PAAS* della Toscana<sup>4</sup>, del progetto *P3@Veneti*<sup>5</sup>, del progetto *Pane e Internet* dell'Emilia-Romagna<sup>6</sup> o dei *Punti Genova Digitale*<sup>7</sup> e del precedente progetto pilota *Genova Digitale*<sup>8</sup> che, con modalità del tutto simili, hanno permesso di sperimentare sul campo la cooperazione tra istituzioni e Terzo Settore, portando le attività di *alfabetizzazione digitale sul territorio*, in prossimità dei luoghi fisici vissuti giornalmente dai cittadini. Le sedi di varie associazioni del territorio oppure, come nel caso di molti progetti sopra menzionati, i locali messi

---

3 <https://www.comune.roma.it/web/it/partecipa-punti-roma-facile.page>

4 <https://www.regione.toscana.it/-/paas-cittadini>

5 <https://www.regione.veneto.it/web/informatica-e-e-government/p3a-veneti>

6 <https://www.paneeinternet.it/>

7 <https://associazione.opengenova.org/progetti/punti-genova-digitale/>

8 <http://genovadigitale.comune.genova.it/>

a disposizione da municipi o enti pubblici sono diventati naturali punti di contatto tra il cittadino e i *facilitatori digitali* che mettono a disposizione le proprie conoscenze e competenze.

È degna di nota anche l'attività svolta in collaborazione tra l'associazione Informatici senza Frontiere (ISF) e il Liceo Montessori di Roma [2]. Per questo progetto, gli studenti hanno lavorato come *tutor nei corsi di informatica* realizzati per gli anziani da ISF. Il gruppo di lavoro in fase di progettazione era composto da soci ISF e insegnanti, ed ha definito le linee generali di progetto, gli obiettivi e le modalità di erogazione della formazione.

Un'altra recente iniziativa nell'ambito delle *competenze digitali* è la stesura del libretto *#DIGIfenditi, manuale di autodifesa personale* [3], progettato dal Municipio Roma VII insieme a una serie di gruppi e società del settore (tra cui TSCAI – The Smart City Association Italy e Stati Generali dell'Innovazione). Nel vademecum sono affrontati i temi dell'utilizzo consapevole di smartphone e tablet, dei social network, delle password, della posta elettronica e di tutti ciò oramai che fa parte della vita quotidiana delle persone, ma necessita di attenzione per far fronte alle insidie digitali nascoste dietro lo schermo.

## 6 RISULTATI OTTENUTI

Non è stato possibile reperire informazioni dettagliate sui risultati ottenuti da tutti i progetti sopra menzionati, si ritiene tuttavia che, anche grazie alle caratteristiche di replicabilità di questi progetti, i risultati possano essere facilmente perseguiti da qualsiasi associazione che voglia impegnarsi attivamente sul proprio territorio (anche con un numero limitato di *facilitatori digitali*).

I *Punti Roma Facile* sono strutturati in 26 punti operativi<sup>9</sup> in tutto il territorio di Roma, di cui 17 attivati presso luoghi istituzionali messi a disposizione dai municipi. Altri 7 sono invece stati predisposti presso il Sistema Bibliotecario di Roma, e ulteriori 2 presso associazioni che hanno aderito all'iniziativa. Nel periodo novembre 2016-luglio 2017 ogni singolo *Punto Roma Facile* ha mediamente erogato 12 ore settimanali di formazione, per un totale di 2.500 cittadini che hanno usufruito di questa possibilità.

Il progetto *PAAS* della Toscana, nato tra il 2004 e il 2005, si è sviluppato negli anni successivi fino a rimanere in stand by a cavallo tra il 2018 e l'inizio del 2019. Nel 2018 la rete territoriale era composta di 156 punti *PAAS* collocati in sedi associative e di enti, diffusi in 95 comuni su 273 totali dell'intera regione. Negli anni precedenti (nel periodo 2007-2015), quando erano stati resi disponibili alcuni finanziamenti di sostegno specifici, si era arrivati ad avere 298 punti in 192 comuni (che rappresentano circa il 67% dei 287 comuni presenti all'epoca sul territorio toscano).

I risultati del progetto pilota *Genova Digitale* sono stati presentati nel novembre 2015 alla Reggia di Venaria (TO), durante l'Italian Digital Day, davanti alla presidenza del Consiglio dei Ministri. L'iniziativa genovese ha visto il coinvolgimento di 3 diversi assessorati (Sviluppo Economico, Informatica e Rapporti con i Municipi) e di tutti i municipi che compongono il capoluogo ligure (9). La durata dell'iniziativa di formazione è stata di circa 1 mese, ed ha visto l'organizzazione di 74 incontri pubblici dove sono stati affrontati alcuni argomenti di base sull'utilizzo di alcune applicazioni per smartphone di pubblica utilità. Tutto il progetto è stato gestito grazie al contributo di 27 associazioni del territorio e di 23 facilitatori digitali. Inoltre, tra novembre 2017 e novembre 2019 sono stati organizzati alcuni

---

9 Dati aggiornati a giugno 2018.

ulteriori eventi formativi sul territorio con il nome *Punti Genova Digitale*, che hanno visto un totale di 84 ore di formazione one-to-one erogate ai cittadini che ne hanno fatto richiesta.

L'attività di *tutoraggio nei corsi di informatica* organizzati da ISF si sono svolte presso due Centri Anziani in due diverse zone di Roma (Trastevere e Conca d'Oro). Gli studenti/tutor sono stati guidati e seguiti dai volontari ISF che già operano con queste finalità presso le due strutture coinvolte. Nel 2018-2019 il progetto ha coinvolto 22 studenti e circa 25 anziani mentre nel 2019-2020 il progetto ha coinvolto 30 studenti e circa 35 anziani. Un ulteriore gruppo di 15 studenti ha iniziato il percorso a fine gennaio 2020 ma dopo le prime due giornate di formazione è stato necessario interromperlo a causa delle restrizioni Covid-19. Gli studenti sono stati impegnati in una prima fase dedicata alla loro formazione allo scopo di prepararli al tipo di insegnamento a cui avrebbero dovuto fare fronte, di durata pari a circa 10 ore. Successivamente si sono svolte le attività di formazione/tutoring, per una durata complessiva di circa 16-18 ore. Infine è stato loro richiesto di elaborare l'esperienza svolta redigendo una relazione, con un impegno stimato pari a circa 4-6 ore.

## **BIBLIOGRAFIA**

1. European Commission. «The Digital Economy and Society Index (DESI)», 2020.
2. Informatici Senza Frontiere. «Quando i giovani fanno da tutor agli anziani: racconto di un'esperienza», 2020.
3. Municipio Roma VII. *#DIGIfenditi, manuale di autodifesa personale*. Roma: Municipio Roma VII, 2020.

# An Ontology for the Philosophy of Early Middle Ages.

Armando Bisogno<sup>1</sup>, Gianmarco Bisogno<sup>2</sup>, Renato De Filippis<sup>3</sup>, Giulio d'Onofrio<sup>4</sup>, Raffaella d'Urso<sup>5</sup>,  
Giuseppe Palermo<sup>6</sup> and Timothy Tambassi<sup>7</sup>

<sup>1</sup> Università di Salerno - arbisogno [at] unisa.it

<sup>2</sup> Università di Salerno - gibisogno [at] unisa.it

<sup>3</sup> Università di Salerno - rdefilippis [at] unisa.it

<sup>4</sup> Università di Salerno - gdonofrio [at] unisa.it

<sup>5</sup> Università di Salerno - rURSO [at] unisa.it

<sup>6</sup> Università di Salerno - gipalermo [at] unisa.it

<sup>7</sup> Università di Salerno - ttambassi [at] unisa.it

## ABSTRACT

One of the main assumptions of Open Culture is to provide digital tools that might increase the opportunities to share information on cultural heritage. Philosophy, we claim, is an important part of such a common background. In this context, the purpose of OPhEMA project is to develop an IT ontology that facilitates processes of browsing and searching annotated texts in order to define the meaning (if any) of philosophy in the Early Middle Ages. The poster is meant to show the preliminary steps of the project, namely: a literature review aimed at identifying the occurrences of "philosophia", the creation of an (Open Access) Google Sheets file systematizing all the occurrences, and the design of an IT ontology referred to the Philosophy of the Early Middle Ages which combines such data.

## KEYWORDS

philosophy of the early middle ages, IT ontology, ontology of philosophy, Latin literature

## POSTER

The aim of this poster is to present the methodology and some preliminary results of the project in Digital Humanities *OPhEMA* (Ontology for the Philosophy of Early Middle Ages), promoted by the FiTMU centre and the RAMUS doctoral school of the University of Salerno. The purpose of the project is to establish what is meant by *philosophy* in the Early Middle Ages [2], through the analysis of the literature produced between the II and the XII century<sup>1</sup>.

---

1 <https://www.dmgh.de/>

<https://www.mlat.uzh.ch/MLS/>

The main premise is that, in order to reconstruct any particular meta-philosophical view on the Philosophy of the Early Middle Ages, it is necessary to consider how the term "philosophia" (and related terms such as "philosophus", "philosophi", etc.) occurs in the literature.

The idea is that a survey of the occurrences of "philosoph\*" might help us to define the meaning of *philosophy* for a single author, in a specific period of time, and then – by tracking traces of any change of perspective and recurring definition – in the whole Early Middle Ages.

In order to do this, we plan to develop a specific IT ontology for the Philosophy of the Early Middle Ages [3, 4], which might be suitable to grasp the philosophical knowledge contained in the literature of the field. The scope of this tool is to make available ontologically annotated texts for a variety of scientifically relevant uses, going beyond the sole meta-philosophical browsing: for example, the ontology might support the production of digital and critical editions [4].

The main methodological steps of this project are the following:

- 1 working on a literature review aimed at identifying the occurrences of "philosoph\*", by starting from authors, such as Augustine of Hippo, Gregory the Great, Anselm of Canterbury, as representatives of the area of specialization of the research-team's members,
- 2 creating a shared Google Sheets file, within which all the occurrences of "philosoph\*" are systematized by indicating authors, bibliographic data, translations, comments and URIs,
- 3 revising the structure of such a systematization so as to verify the functionality and completeness of the identified fields,
- 4 combining those data in an IT ontology of the Philosophy of the Early Middle Ages. This might facilitate processes of browsing and searching texts based on internal philosophical and meta-philosophical contents and knowledge.

## REFERENCES

1. Bisogno, Armando. 'Dall'ebook al living document: il ruolo delle Humanities nell'open science'. In *Ebook e Dati Della Lettura*, edited by Maria Senatore Poliseti and Maria Rosaria Califano, Ledizioni. Milano, 2020.
2. D'Onofrio, Giulio. *Storia Della Filosofia Medievale*. Roma: Città Nuova, 2011.
3. Grenon, Pierre, and Barry Smith. 'Foundations of an Ontology of Philosophy'. *Synthese* 182, no. 2 (2011): 185–204.
4. Tambassi, Timothy. 'Intrecci Meta-Ontologici Tra Filosofia e Informatica'. *Giornale Di Metafisica* In stampa (2021).

# Citizen curation and NLP technologies for museums in the SPICE Project

Andrea Bolioli<sup>1</sup>, Alessio Bosca<sup>2</sup>, Anna Maria Marras<sup>3</sup>, Rossana Damiano<sup>4</sup>, Antonio Lieto<sup>5</sup>

<sup>1</sup>CELI - Language Technology - andrea.bolioli [at] celi.it

<sup>2</sup>CELI - Language Technology - alessio.bosca [at] celi.it

<sup>3</sup>Università di Torino - annamaria.marras [at] unito.it

<sup>4</sup>Università di Torino - rossana.damiano [at] unito.it

<sup>5</sup>Università di Torino - antonio.lieto [at] unito.it

## ABSTRACT

In this paper, we briefly present some aspects of the EU H2020 Research & Innovation project “SPICE - Social cohesion, Participation, and Inclusion through Cultural Engagement” (2020-2023), which aims to develop methods and technologies to support citizen curation of museums. In this paper, we will focus on the topic of multilinguality, connected with Natural Language Processing (NLP) and semantic technologies, and the case study located in the Turin Gallery of Modern Art (GAM).

## KEYWORDS

inclusion, museums, natural language processing, citizen curation, user engagement

## POSTER

### 1 INTRODUCTION

The social impact of museums and their role in promoting social inclusion and cohesion is directly linked with their audience engagement. Museums, rather than providing an authoritative view, increasingly attempt to present multiple voices related to their collection and exhibitions, including from the museum visitors themselves, both in presence and at a distance [2]. Multilingual NLP and semantic technologies can facilitate the inclusion of people and accessibility.

### 2 MUSEUM INCLUSION AND CITIZEN CURATION

Digital technologies can support not only the museum visitors' experience but also the remote engagement of people and the inclusion for all. Also due the COVID-19 pandemic, museums have to rethink the use of digital and remote interaction and its crucial role to open not only their collection but also to have an open dialogue with “visitors” [8]. Museum mission is changing and it is changing its role inside the society.



This challenge is difficult and troubled as demonstrated by the still active discussion inside the International Council of Museums (ICOM) community regarding the new proposal for a definition of museum *“Museums are democratising, inclusive and polyphonic spaces for critical dialogue about the pasts and the futures. Acknowledging and addressing the conflicts and challenges of the present, they hold artefacts and specimens in trust for society, safeguard diverse memories for future generations and guarantee equal rights and equal access to heritage for all people...”*<sup>1</sup> This definition includes many aspects and points out also the principles of citizen curation and participation. These elements are crucial also in the Faro convention<sup>2</sup>, recently ratified by Italy, that underlines these important aspects of heritage as they relate to human rights and democracy.

With regard to museums, the term “citizen curation” or “citizen curator” has been used for many years. In 2010 for example, Proctor [8] used it in questioning where the boundaries lie between digital and analog, in-gallery and online, curator, interpreter, and agent of social media. In 2013, Ride [9] involved citizens sharing contributions via Twitter which were later used in a video installation developed by the museum. The term was used in the context of the Paleo project [4], Hill et al. [5] described citizen curation as a process in which citizens with little or no background in museum curation were taught and guided to create their own physical and virtual exhibitions. In the SPICE project, “citizen curation” is intended as a way of supporting visitors, both in presence or at a distance, to share their own interpretations of museum artworks and reflect on the variety of interpretations contributed by others, as described below.

### 3 SPICE PROJECT

The SPICE project brings together 13 partners from 7 countries. The consortium comprises: seven research centres (Universities of Bologna, Aalto, Aalborg, Open University, Complutense University of Madrid, Turin, Haifa) with expertise in codesign, museology, HCI, Linked Data, narratology, ontologies, visualisation and user modelling, three enterprises focused on visitor guides (GVAM), mobile games (PadaOne) and Natural Language Processing (CELI), four heritage institutions (Design Museum Helsinki, Irish Museum of Modern Art, Gallery of Modern Art Turin, Hecht Museum).

The aim of SPICE is to build social cohesion, both between and within citizen groups by developing tools and methods to support citizen curation [2]. We define “citizen curation” as a process in which cultural objects are used as a resource by citizens to develop their own personal interpretations. Those interpretations are then shared and used within and across groups to reflect on similarities and differences in perspective. This development of a shared heritage can strengthen social ties within the group, thereby promoting social cohesion. These intra-group social ties can be described as bonding capital: the strength of social connections within a group.

Interpretations can also be used to support social cohesion across groups, by both promoting tolerance and understanding of their differences and also recognising what they have in common.

---

1 ICOM Museum Definition

<https://icom.museum/en/resources/standards-guidelines/museum-definition/>

2 Council of Europe Framework Convention on the Value of Cultural Heritage for Society (Faro, 27<sup>th</sup> October 2005)  
<https://www.coe.int/en/web/conventions/full-list/-/conventions/treaty/199>

## 4 SEMANTIC TECHNOLOGIES AND NATURAL LANGUAGE PROCESSING

The project will deliver a Linked Data infrastructure that will enable contributions to be shared and analyzed collectively across different authoring and social media platforms in order to avoid a technologically driven fragmentation of contributions. The infrastructure will enable communities to build and control their own perspective on culture and determine how this is shared with others.

Language barriers are a major challenge to overcome, as we have to support all languages spoken or used in case study countries, i.e. English, Italian, Spanish, Finnish, Hebrew and Italian Sign Language (LIS). We will therefore design a solution that provides both a native multilingual backend and high cross-language portability. As for LIS, an interpreter has been involved in the project. In addition to the local language, English will be used as a common language through all the case studies.

We will develop a real-time semantic annotation service for analysing multilingual texts, supporting the official languages of the case studies. The service will perform, via an API, term extraction, named entity recognition, sentiment analysis, emotion recognition, opinion mining, event detection and classification. We will initially train the service on multilingual corpora of relevant contents (e.g. comments, short stories, artwork descriptions) and then we will validate it within the 5 museums case studies. We will carry out the research in the application of neural networks and word embeddings techniques, e.g. BERT [3] and Laser [1], to new multilingual datasets, extending it beyond the state-of-the-art, and we will integrate these technologies with rule-based NLP components and multilingual domain ontologies. An Interlinked Knowledge Graph obtained by aligning and linking different ontologies and data infrastructure will build on existing cultural, narrative, curatorial and social media ontologies, advancing the state of the art through knowledge integration and ontology engineering. In particular, the Knowledge Graph will provide a modular approach for: semantically describing interpretations across different media forms, ranging from collections to social media stories; semantically consuming, storing and reproducing interpretations; integrating interpretations descriptions with information about cultural objects; and associating interpretations with social activities, users and user profiles (e.g. authors, contributors, commenters). The SPICE ontologies will be available as a Linked Data vocabulary in order to promote exploitation and reuse by other institutions and projects. The ontologies, being oriented by design to the creation and reuse of interpretations of cultural objects, will provide an effective way to manipulate and retrieve narrative data by multiple (i.e. narrative, social, cultural, affective) dimensions.

## 5 GAM CASE STUDY

One of the project case studies is carried out by Gallery of Modern Art (GAM) in Turin. Drawing on the existing cooperation with the Turin Institute for the Deaf, GAM will work closely with the Institute to develop interpretation tools specifically tailored to the characteristics of the Deaf.

An app - specifically designed for the case study - will allow visitors to create and share simple stories during and after the visit of the museum, inspired by the exhibits. In this app, the use of text will be kept to a minimum, thanks the emphasis put on the visual features of the artworks and the use of a direct manipulation approach in the interface design, so as to encourage its use by the Deaf.

By leveraging the narrative and emotional representation of the artworks encoded in the project ontologies, the app will assist the users in creating their own stories, suggesting suitable narrative elements and artworks and helping them to explore the stories created by the other users, with the ultimate goal of encouraging the appreciation of diverse perspectives on art.

## 6 ACKNOWLEDGEMENT

SPICE project has received funding from the European Union's Horizon 2020 research and innovation programme (Grant Agreement N. 870811).

## REFERENCES

1. Artetxe, Mikel, and Holger Schwenk. 'Massively Multilingual Sentence Embeddings for Zero-Shot Cross-Lingual Transfer and Beyond'. *Transactions of the Association for Computational Linguistics* 7 (2019): 597–610.
2. Bruni, Luis Emilio, Enrico Daga, Rossana Damiano, Lily Diaz, Tsvi Kuflik, Antonio Lieto, Aldo Gangemi, Paul Mulholland, Silvio Peroni, and Sofia Pescarin. 'Towards Advanced Interfaces for Citizen Curation', 2018.
3. Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 'Bert: Pre-Training of Deep Bidirectional Transformers for Language Understanding'. *ArXiv Preprint ArXiv:1810.04805*, 2018.
4. Ellwood, Elizabeth, Kathryn Estes-Smargiassi, Noel Graham, Gary Takeuchi, Austin Hendy, Molly Porter, and Emily Lindsey. 'Project Paleo: Citizen Curation and Community Science at the Natural History Museum of Los Angeles County'. *Biodiversity Information Science and Standards*, 2018.
5. Hill, Amanda, Mark Kretschmar, David Morton, and Sara Raffel. 'Eenie Meenie Miney Mose: Using Experimental Citizen Curating to Engage Visitors with Racial Ephemera'. *Florida Studies Review*, 2018.
6. ICOM. 'What Definition Do Museums Need?' ICOM, 2020.
7. Najbrt, Lukáš, and Jana Kapounová. 'Categorization of Museum Visitors as Part of System for Personalized Museum Tour'. *International Journal of Information and Communication Technologies in Education* 3 (2014): 17–27.
8. Proctor, Nancy. 'Digital: Museum as Platform, Curator as Champion, in the Age of Social Media'. *Curator: The Museum Journal* 53, no. 1 (2010): 35–43.
9. Ride, Peter. 'Creating# Citizencurators: Putting Twitter into Museum Showcases.' In *Proceedings of the 19th International Symposium of Electronic Art*. Brighton: ISEA, 2013.

# Human digital library for inclusion

Luca Bravi  
Università di Firenze – luc.bravi [at] unifi.it

## SINTESI

Il livello di odio diffuso in Italia verso rom e sinti è tra i più alti in Europa e riguarda l'86% della popolazione del nostro Paese [3]. Il poster intende offrire uno strumento didattico legato all'ambito della Public history of education, in grado di fornire dati corretti rispetto alla presenza ed alla caratterizzazione delle comunità rom presenti nella nostra nazione, ma in particolare capace di costruire spazi di testimonianza diretta che scardinino gli stereotipi massificanti. I documenti in grado di decostruire lo stereotipo collettivo dello “zingaro” sono fatti dialogare con testimonianze orali che compongono una biblioteca digitale umana, fatta di persone appartenenti alle comunità che narrano la propria storia personale, ma anche le vicende d'inclusione/esclusione e di scolarizzazione con l'obiettivo di costruire nuovi quadri di memoria collettiva, privi di facili generalizzazioni e pregiudizi.

## PAROLE CHIAVE

inclusione, public history of education, hate speech, antidiscriminazione, didattica

## POSTER

### 1 STATO DELL'ARTE

Gli elementi offerti dalle analisi dell'Ufficio Nazionale Antidiscriminazioni Razziali (UNAR) confermati da recenti monitoraggi del consiglio d'Europa [8] sottolineano l'imponente costruzione e diffusione mediatica del pregiudizio sugli “zingari”, etichetta denigrante irrobustitasi e mai decostruita dal periodo dei totalitarismi novecenteschi. I dati raccolti dal Ministero dell'interno dimostrano che solo il 20% della popolazione di questo gruppo vive in campi o baraccopoli e che l'80% di essi (sono circa 180 mila sul territorio italiano [14]) è già incluso nel contesto sociale nazionale, ma evita di dichiararsi rom, per evitare di attirare l'odio razzista. Dal punto di vista delle comunità rom, la storia personale, familiare e comunitaria è da sempre parte del racconto interno al gruppo, ma non è mai riconosciuta o conosciuta all'esterno, dalla popolazione maggioritaria, neppure quando si tratti del genocidio subito da questa minoranza nel campo di sterminio di Auschwitz. Il non riconoscimento come attori sociali in Europa riproduce ed alimenta costantemente lo stereotipo dello zingaro nomade ed asociale.<sup>1</sup> La conoscenza che si è sviluppata attraverso specifiche ricerche antropologiche, storiche, sociologiche e peda-

---

1 E. Rizzin (a cura di), *Attraversare Auschwitz*, Gangemi, Roma, 2020

gogiche in ambito accademico non ha ancora raggiunto un livello adeguato di comunicazione, tale da modificare l'opinione pubblica. Il poster intende costruire uno spazio di ascolto della testimonianza diretta come elemento in grado di riattivare la riflessione critica e, indirettamente, decostruire gli stereotipi massificanti.

## **2 L'AMBITO METODOLOGICO**

Dal punto di vista metodologico, il poster propone dati relativi alla presenza delle comunità rom e sinti in Italia secondo le recenti ricerche a livello nazionale ed internazionale.

Utilizza una selezione di fonti orali caricate su un archivio online che possano garantire l'ascolto di storie inattese e lontane dalle stereotipie.

La selezione proposta per il poster sarà composta da 10 interviste che possano rappresentare un estratto rappresentativo delle differenti generazioni di persone appartenenti a queste comunità, tutte nate tra il 1950 ed il 2000.

Le interviste sono state raccolte sulla base di questionari standard che hanno preso in considerazione quattro tematiche: le origini, la storia, la scuola, l'inclusione. Una parte del poster sarà dedicata inoltre ai testimoni della persecuzione nazifascista, facendo riferimento in particolare agli archivi della «Shoah Foundation» e a quelli molteplici luoghi di memoria presenti in Europa. In questo caso saranno evidenziate le «parole dell'esclusione» in modo da individuare continuità e fratture dello stereotipo tra passato e presente.

Precedenti esperienze di raccolta di fonti orali della minoranza rom e sinti sono praticamente assenti nel territorio italiano, in particolare per quanto concerne la specifica attenzione relativa alla narrazione della storia comunitaria nel contesto italiano.

## **3 I RISULTATI ATTESI**

Il poster vuole rappresentare un'esperienza di Public History of Education<sup>1</sup> per attivare un processo di costruzione dal basso delle narrazioni socializzate e socializzanti e per fornire uno strumento di azione educativa e formativa rivolto alla scuola. Il poster ha quindi l'obiettivo di costruire uno strumento utile alla didattica, ma anche ad una conoscenza di elementi inattesi, ma percepibili in modo immediato attraverso una Human digital library: una biblioteca digitale e virtuale di storie di rom e sinti presenti in Italia e appartenenti a differenti generazioni. La testimonianza diretta come fonte orale è qui inserita in un archivio in continua implementazione che possa diventare elemento di conoscenza, d'interrogativi e di confronto con le criticità, volto alla decostruzione dell'immagine stereotipata dello «zingaro».

Gli obiettivi attesi possono essere differenziati secondo il seguente schema:

all'interno della comunità rom e sinti:

- 1 la possibilità di narrare la storia personale, familiare, comunitaria, ma anche nazionale ed internazionale abbandonando il ruolo di problematicità sociale connesso in maniera diffusa all'appartenenza etnica,
- 2 avere l'opportunità di confrontarsi con le risorse storiche per portare a livello cosciente i rimandi con il passato e rintracciando elementi di permanenza del pregiudizio e di superamento dell'odio,
- 3 rielaborare un racconto che inserisca anche i rom all'interno della storia di costruzione delle democrazie in Europa e non soltanto connotato come vittima di persecuzione,

all'interno della società maggioritaria:

- 1 confrontarsi con immagini di rom e sinti lontane dallo stereotipo,
- 2 elaborare una narrazione multifocale della storia italiana ed europea,
- 3 elaborare la relazione con le comunità non soltanto sul piano della problematicità sociale, ma all'interno di una relazione sul piano culturale,
- 4 avere a disposizione risorse utili per la costruzione di percorsi di formazione all'interno dei contesti educativi.

## BIBLIOGRAFIA

1. Bandini, Gianfranco. *Frontiere della formazione postmoderna. Neopragmatismo americano e problemi educativi*. Roma: Armando Editore, 2002.
2. Biscioni, R. *Fotografia e public history*. Pisa: Pacini, 2019.
3. Bonetti, Paolo, A. Simoni, e T. Vitale. *La condizione giuridica di Rom e Sinti in Italia: atti del convegno internazionale, Università degli studi di Milano-Bicocca, 16-18 giugno 2010*. Vol. 1. Milano: Giuffrè, 2011.
4. Bonomo, Bruno. *Voci della memoria. L'uso delle fonti orali nella ricerca storica*. Roma: Carocci, 2014.
5. Bravi, L. *Tra inclusione ed esclusione*. Milano: UNICOPLI, 2009.
6. Bravi, Luca. *Percorsi storico-educativi della memoria europea*. Milano: Franco Angeli, 2014.
7. Colazzo, S., G. Iurlano, e D. Ria. *Public History fra didattica e comunicazione*. Università del Salento, 2019.
8. European Commission. «Rapporto di monitoraggio della società civile sull'implementazione della Strategia nazionale di inclusione Rom, Sinti e Caminanti in Italia». European Commission, 2019.
9. Farnetti, P. B., L. Berucelli, e Botti A. *Public history. Discussioni e pratiche*. Milano: Mimesis, 2017.
10. Oliviero, Stefano, e G. Bandini. *Public History of Education*. Firenze: FUP, 2019.
11. Piasere, Leonardo. *I rom d'Europa. Una storia moderna*. Roma-Bari: Laterza, 2009.
12. ———. *L'Antiziganismo*. Macerata: Quodlibet, 2005.
13. Pontrandolfo, S., e E. Rizzini. *Antiziganismo in Italia 1990/2019. La produzione dell'odio verso rom e sinti nei discorsi politici e nelle rappresentazioni mediatiche dell'Italia contemporanea*. Roma: Cisu, 2020.
14. Presidenza del Consiglio dei Ministri, a c. di. «Strategia Nazionale d'inclusione dei rom, dei sinti e dei caminanti in Italia (2012-2020)». Governo Italiano, 2012.

# Both Digital Edition and Corpus Archive: the works of Kristijonas Donelaitis

Philipp Büch<sup>1</sup>, Mortimer Drach<sup>2</sup>, Jolanta Gelumbeckaitė<sup>3</sup>, Armin Hoenen<sup>4</sup>,  
Adriano Cerri<sup>5</sup>, Vanessa Wadlinger<sup>6</sup>

<sup>1</sup> Emp. Ling. Inst., GU Frankfurt – buech [at] em.uni-frankfurt.de

<sup>2</sup> Emp. Ling. Inst., GU Frankfurt – drach [at] em.uni-frankfurt.de

<sup>3</sup> Emp. Ling. Inst., GU Frankfurt – gelumbeckaite [at] em.uni-frankfurt.de

<sup>4</sup> Emp. Ling. Inst., GU Frankfurt – hoenen [at] em.uni-frankfurt.de

<sup>5</sup> Università di Pisa

<sup>6</sup> Emp. Ling. Inst., GU Frankfurt

## ABSTRACT

In this paper and on the poster we present the CorDon project and its web-presence<sup>1</sup>.

## KEYWORDS

donelaitis, online edition, corpus, old Lithuanian

## POSTER

### 1 INTRODUCTION

Subject of the project are the texts in metrotonic hexameters of the Lithuanian poet Kristijonas Donelaitis (\*1714 - †1780), whose works have been linguistically deeply annotated in this and previous projects. The works consist of 4 texts from the cycle *Metai* (The Seasons), 2 associated fragments, and 6 fables (*Pasakos*). These had to be checked, analyzed, and made available on the web.

### 2 ANNOTATIONS, EDITIONS AND TRANSLATIONS

Donelaitis' texts, comprising around 30,000 tokens with around 9,500 types, are the most extensive autochthonous Old Lithuanian texts. They have been annotated with approx. 330,000 annotations on 11 annotation layers: *transliteration*, *modernized form*, *lemma*, *accented lemma*, *language*, *part of speech lemma*, *part of speech form*, *morphology lemma*, *morphology form*, *inflection*, *gloss*. In addition to the representation of these base texts and their annotations, the project also provides two other early parallel Lithuanian text editions, two early translations into German, and two modern translations into Italian and English.

---

1 <http://titus.fkidg1.uni-frankfurt.de/cordon/docs/posterAIUCD2021Pisa.pdf>



### 3 FUNCTIONALITIES, EXPECTATIONS, USERS

The website is primarily designed to serve the needs of researchers and teachers of (Old) Lithuanian, linguists and Indo-Europeanists as well as interested laypersons. The texts were digitized according to the principles of diplomatic literal transcription. The annotation was carried out using the program *Toolbox*<sup>2</sup>, and further revised in the annotation software ELAN<sup>3</sup> [1] which then exports the data to a table-based HTML structure, easily customizable through some JavaScript functionality so as to dynamically toggle any annotation layer. This potentially generic way to create a web corpus had been provided to us by P. Hinkelmanns (Salzburg) who developed the concept for the *Lesekorpus Altdeutsch* (Old High German Reading Corpus) [2].<sup>4</sup> Within CorDon, we included a facsimile (as [3, p.27] discusses, this is expected for digital editions) and the Lithuanian editions and/or all translations (parallel text edition). Using a database connected to the web interface, we provide a comprehensive search function with a PDF export so as to search for specific occurrences or patterns within the texts. The search mask is GUI-based and has 2 modes: basic and extended, the latter allowing the specification of any annotation configuration for up to 5 tokens in a row. The same data ordered in a different way is presented as a lexicon/concordance and augmented by additional information such as each lemma's neighbors in lemma vector space. A subset of this data is online as RDF using OntoLex-Lemon. Finally, we provide some visualizations such as the ubiquitous word cloud as well as some more analytical types.

### 4 ACKNOWLEDGEMENTS

We would like to thank the Fritz Thyssen Stiftung for their financial support.

### REFERENCES

1. Brugman, Hennie, Albert Russel, and Xd Nijmegen. 'Annotating Multi-Media/Multi-Modal Resources with ELAN.' In *LREC*, 2004.
2. Mittmann, Roland, and Ralf Plate. 'Das ,Referenzkorpus Altdeutsch 'als Lesekorpus. Grammatisch Annotierte Und Mit Wörterbüchern Verknüpfte Texte Für Lehre Und Selbststudium'. *Das Mittelalter* 24, no. 1 (2019): 173–187.
3. Sahle, Patrick. 'What Is a Scholarly Digital Edition'. *Digital Scholarly Editing: Theories and Practices*, 2016, 19–39.

---

2 <https://software.sil.org/toolbox>

3 <https://archive.mpi.nl/tla/elan>

4 <http://titus.uni-frankfurt.de/lea>

# DNT: un Corpus Diacronico e Multigenere di Testi in Lingua Inglese

Tommaso Caselli<sup>1</sup>, Rachele Sprugnoli<sup>2</sup>

<sup>1</sup> University of Groningen – t.caselli [at] rug.nl

<sup>2</sup> Università Cattolica del Sacro Cuore – rachele.sprugnoli [at] unicatt.it

## SINTESI

In questo contributo presentiamo il Diachronic News and Travel (DNT) corpus, una collezione di testi in lingua inglese che copre diversi generi testuali e due ampi periodi temporali. Più precisamente, il corpus è formato da testi sia contemporanei che storici di tre generi, ovvero articoli di quotidiano, resoconti di viaggio e guide turistiche. Il corpus, liberamente disponibile, è già stato usato per l'analisi di vari fenomeni linguistici e per lo sviluppo di sistemi per il Trattamento Automatico del Linguaggio che vengono qui brevemente descritti.

## PAROLE CHIAVE

corpus, annotazione, trattamento automatico del linguaggio

## POSTER

### 1 INTRODUZIONE

Lo sviluppo e la diffusione di corpora e di risorse linguistiche ha visto una rapida crescita a partire dalla prima metà degli anni 2000. La *Language Resource and Evaluation Map*<sup>1</sup> riporta la documentazione di oltre 2,920 corpora in diverse lingue ma questa stima, basata su uno sforzo volontario della comunità, è destinata a crescere.

Sebbene la disponibilità di corpora sia stata favorita dal crescente sviluppo di tecnologie per il Trattamento Automatico del Linguaggio (TAL), un aspetto centrale dei corpora è quello di essere istanze di *performance* linguistiche [3]. La combinazione di disponibilità e uso di corpora ha introdotto un profondo cambiamento negli studi sul linguaggio, agevolando la verifica empirica di teorie e fenomeni linguistici [11, 2, 5].

---

1 [http://lremap.elra.info/?&selected\\_facets=resourceTypeFilter\\_exact%3ACorpus](http://lremap.elra.info/?&selected_facets=resourceTypeFilter_exact%3ACorpus)

Questo contributo introduce il Diachronic News and Travel (DNT) corpus, composto da testi scritti in lingua inglese. La sua caratteristica distintiva è la copertura sia di diversi generi testuali (i.e., articoli di quotidiano e scrittura di viaggio, quest'ultima a sua volta divisa tra resoconti in prima persona e guide turistiche) sia di diversi periodi temporali (i.e., 1860-1939 e 1998-2017).

## 2 DESCRIZIONE DEL CORPUS

La combinazione di periodi temporali e generi testuali diversi permette di dividere il corpus in 6 sotto-corpora le cui dimensioni in termini di numero di documenti e di token sono riportate nella Tabella 1.

| Genere e Periodo Temporale       | # Documenti | # Token |
|----------------------------------|-------------|---------|
| Articoli di quotidiano 1998-2017 | 84          | 32.086  |
| Articoli di quotidiano 1860-1939 | 50          | 29.717  |
| Resoconti di viaggio 1998-2017   | 23          | 30.747  |
| Resoconti di viaggio 1860-1939   | 25          | 31.690  |
| Guide turistiche 1998-2017       | 58          | 29.950  |
| Guide turistiche 1860-1939       | 39          | 29.327  |
| TOTALE                           | 279         | 183.517 |

Tabella 1: statistiche generali sul corpus

Nel creare la risorsa, si è cercato di bilanciare la dimensione dei vari sotto-corpora anche se un bilanciamento perfetto non è stato possibile a causa della difficoltà nel reperire sia testi storici puliti da eventuali errori dovuti all'OCR che testi contemporanei liberi da copyright. Per quanto riguarda questi ultimi, gli articoli di giornale sono stati estratti da altre risorse linguistiche (come il Wall Street Journal Corpus [6]), i resoconti di viaggio da post pubblicati online su blog di viaggiatori mentre i testi delle guide derivano dal portale WikiTravel<sup>2</sup> e dal sito della Lonely Planet<sup>3</sup>: per ciascun post e per le pagine della Lonely Planet abbiamo ottenuto il permesso di riuso. Relativamente ai testi storici, invece, gli articoli di giornale sono stati presi dal portale Wikisource<sup>4</sup> mentre resoconti di viaggio e guide sono stati scaricati dal sito del Progetto Gutenberg<sup>5</sup>.

## 3 CASI D'USO

Il DNT è stato usato prevalentemente per lo sviluppo di tecnologie di TAL. Questo ha di fatto imposto sullo stesso insieme di dati, o su un suo sottoinsieme, annotazioni manuali su diversi livelli, offrendo la possibilità di studiare l'interazione di fenomeni linguistici all'intersezione delle dimensioni temporali e/o di genere. Di seguito presentiamo brevemente tre casi di studio che mostrano come DNT possa essere utile per diverse analisi e applicazioni.

---

2 [https://wikitravel.org/en/Main\\_Page](https://wikitravel.org/en/Main_Page)

3 <https://www.lonelyplanet.com/>

4 [https://en.wikisource.org/wiki/Main\\_Page](https://en.wikisource.org/wiki/Main_Page)

5 <http://www.gutenberg.org/>

Il più complesso tra i casi di studio riguarda l'annotazione di micro-atti illocutori a livello di proposizioni realizzati tramite tipi di contenuto (o content types)<sup>6</sup> [12]. A differenza di lavori precedenti [4] [10], un unico schema di annotazione è stato elaborato e applicato sull'intero corpus. L'analisi delle annotazioni ha evidenziato una distribuzione statisticamente significativa dei diversi tipi di contenuto nei diversi generi, ma non nei diversi periodi temporali. Al contrario, l'applicazione di un sistema supervisionato ha evidenziato come una certa differenza nella distribuzione dei tipi di contenuto sembra essere presente anche per la dimensione diacronica [8].

Gli altri due casi riguardano invece l'analisi dei soli testi storici. Lo studio della realizzazione linguistica degli eventi [1] ha visto l'integrazione di più schemi di annotazione in un unico modello e la sua applicazione ai vari generi presenti nel corpus. Questo ha evidenziato una distribuzione di classi di eventi diversi statisticamente significativa nelle scritture di viaggio e negli articoli di giornale. Un sistema supervisionato è poi stato addestrato sui dati annotati per il riconoscimento e la classificazione semantica degli eventi in testi storici<sup>7</sup> [9].

Infine, lo studio di entità nominate si è concentrato su guide e resoconti di viaggio<sup>8</sup>. L'applicazione di un sistema TAL allo stato dell'arte<sup>9</sup> ha messo in evidenza i limiti dell'applicazione di queste tecnologie a dati storici e/o generi diversi da quelli usati per sviluppare il sistema. In particolare, la presenza di fenomeni specifici quali variazioni ortografiche di nomi propri, frequenza di grafemi latini, uso di forme abbreviate ha richiesto l'annotazione manuale di nuovi testi e lo sviluppo di un nuovo sistema per il riconoscimento automatico di nomi propri di luoghi geografici (e.g., fiumi, monti), politici (e.g., città, nazioni) e funzionali (e.g., monumenti, siti archeologici) [7].

## 4 CONCLUSIONI

Nell'ottica di favorire e promuovere una cultura e una scienza aperta, la condivisione di dati e la replicabilità degli esperimenti, tutte le risorse descritte in questo contributo sono rilasciate liberamente. Nello specifico, il DNT corpus in formato testo è disponibile online<sup>10</sup> come lo sono anche le varie annotazioni prodotte nei casi d'uso sopra menzionati ed i modelli di Trattamento Automatico del Linguaggio sviluppati. Una versione tokenizzata del corpus sarà presto rilasciata per facilitare l'applicazione di altri livelli di annotazione da parte della comunità e promuovere lo sviluppo di un'ampia risorsa aperta e multi-livello.

---

6 <https://github.com/tommasoc80/ContentTypes>

7 <https://github.com/dhfbk/Histo>

8 <https://github.com/dhfbk/Detection-of-place-names-in-historical-travel-writings>

9 <https://stanfordnlp.github.io/CoreNLP/>

10 [https://osf.io/g6w4e/?view\\_only=1d67431f4b0345f780d0e4c0b370d94a](https://osf.io/g6w4e/?view_only=1d67431f4b0345f780d0e4c0b370d94a)

## BIBLIOGRAFIA

1. Bach, Emmon. «The algebra of events». *Linguistics and philosophy* 9, n. 1 (1986): 5–16.
2. Chiari, Isabella, e Elisabetta Jezek. «Dati Empirici e Risorse lessicali». *Ricognizioni. Rivista di Lingue e Letterature straniere e Culture moderne* 2, n. 6 (2016): 9–11.
3. Chomsky, Noam. *Syntactic structures*. The Hague/Paris: Mouton, 1957.
4. Guo, Yufan, Anna Korhonen, Maria Liakata, Ilona Silins, Lin Sun, e Ulla Stenius. «Identifying the information structure of scientific abstracts: an investigation of three different schemes». In *Proceedings of the 2010 Workshop on Biomedical Natural Language Processing*, 99–107, 2010.
5. Marotta, Giovanna. «Strutture linguistiche e dati empirici in diacronia e sincronia». 2017.
6. Paul, Douglas B., e Janet Baker. «The design for the Wall Street Journal-based CSR corpus». In *Speech and Natural Language: Proceedings of a Workshop Held at Harriman, New York, February 23-26, 1992*, 1992.
7. Sprugnoli, Rachele. «Arretium or Arezzo? a neural approach to the identification of place names in historical texts». In *Fifth Italian Conference on Computational Linguistics (CLiC-it 2018)*, 360–365. Accademia University Press, 2018.
8. Sprugnoli, Rachele, Caselli Tommaso, Tonelli Sara, e Moretti Giovanni. «The Content Types Dataset: a New Resource to Explore Semantic and Functional Characteristics of Texts». In *15th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2017*, 2:260–266. Association for Computational Linguistics, 2017.
9. Sprugnoli, Rachele, e Sara Tonelli. «Novel event detection and classification for historical texts». *Computational Linguistics* 45, n. 2 (2019): 229–265.
10. Teufel, Simone, e Marc Moens. «Summarizing scientific articles: experiments with relevance and rhetorical status». *Computational linguistics* 28, n. 4 (2002): 409–45.
11. Voghera, Miriam. *Teorie linguistiche e dati di parlato*. A cura di F. Albano Leoni, E. Stenta Krosbakken, R. Sornicola, e C. Stromboli. Roma: Bulzoni, 2001.
12. Werlich, Egon. *A text grammar of English*. Quelle & Meyer, 1976.

# MUTANT: MULTImodAl, TrAcked aNd parTecipated e- learning

Valentina Cesari<sup>1</sup>, Angelo Gemignani<sup>2</sup>, Danilo Menicucci<sup>3</sup>

<sup>1</sup> Università di Pisa – valentina.cesari [at] phd.unipi.it

<sup>2</sup> Università di Pisa – angelo.gemignani [at] unipi.it

<sup>3</sup> Università di Pisa – danilo.menicucci [at] unipi.it

## ABSTRACT

MULTImodAl, TrAcked aNd parTecipated e-learning (MUTANT) is a concept prototype for optimize skills acquisition via e-learning platforms taking advance from three essential psychological assumptions: multimodal learning to enhance sense of presence, eye-tracking to detect fatigue/cognitive load, interactive learning to encourage participated decision making.

## KEYWORDS

e-learning, eye-tracking, multimodal integration, participated learning

## POSTER

### 1 INTRODUCTION

E-learning is the unique alternative to guarantee the continuation of educational services due to COVID-19 pandemic restrictions. However, it has some limitations: it requires motivation and time management abilities to face absence of personal relations and remoteness, it is not effective when hands-on practical exercises are required [2], it leads to excessive fatigue during demanding and complex teaching [5].

The implementation of MUTANT could overcome these limitations, since it is anchored to strong assumptions.

### 2 BEHIND MUTANT

1. Multimodal. The enhancement of external perception *via* multimodal integration is important for cognition and thinking [1]. Skills acquisition could be optimized adopting embodiment-based learning (e.g by integrating haptic feedbacks), since individuals experience a higher control of supplementary sensory information as their own, thus increasing the sense of agency [3].

2. Tracked. Accumulating evidences indicate that eye tracking is a reliable method to detect signs of cognitive load and fatigue, that often prevent skills acquisition [4].

Parallely, it provides a real-time feedback to the user. Individuals could be trained to detected signs of “nonoptimal state” by receiving, initially, feedbacks signaling a state of fatigue/overload. For enhancing self-control and state awareness, individuals will recognize non-optimal state without being informed by feedbacks.

3. Participated. A virtual environment in which individuals could manipulate the same object by receiving tactile feedback could augment the sense of group presence. Individuals will be also stimulated with group decision-making learning tasks, in which super-ordinate goals will be reached after a process of exchange and integrate information.

MUTANT will take advantage from internet-based learning to allow small groups to interact real-time via haptic interface in conjunction with visual/auditory one, while subjects wear non-intrusive glasses for eye-tracker. Educator (i.e., teacher) will supervise the entire sessions promoting activities which stimulates group decision-making.

MUTANT will optimize intellectual and perceptual-motor skills during e-learning sessions, offering the opportunity to minimize side effects related to the absence of physical environment, thus guarantying an optimal learning process.

### 3 ACKNOWLEDGEMENTS

Progetti di Ricerca di Ateneo 2020-2021 - Servizi pubblici e piattaforme digitali. Categorie giuridiche, processi, garanzie (PRA\_2020\_73)

### REFERENCES

1. Anastopoulou, Stamatina, Mike Sharples, e Chris Baber. «An evaluation of multimodal interactions with technology while learning science concepts». *British Journal of Educational Technology* 42, n. 2 (2011): 266–290.
2. Arkorful, Valentina, e Abaidou Nelly Abaidou. «The Role of E-Learning, the Advantages and Disadvantages of Its Adoption in Higher Education». *International Journal of Education and Research* 2, n. 12 (2014): 397–404.
3. Caspar, Emilie A., Axel Cleeremans, e Patrick Haggard. «The relationship between human agency and embodiment». *Consciousness and cognition* 33 (2015): 226–236.
4. El Haddioui, Ismail. «Eye Tracking Applications for E-Learning Purposes: An Overview and Perspectives». In *Cognitive Computing in Technology-Enhanced Learning*, 151–174. IGI Global, 2019.
5. Pimenta, André, Sergio Gonçalves, Davide Carneiro, Florentino Fde-Riverola, José Neves, e Paulo Novais. «Mental workload management as a tool in e-learning scenarios». In *2015 International Conference on Pervasive and Embedded Computing and Communication Systems (PECCS)*, 25–32. IEEE, 2015.

# DEMOTICON. Per un'edizione semantica dei Malavoglia

Salvatore Cristofaro<sup>1</sup>, Christian D'Agata<sup>2</sup>, Antonio Di Silvestro<sup>3</sup>,  
Giuseppe Palazzolo<sup>4</sup>, Pietro Sichera<sup>5</sup>, Daria Spampinato<sup>6</sup>

<sup>1</sup>CNR-ISTC – salvatore.cristofaro [at] istc.cnr.it

<sup>2</sup>Università di Catania – christian.dagata [at] gmail.com

<sup>3</sup>Università di Catania – silvanto [at] unict.it

<sup>4</sup>Università di Catania – giuseppe.palazzolo [at] unict.it

<sup>5</sup>CNR-ISTC – pietro.sichera [at] istc.cnr.it

<sup>6</sup>CNR-ISTC – daria.spampinato [at] cnr.it

## SINTESI

Questo contributo descrive l'idea di un prototipo di ambiente web integrato per la fruizione delle opere letterarie di Giovanni Verga, in primo luogo dei Malavoglia, e dei documenti epistolari e d'archivio (anche inediti) connessi con la produzione letteraria dell'autore. Si propone un'organizzazione dei contenuti che coniughi i modelli concettuali della TEI e delle ontologie formali, integrando la concezione ordinata e gerarchica (OHCO) del testo con l'organizzazione semantica a grafo dei contenuti, senza ridondanza dei dati. L'acronimo DEMOTICON [Digital Editions, Multimedia (and) Ontologies (for) Texts (of) Italian Culture On (line)] vuole rappresentare la sinergia tra edizioni digitali, contenuti multimediali e ontologie in funzione di una fruizione «aumentata» dei testi letterari nell'ambiente digitale.

## PAROLE CHIAVE

edizioni scientifiche digitali, ontologie, didattica della letteratura, Giovanni Verga

## POSTER

### 1 INTRODUZIONE

Il progetto propone la realizzazione di un'edizione digitale del capolavoro di Verga, *I Malavoglia*, attraverso la pubblicazione di testi codificati in XML/TEI, integrati con basi di conoscenza ontologiche RDF/OWL per la rappresentazione dei concetti semantici e delle relazioni con le entità nominate (persone, luoghi, eventi). La complessa articolazione dei materiali preparatori che fa da sfondo alla scrittura dei *Malavoglia*, con un rilievo particolare dato alla descrizione dei personaggi (comprensiva di una breve presentazione dei tratti fisici, caratteriali e degli eventi di cui sono protagonisti) e alla lista dei proverbi presenti nel manoscritto verghiano, ha infatti suggerito lo sviluppo di un prototipo che integrasse coerentemente la codifica TEI e l'ontologia finalizzato all'arricchimento della fruizione e dell'interrogazione del testo.



Attualmente non esistono sul web né edizioni filologicamente sicure dei *Malavoglia*, né risorse utilizzabili per navigazioni non sequenziali e ricerche testuali. Peraltro è disponibile una enorme quantità di materiali, sovente inediti, conservata soprattutto in archivi e biblioteche catanesi (manoscritti delle opere, epistolari, documenti familiari, libri della biblioteca personale), scarsamente digitalizzata e non consultabile via web. D'altronde, le biblioteche digitali non offrono un valido modello per la rappresentazione di materiali di questo tipo, in quanto o presentano semplicemente i testi in vari formati ma senza strumenti di interrogazione, oppure offrono una modalità di ricerca elementare. L'obiettivo che dunque ci si propone è quello di fornire diversi livelli di rappresentazione dei testi, integrati da una serie di strumenti e risorse digitali utilizzabili da studiosi, docenti e studenti delle scuole e delle università italiane, ma anche da un pubblico di lettori comuni e di appassionati dell'opera verghiana.

## 2 ORGANIZZAZIONE CONCETTUALE

La codifica del testo dei *Malavoglia* ha dovuto affrontare in prima battuta la complessità dei rapporti tra i materiali preparatori (proverbi, liste dei personaggi) e il testo stesso. Rispettando l'ordine dato nel manoscritto, a livello di codifica TEI, si è deciso di inserire nel *front* tutti i materiali preparatori e nel *body* il testo, individuando con tag specifici le occorrenze dei proverbi, dei personaggi, dei luoghi e dei discorsi diretti. È opportuno notare come la codifica in TEI dei materiali preparatori e del testo rispetti la forma filologicamente definita dall'edizione di riferimento senza esplicitare le varie connessioni, che invece verranno demandate all'ontologia *demoticon* che si sta sviluppando. Così, ad esempio, nel caso dei proverbi, nel *front* viene inserita la codifica del proverbio nella sua *forma originale*, in siciliano, così come puntualmente riportata nei materiali preparatori verghiani e, all'uopo, la codifica della variante tradotta dello stesso proverbio in altra lingua, ogniqualevolta quest'ultima è anch'essa contenuta nei materiali preparatori (più specificamente, viene codificata la lista ordinata di proverbi citati dal Pitrè, motivo per cui si è scelto di usare il tag *cit* per l'annotazione dei proverbi contenuti nei materiali preparatori). Si osservi che il Verga riporta nel testo del romanzo la traduzione in italiano dei proverbi, talvolta anche di proverbi non contenuti nei materiali preparatori. La complessità di relazioni tra le diverse varianti e traduzioni dei proverbi utilizzati da Verga nel romanzo viene catturata nell'ontologia attraverso apposite proprietà (*role*) tra le istanze di classe che rappresentano le differenti formulazioni intralinguistiche e interlinguistiche dei proverbi stessi.

L'ambiente software sarà sviluppato in conformità con gli standard del W3C per assicurare la persistenza dei dati e l'interoperabilità in accordo con i paradigmi del Web Semantico. Verranno integrati e riutilizzati strumenti open source standard per la codifica, la visualizzazione e l'organizzazione dei dati. Secondo le buone pratiche dei Linked Open Data verranno implementati collegamenti a vocabolari LOD sul web (Geonames, VIAF) e riutilizzati schemi ontologici consolidati, come il CIDOC CRM.

Per entrare nel merito, attraverso l'esempio di un proverbio presente nel paragrafo 5 del primo capitolo, *il motto degli antichi mai menti*, è possibile vedere come l'occorrenza del proverbio nel testo venga segnalata con il tag *distinct* (Figura 1), in quanto occorrenza di un proverbio linguisticamente distinto rispetto alla formulazione verghiana, e l'attributo *type* (per indicare che si tratta appunto di un proverbio così come definito nell'ontologia *demoticon*). Inoltre viene usato il tag *ref* per referenziare le URI delle particolari istanze di classe che rappresentano le versioni del proverbio, ossia istanza di *tipo A* per l'originale in siciliano del proverbio, istanza di *tipo B* per la variante italiana presente nella lista preparatoria, e istanza di *tipo C* per la traduzione da A (operata da Verga) presente nel testo (Figura 2).

```

<front>
  <div type="Personaggi_lavori_preparatori_Manuscritto">
    <listPerson> [...] </listPerson> </div>
    <div type="Proverbi_lavori_preparatori_Manuscritto">
      [...]
      <cit type="demoticon:proverbio" subtype="A" xml:id="P1_A" source="#Pitrè" ana="1">
        <ref target="URI_istanza_proverbio_P1_A">
          <quote>Lu muttu di l'anticu mai minti</quote>
        </ref></cit>
      <cit type="demoticon:proverbio" subtype="B" xml:id="P1_B" source="#Pitrè" ana="1">
        <ref target="URI_istanza_proverbio_P1_B">
          <quote>Proverbio non falla</quote>
        </ref></cit>
      [...]
    </div>
  </front>
</body>
  [...]
  <p xml:id="GV_MA_C1_P5" n="cap1_p5">
    <persName ref="#P_Ntoni"> <roleName>Padron</roleName> <addName>'Ntoni</addName></persName>
    sapeva anche certi <emph>motti</emph> e proverbi che aveva sentito dagli <emph>antichi
    </emph>, <q who="#P_Ntoni">
      <distinct type="demoticon:proverbio" subtype="C" xml:id="P1_C">
        <ref target="URI_istanza_proverbio_P1_C">
          «perché il motto degli antichi mai menti»
        </ref></distinct></q>
    [...]
  </p>
  [...]

```

Figura 1: parte della codifica TEI dei Malavoglia relativa ai proverbi.

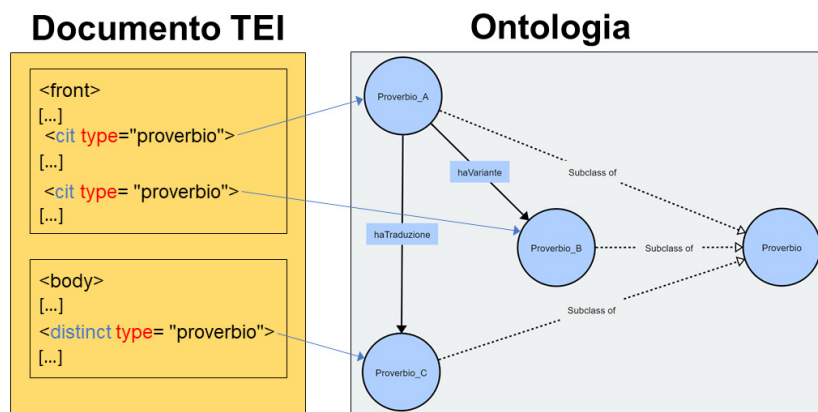


Figura 2: rappresentazione del modello di conoscenza relativa ai proverbi.

## BIBLIOGRAFIA

1. Alfieri, Gabriella. *Il motto degli antichi. Proverbio e contesto nei "Malavoglia"*. Catania: Fondazione Verga, 1985.
2. Cristofaro, Salvatore, e Daria Spampinato. «Una proposta di ontologia basata su RDA per il patrimonio culturale di Vincenzo Bellini». *Proceedings of AIUCD2020*, 2020, 82–88.
3. Daquino, Marilena, Francesca Giovannetti, e Francesca Tomasi. «Linked Data per le edizioni scientifiche digitali. Il workflow di pubblicazione dell'edizione semantica del quaderno di appunti di Paolo Bufalini». *Umanistica Digitale*, n. 7 (18 dicembre 2019). <https://doi.org/10.6092/issn.2532-8816/9091>.
4. Grosso, Angelo Mario Del, Erica Capizzi, Salvatore Cristofaro, Maria Rosa De Luca, Emiliano Giovannetti, Simone Marchi, Graziella Seminara, e Daria Spampinato. «Bellini's Correspondence: A Digital Scholarly Edition for a Multimedia Museum». *Umanistica Digitale*, n. 7 (18 dicembre 2019). <https://doi.org/10.6092/issn.2532-8816/9162>.
5. Italia, Paola. *Editing Duemila*. Roma: Salerno, 2020.
6. Subialka, Michael. «The future of Pirandello: On the New Digital Edition of Pirandello's Opera omnia». *The Journal of The Pirandello Society of America XXXII* (2019).
7. Verga, Giovanni. *I Malavoglia. Edizione critica a cura di F. Cecco*. Novara: Interlinea, 2014.

# Madeleine in Biblioteca

## Un laboratorio digitale di ricordi e storie di lettura

Monica Dati  
Università di Firenze – monica.dati [at] unifi.it

### SINTESI

Il presente contributo vuole essere un resoconto del progetto Madeleine in biblioteca, volto al recupero di memorie e ricordi in relazione al molteplice e variopinto rapporto che le persone hanno con la lettura, dall'infanzia all'età adulta. Svolto in collaborazione con la Biblioteca Civica di Lucca, il percorso ha previsto anche la successiva realizzazione dell'omonimo sito web<sup>1</sup>, uno spazio digitale con diverse finalità: promuovere innanzitutto l'importanza della lettura e riflettere sulla sua storia, restituire alla collettività le testimonianze raccolte, infine ricostruire il passato attingendo dalle narrazioni storiche dei non specialisti secondo un approccio di public history. In questo contesto le biblioteche possono svolgere un ruolo centrale nel collegare mondo accademico e comunità, presentandosi non solo come luoghi dove raccogliere e conservare libri ma centri di esperienza e conoscenza al servizio del pubblico.

### PAROLE CHIAVE

public history, storia digitale, biblioteche, storia della lettura

### INTERVENTO

La lettura, nella sua affascinante complessità, è una delle pratiche su cui si fondano aspetti centrali della nostra tradizione culturale, della nostra vita emotiva ed intellettuale e il campo degli studi relativi alla sua storia è estremamente ampio, articolato, diversificato, ricco di fonti documentarie estremamente eterogenee. Tra queste rivestono un ruolo non secondario le testimonianze autobiografiche dei lettori, le loro esperienze intime, uniche e soggettive, le loro storie individuali che partecipano di una storia collettiva possono metterne in luce gli aspetti meno indagati e trascurati, quelli che hanno a che fare con i sentimenti, le emozioni, le spinte ideali, i sogni. Da qui l'idea di mettersi in ascolto e di prestare attenzione alle parole dei lettori così come suggerito da Michele Petit che da decenni intervista uomini e donne di tutte le estrazioni sociali e di aree geografiche diverse e da Alberto Manguel che in suo famoso volume afferma: “la storia della lettura è la storia di ciascun lettore”.

---

1 [www.madeleineinbiblioteca.it](http://www.madeleineinbiblioteca.it)

Con questa semplice affermazione l'autore argentino vuole sottolineare come tante piccole storie di apprendimento informale vadano a comporre in realtà un mosaico più grande, una galleria che incede di pari passo con la Storia e parla di promozione sociale, banchi di scuola, biblioteche private, ricordi di infanzia e famiglia, di libri amati ed odiati, letti di nascosto.

È in questo contesto, brevemente delineato nelle sue caratteristiche principali, che si colloca il progetto Madeleine che ha visto gli utenti della biblioteca donare storie e racconti da restituire alla comunità per promuovere la riflessione individuale e collettiva secondo un approccio di public history, generata prima di tutto dal desiderio del pubblico di partecipare alla scrittura, documentazione, discussione sul passato. In quest'ottica il coinvolgimento della Biblioteca civica- un luogo dove cultura, tempo libero, scambi umani ed educazione permanente si intrecciano- è stato imprescindibile per evidenziare come i saperi accademici possano essere effettivamente messi in contatto con molteplici contesti educativi per creare nuove proficue relazioni tra istituzioni e soggetti diversi.

Con quali modalità? Per rendere effettivo lo scambio e il coinvolgimento del pubblico e garantire un'efficace restituzione della memoria oltre all'organizzazione di seminari dedicati alla storia della lettura e costruiti con le fonti orali raccolte l'attenzione è stata posta sullo spazio offerto dalla Rete. terminate le prime quaranta interviste, effettuate tramite registrazioni audio, si è pertanto proceduto all'ideazione e progettazione di un sito web<sup>2</sup>, collegato ad una pagina Instagram per allargare il bacino di partecipazione e condivisione digitale dell'esperienza e avvalersi delle potenzialità offerte dalla digital history. Quest'ultima si rivela una prospettiva particolarmente congeniale agli studi sulle memorie educative soprattutto nel favorire un legame tra la ricerca accademica, i luoghi di apprendimento ed i contesti di vita. Un approccio utilissimo ad incorporare storie dal basso nella narrazione storica con modalità non sempre possibili nella stampa, nelle esposizioni e ovviamente nei manuali scolastici che si avvale delle nuove tecnologie di comunicazione, aiutando a dare parola non solo ai singoli ma anche ad associazioni e gruppi.

Il sito, realizzato tramite la piattaforma WordPress e indirizzato a tutti coloro che amano la lettura e quanti siano interessati agli aspetti educativi ad essa connessi, ruota principalmente intorno a due sezioni: la prima, "Storie di Lettura" raccoglie le testimonianze degli utenti della biblioteca volte a far comprendere come accanto ad una più generica ed ampia Storia della lettura esista un articolato mondo fatto di tante piccole storie, soggettive ed uniche, la seconda "Appunti di Storia" si propone invece di esplorare la storia sociale del libro, i contesti sociali della lettura, i luoghi di apprendimento informale come le biblioteche e i circoli, il mondo e le categorie dei lettori, i libri proibiti e censurati. Il repository che prevede la condivisione dei contenuti con altri utenti, una barra di ricerca per parole chiave e un form di contatto per partecipare, ha come intento non solo la piena e completa disponibilità online delle testimonianze ma anche l'interazione, lo scambio e il dibattito, a tal proposito è stato inserito un plugin per aggiungere la relativa pagina Instagram "Madeleine.Lab"<sup>3</sup>. L'integrazione con questo social network, basato notoriamente sulle immagini e contenente numerose pagine dedicate al mondo dei libri, ha reso possibile anche la raccolta di materiale fotografico postato e condiviso dagli utenti, arricchendo le fonti a nostra disposizione e la narrazione sulla storia della lettura.

---

2 [www.madeleineinbiblioteca.it](http://www.madeleineinbiblioteca.it)

3 [www.instagram.com/madeleinelab/?igshid=12id233z41u82](https://www.instagram.com/madeleinelab/?igshid=12id233z41u82)

Quest'ultima all'interno del sito si sviluppa attraverso una pluralità di tematiche: dai ricordi di infanzia alla letteratura di genere, dal ruolo della famiglia e dei genitori a quello della scuola, spingendosi ad indagare anche il tema della lettura di nascosto, oggetto recente di un seminario realizzato interamente grazie alle fonti orali raccolte ("Libri clandestini: letti di nascosto censurati proibiti. Un excursus storico").

Concludendo l'esperienza vuole essere un primo tentativo di applicare i principi della public history nel suo incontro con la Storia Digitale ad un terreno inesplorato come la storia della lettura e del libro. Mettere in pratica un dialogo tra interlocutori diversi per stabilire una conoscenza diffusa e condivisa del passato, capace altresì di valorizzare la storia e l'importanza di "pensare storicamente", ritenuto un pensare utile per tutta la collettività. In questo processo le biblioteche possono svolgere un ruolo centrale, presentandosi non solo come meri luoghi di prestito e richiesta informazioni ma centri di esperienza e conoscenza al servizio del pubblico. Uno spazio dove agiscono processi formativi di carattere formale, non formale ed informale e dove possono essere realizzati anche progetti di public history la cui demarche prevede un forte coinvolgimento delle persone, della collettività, e di tutto il loro patrimonio di ricordi, immagini, oggetti: un contesto privilegiato per attivare momenti e processi costruttivi di discussione, partecipazione, approfondimento e riflessione su discorsi storico-educativi, compreso il mondo della lettura e la sua storia.

Alla Biblioteca Civica Agorà di Lucca, uno spazio libero ed aperto ad una pluralità di soggetti ed iniziative, vanno i nostri ringraziamenti per aver accolto questa ricerca e progetto.

## BIBLIOGRAFIA

1. Bandini, Gianfranco, e Stefano Oliviero. *Public History of Education: riflessioni, testimonianze, esperienze*. Firenze: Firenze University Press, 2019.
2. Cauvin, Thomas. *Public history: a textbook of practice*. Routledge, 2016.
3. Cavallo, Guglielmo, e Roger Chartier. *Storia della lettura nel mondo occidentale*. Roma: Laterza, 1995.
4. De Sanctis, Filippo Maria, e Paolo Federighi. *Pubblico e biblioteca*. Roma: Bulzoni, 1981.
5. Federighi, Paolo. *Le condizioni del leggere*. Milano: Il cantiere biblioteca, 1996.
6. Lyons, Martyn. *A history of reading and writing in the western world [Book Review]*. London: Palgrave Macmillan, 2010.
7. Manguel, Alberto. *Una storia della lettura*. Milano: Feltrinelli, 2010.
8. Petit, Michele. *Elogio della lettura*. Milano: Ponte alle grazie, 2010.
9. Salvatori, Enrica. «Digital (Public) History: la nuova strada di una antica disciplina». *RiMe. Rivista dell'Istituto di Storia dell'Europa Mediterranea (ISSN 2035-794X)*, 2017, 57–94.
10. Sayer, Faye. *Public history: a practical guide*. London: Bloomsbury Publishing, 2015.

# Political variants mining: computational investigations on authorial variants

Università degli studi di Napoli "L'Orientale" - mdeblasi [at] unior.it  
Università di Bologna - annasofia.lippolis [at] studio.unibo.it  
Università di Bologna -gabriella.totaro [at] studio.unibo.it

## ABSTRACT

How can critical Digital Editions act on the community in the construction of collective ideas and thoughts? The representation and study of Authorial corrections, usually applied to literary texts, becomes of crucial collective interest when involves controversial works of modernity, such as Pietro Verri's Observations on Torture. Thanks to the potential of the Digital Edition, which makes it possible to represent the change of the Author's will, self-censorship or the hands of different authors, the project of Political Variants Mining applies the following analysis to a political text. It provides scholars with statistics on corrections, automatically computed starting from the TEI encoding of the text and made easily understandable through data visualization, allowing us to find out not only what has changed in the rewriting of a great political text but also to extend the resulting methodology to other works.

## KEYWORDS

authorial philology, authorship, digital humanities, TEI encoding, text mining

## POSTER

### 1 INTRODUCTION

Political Variants Mining has the double purpose to participate and to involve the largest number of people in the historical debate on torture, exploiting the new possibilities offered by innovative digital technologies and methodologies on the basis of the publicly accessible Digital Edition Observations on Torture by P. Verri, developed by M. De Blasi, and the qualitative and quantitative analysis of its corrections.

In the era of integrated communication and collective knowledge construction, the new possibilities offered by the digital paradigm emphasize the dynamic character of literary texts and their social value.

Authorial Philology shares the same conception of texts as in fieri works making it possible to provide new critical interpretations through variants that denote changes in the author’s will, self-censorship or the hands of different authors. Being openly available resources, Scholarly Digital Editions have therefore become fundamental in shaping collective ideas and thoughts, especially when it comes to political texts that have had a huge impact over the centuries.

## 2 HISTORY OF THE OBSERVATIONS ON TORTURE

Despite the several editions available, it was only in 2018 that the *Observations on Torture* has been published with an apparatus that included authorial variants.

As a result, Verri’s work offered the possibility to study authorial corrections under an innovative perspective with respect to the stylistic one of a traditional literary text: the choice to reduce his own explicit considerations for the mere presentation of facts, in order to guide the reader’s point of view through the way the events unfolded, represents a clear example of Verri’s self-censorship.

## 3 RESULTS OF THE ANALYSIS

With Political Variants Mining, we make available an innovative technique able to deepen the conventional philological reading, making it viable to gain and display in an automatic way linguistic and empirical insight into the creative process of an Author directly either from the PDF version of the text or from its encoding in an ad hoc TEI model. As shown in Fig.1, more than half of Verri’s corrections are late variants (additions, written above, overwritten, derived) thus witnessing the aim of decreasing his own voice to obtain an unbiased report.

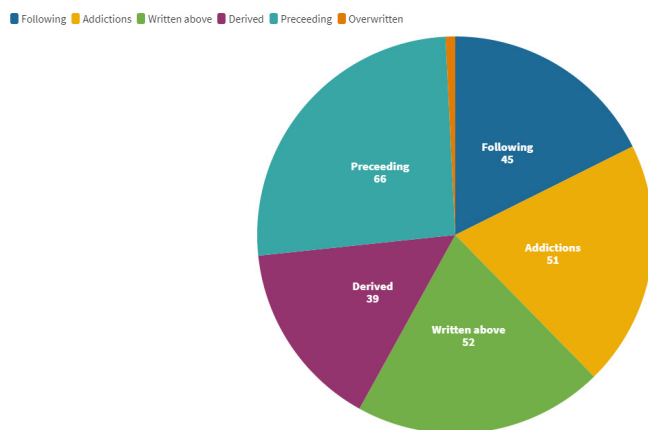


Figure 1: observations on torture frequency of correction

The technical immediacy of the model allows to widen the study to other Authors to make a complete evidence-based analysis of the types of Authorial corrections and compare these different conceptions of writing statistically. Therefore, an open source detailed guide to achieve the same results and the code will be made available in a dedicated website.

In fact, Political Variants Mining exploits new technologies both as means of information representation and as methodological frameworks to widen traditional philology to new perspectives of analysis, creating new possibilities towards the social democratization of knowledge.

## 4 CONCLUSION

Political Variants Mining makes evident Verri's idea in its very developmental process and highlights the importance of the conservation of a foundational text whose contribution was essential to raise social awareness on such a noteworthy topic.

## REFERENCES

- Nava, Beatrice. *Varianti politiche d'autore: da Verri a Manzoni*. Bologna: Pàtron editore, 2019.
- Rosselli Del Turco, Roberto. «EVT-Edition Visualization Technology». Laboratorio di Cultura Digitale Università di Pisa, 2015. <http://www.labcd.unipi.it/progetti/evt-edition-visualization-technology/>.
- Verri, Pietro. *Osservazioni sulla tortura*. A cura di Margherita De Blasi. Roma-Padova: Editrice Antenore, 2018.
- . «Osservazioni sulla tortura». In *Osservazioni sulla tortura*, a cura di Margherita De Blasi, 2020. [www.edizionicritiche.it/verrievt](http://www.edizionicritiche.it/verrievt).



# Verso la descrizione automatica delle immagini nell'editoria digitale accessibile: proposta di una tassonomia di immagini per gli algoritmi di IA

Chiara De Martin<sup>1</sup>, Barbara Leporini<sup>2</sup>, Gregorio Pellegrino<sup>3</sup>

<sup>1</sup>Università di Pisa – c.demartin1 [at] studenti.unipi.it

<sup>2</sup>ISTI-CNR – barbara.leporini [at] isti.cnr.it

<sup>3</sup>Fondazione LIA - gregorio.pellegrino [at] fondazionelia.org

## SINTESI

In questo contributo viene proposta una tassonomia di possibili tipologie di immagini da poter utilizzare per migliorare la composizione dei dataset di addestramento degli algoritmi di intelligenza artificiale da applicare ai task di classificazione e descrizione automatica delle immagini. Viene altresì sinteticamente illustrato il processo di validazione effettuato nello studio. 25 categorie sono state identificate nella tassonomia proposta.

## PAROLE CHIAVE

editoria digitale accessibile, accessibilità immagini, descrizione alternativa immagini, tassonomia immagini, intelligenza artificiale, Fondazione LIA

## POSTER

### 1 INTRODUZIONE

Il problema dell'accesso ai contenuti non testuali è molto sentito nella comunità delle persone con disabilità. L'accesso alle immagini da parte di utenti che usano uno screen reader è stato risolto dalle linee guida sull'accessibilità con l'adozione di descrizioni alternative del contenuto degli oggetti non testuali, che rappresentano un requisito fondamentale per l'accessibilità di qualunque pubblicazione digitale. Descrivere le immagini è un'attività qualificata, costosa e lunga e si configura come uno dei requisiti più difficile da soddisfare. Una possibile soluzione è quella di automatizzare o semi-automatizzare il processo tramite l'intelligenza artificiale [5]. Pertanto, proponiamo una tassonomia per la classificazione delle immagini, con lo scopo di aiutare gli algoritmi a produrre descrizioni efficaci in base al tipo di immagine.

Per ogni categoria della tassonomia vengono proposti degli attributi che ne descrivano le possibili caratteristiche di forma e di contenuto.

## 2 METODOLOGIA

Per definire la tassonomia sono state consultate le principali linee guida internazionali sulla descrizione alternativa delle immagini<sup>1</sup>, [3] e [4]. Le categorie da queste analizzate sono state combinate con quelle emerse dall'esame di metodi manuali, semi-automatici e automatici per la descrizione alternativa delle immagini. Sono stati esaminati allo scopo anche tool e strumenti online dedicati al task<sup>2</sup>, banche dati di immagini<sup>3</sup> e dataset gratuiti<sup>4</sup>. La tassonomia è stata ottenuta combinando le categorie prese in considerazione dalle diverse fonti e basandosi sulle caratteristiche associate a ciascuna di esse per definire i relativi attributi e i rispettivi possibili valori.

La tassonomia è stata validata su alcuni manuali di editoria scolastica, la metodologia applicata è stata di tipo qualitativo. La validazione è stata condotta cercando di collocare ogni immagine contenuta nei manuali in una categoria e di descriverla mediante gli attributi corrispondenti. Nuove categorie e attributi sono stati aggiunti per colmare lacune e inadeguatezze.

## 3 PROPOSTA DI UNA TASSONOMIA

La tassonomia prevede venticinque categorie: fotografie, disegni, dipinti, fumetti/comic, cartoon, mappe, grafici, diagrammi, tabelle, timeline, cover, icone, loghi, bandiere, firme, testo / testo riempito con immagini, typeface, screenshot, poster, immagini complesse / infografica / organigrammi, meme, segnaletica, pubblicità, collage, murales. Alcune categorie, come grafici o diagrammi, comprendono sottocategorie (grafici a barre, a linea, a torta, ...).

Per ogni categoria sono proposti degli attributi, formulati come campi per i quali vengono previsti una serie di possibili valori. La selezione degli attributi è stata guidata dall'obiettivo di catturare nel modo più dettagliato possibile gli aspetti rilevanti che possono contraddistinguere le immagini delle varie categorie. Alcuni attributi sono sprovvisti di valori predefiniti, poiché la loro variabilità è tale da rendere impossibile fornire a priori delle indicazioni.

---

1 <http://diagramcenter.org/table-of-contents-2.html>

2 Come <https://poet.diagramcenter.org/>

3 iStock by Getty Images

<https://www.istockphoto.com/it>

Pixabay

<https://pixabay.com/it/>

Internet Archive, "Community Images"

[https://archive.org/details/opensource\\_image](https://archive.org/details/opensource_image)

Flickr

<https://flickr.com>

4 MSCOCO

<http://cocodataset.org>

Flickr8k Dataset

<http://academictorrents.com>

Flickr30k

<https://www.kaggle.com/hsankesara/flickr-image-dataset>

PASCAL

<http://host.robots.ox.ac.uk/pascal/VOC/>

Alcuni attributi sono di carattere generale e applicabili a più di una tipologia di immagini, altri sono più specifici e hanno un'applicabilità più ristretta. Alcuni valori svolgono a loro volta il ruolo di attributi: in questo caso, il loro valore è rappresentato da un'altra etichetta, o da una combinazione di più etichette.

Nella tassonomia sono inclusi due attributi di validità globale: “rilevanza dell'immagine” e “tipo di descrizione”. Il primo serve per tenere conto del fatto che un'immagine può essere, nel contesto in cui è inserita, *significativa* o *non significativa* e dunque avere bisogno o meno di essere descritta, il secondo specifica il tipo di descrizione più adatto all'immagine: questa può essere breve o lunga, ma composta di solo testo, oppure lunga e strutturata, richiedendo una formattazione che prevede l'uso di liste o tabelle.

## **4 LA VALIDAZIONE**

La tassonomia è stata validata su quattro libri di testo scolastici, di materie diverse: biologia, comunicazione, tecnologia, storia. Per ogni immagine, si è cercato di classificarla utilizzando le categorie individuate e di descriverla usando i campi degli attributi. Nuove categorie sono state aggiunte per coprire immagini che altrimenti non avrebbero trovato un'etichetta adeguata (come “emoticon” o “francobollo”), nuovi attributi sono stati aggiunti per raffinare le categorie esistenti.

Il riscontro ottenuto in fase di validazione si è rivelato un ottimo feedback per valutare possibili modifiche future.

## **5 CONCLUSIONI**

In questo contributo si propone una possibile tassonomia per il raffinamento della composizione dei dataset per l'addestramento degli algoritmi di intelligenza artificiale, con l'obiettivo di contribuire a un'editoria digitale sempre più accessibile, con particolare riferimento al tema della descrizione alternativa delle immagini. La tassonomia proposta include venticinque categorie. La validazione condotta ha dimostrato come la tassonomia sia in grado di descrivere un'elevata percentuale di immagini ed elementi non testuali. Si ritiene che la stessa possa costituire un valido punto di riferimento per la composizione dei dataset, portando in questo modo a un miglioramento delle prestazioni e dell'accuratezza degli algoritmi nei task di classificazione e descrizione automatica delle immagini.

## **6 RINGRAZIAMENTI**

Si ringrazia l'editore Zanichelli per aver messo a disposizione il suo catalogo digitale per validare la tassonomia.

## BIBLIOGRAFIA

1. Amornchat, Supada. «Complex Images for All Learners», s.d. <https://www.pcc.edu/instructional-support/wp-content/uploads/sites/17/2017/11/complex-images.pdf>.
2. Gleason, Cole, Amy Pavel, Xingyu Liu, Patrick Carrington, Lydia B. Chilton, e Jeffrey P. Bigham. «Making Memes Accessible». In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility*, 367–76. Pittsburgh PA USA: ACM, 2019. <https://doi.org/10.1145/3308561.3353792>.
3. NCAM. «Effective Practices for Description of Science Content within Digital Talking Books». Foundation, 2008. <https://www.wgbh.org/foundation/ncam/guidelines/effective-practices-for-description-of-science-content-within-digital-talking-books>.
4. NWEA. «NWEA Image Description Guidelines for Assessments». Northwest Evaluation Association, 2016. [https://www.nwea.org/content/uploads/2016/10/Image-Description-Guidelines-IDGA\\_June2016.pdf](https://www.nwea.org/content/uploads/2016/10/Image-Description-Guidelines-IDGA_June2016.pdf).
5. Pellegrino, Gregorio. «E-book. Automatizzare le descrizioni alternative delle immagini con l'AI». *Giornale della Libreria*, luglio 2019. <https://www.giornaledellalibreria.it/news-innovazione-e-book-automatizzare-le-descrizioni-alternative-delle-immagini-con-lai-3877.html>.

# La Filologia come sistema dinamico: qualche considerazione preliminare

Riccardo Del Gratta<sup>1</sup>, Federico Boschetti<sup>2</sup>, Angelo Mario Del Grosso<sup>3</sup>, Luigi Bambaci<sup>4</sup>

<sup>1</sup>ILC-CNR - riccardo.delgratta [at] ilc.cnr.it

<sup>2</sup>ILC-CNR & VeDPh - federico.boschetti [at] ilc.cnr.it

<sup>3</sup>ILC-CNR - angelo.delgrosso [at] ilc.cnr.it

<sup>4</sup>Università di Bologna - luigi.bambaci2 [at] unibo.it

## SINTESI

In questo articolo introduciamo un approccio formale all'evoluzione dei documenti con particolare attenzione alla prospettiva filologica e alle problematiche tipiche connesse. Proponiamo un modello/framework matematico in grado di formalizzare diversi fenomeni complessi in vari ambiti di ricerca quali la Linguistica Computazionale, la Filologia Digitale e l'Ingegneria del Software, in particolare quando questa viene applicata all'analisi di documenti e testi di interesse storico-letterario.

## PAROLE CHIAVE

approccio evolucionistico, modello formale, analisi documentale e testuale, sistema dinamico, filologia computazionale

## POSTER

### 1 INTRODUZIONE

I diversi tentativi fatti di formalizzazione dei metodi della Filologia testuale (o critica testuale) sono condizionati dalle specificità dei diversi domini di applicazione (Filologia Classica, Biblica, Romanza etc.). Il metodo del Lachmann è un chiaro esempio di questo problema: essendo modellato principalmente sulla tradizione dei testi greci e latini, esso risulta difficilmente applicabile al di fuori del dominio della filologia classica [7].

Risulta importante, quindi, promuovere una visione unitaria della Filologia testuale che evidenzi, da una parte, i fenomeni comuni alle diverse tradizioni testuali (es. fenomeni di corruzione involontaria o meccanica) e tenga conto, dall'altra, di fenomeni che si presentano diversamente in ciascun dominio, ma che sorgono da abitudini scribali comuni (innovazioni volontarie quali rielaborazioni esegetiche e stilistiche, interpolazioni dettate da motivazioni morali, teologiche e così via).

Il processo di formalizzazione non deve essere limitato alle entità, alle proprietà e alle relazioni che caratterizzano il testo filologico, ma deve anche tenere conto degli attori responsabili della creazione e della trasmissione di tali testi, siano essi tool automatici oppure agenti umani come autori, editori, scribi, traduttori, etc.

In questo articolo introduciamo un approccio formale all'analisi del testo prendendo in considerazione alcuni dei problemi menzionati sopra. In particolare, tenteremo di formalizzare il fenomeno della variazione testuale nel tempo attraverso un approccio di tipo evolucionistico, dove per tale approccio intendiamo esattamente il processo dinamico che lega un documento a un altro, o due differenti versioni dello stesso documento.

## 2 UN SISTEMA DINAMICO, IN BREVE

Il più semplice sistema dinamico è quello della dinamica classica, formalizzato da Newton nei suoi Principia Matematica. Secondo tale sistema, in breve, se si conoscessero tutte le forze e tutti i dettagli di un sistema ad un dato istante  $t_0$ , allora si conoscerebbe il sistema nel suo complesso a qualsiasi istante  $t$ . Occorre notare che il presupposto è che, in ambito classico, descrivere un sistema significa descriverne ogni sua singola parte.

Nella sezione 3, definiamo alcune entità che ci aiutano a vedere alcuni aspetti della Filologia (ma anche della Linguistica Computazionale) come un sistema dinamico. Vedremo, nel corso dell'articolo, fino a che punto il parallelismo con un sistema dinamico classico sia pertinente.

## 3 IL MODELLO PROPOSTO, ALCUNE DEFINIZIONI UTILI

La prima definizione che proponiamo è quella di documento. La definizione proposta è coerente con quelle usate dai principali linguaggi di programmazione [1] a vari livelli di astrazione [2, 4, 6].

Per agevolare la comprensione da parte del lettore, parleremo di documenti testuali *elettronici*. Tale definizione di documenti e testi proposta è utile per la Filologia Digitale e Computazionale poiché modella aspetti noti quali le varianti, le congetture, la ricostruzione delle relazioni tra le fonti, ecc.

Un documento (testuale elettronico)  $D$  è una tripla  $c, f, \{p\}$ :

$$D = D(c, f, \{p\}) \quad (1)$$

dove  $c$  è il contenuto informativo trasportato dal documento,  $f$  è il formato e  $\{p\}$  è un insieme di informazioni paratestuali aggiuntive<sup>1</sup>. Sia  $c$  che  $\{p\}$  possono essere vuoti, questo per consentire la possibilità del documento vuoto<sup>2</sup>. Per esempio, se  $D$  è un documento di testo piano<sup>3</sup> contenente una sequenza di caratteri a formare una frase,  $c$  è la frase,  $f$  è il formato plain text e  $\{p\}$  è vuoto. In documenti più complessi (file XML-TEI, per esempio) la separazione tra  $c$ ,  $f$  e  $\{p\}$  non è così immediata. Anche se non saranno usati in questo articolo, è utile introdurre la definizione di documenti cartesiani.

---

1 Le parentesi  $\{ \}$  indicano il fatto che il paratesto  $p$  sia un insieme di informazioni. Dove è scritto  $\{p\}$  è da intendersi  $\{p_1, p_2, \dots, p_n\}$ .

2 Il documento vuoto deve essere supposto per permettere la chiusura del sistema formale rispetto alle operazioni (definite dopo). Un'operazione, per esempio un tool di estrazione, può non estrarre niente. Il suo risultato è appunto un documento vuoto.

3 Creato con editor testuali come notepad, ad esempio. Oppure con vim su sistemi Linux.

Un documento si dice cartesiano se è separabile nelle sue parti costituenti:

$$D = D(c, f, \{p\}) = c \times f \times \{p\} \quad (2)$$

intendendo con  $X$  il prodotto cartesiano tra le varie parti. In un certo senso, nei documenti cartesiani o separabili,  $c$ ,  $f$  e  $\{p\}$  sono tre entità indipendenti. Il senso della (2) è che  $D$  è identificato dalle proprie parti, esattamente come un punto in un piano cartesiano lo è dalle proprie coordinate.

Molte delle cose che diremo nel presente contributo sono indipendenti dal fatto che  $D$  sia cartesiano o meno, dato che lavoriamo ad un livello di astrazione più alto.

Una ulteriore sotto-definizione importante è quella di documento esteso. Un documento esteso è un documento “accompagnato” da altre informazioni:

$$D_{ex} = D(c, f, \{p\}) \times \{m\} \quad (3) \square$$

$\{m\}$  può essere il periodo storico, l'autore, un riferimento a una versione precedente ecc. Nel caso di documenti digitali  $\{m\}$  può contenere, ad esempio, il nome del file e dove risiede fisicamente.

La seconda definizione è quella di azione o operazione. Una operazione  $op$  è “qualcosa che elabora un insieme di documenti e restituisce un singolo documento”<sup>4</sup>. Ovvero:

$$op(\{D_1, D_2, \dots, D_n\}) = \{D_a\} \quad (4)$$

Il senso della (4) è il seguente:  $op$  lavora su  $n$  documenti in input e crea un nuovo documento  $D_a$  in output. In caso di documenti cartesiani, anche le operazioni che agiscono su questi si considerano cartesiane:

$$op = op_c \times op_f \times op_{\{p\}} \quad (5)$$

con l'ovvio significato di operazioni specifiche per parti specifiche: ad esempio  $op_c$  agisce solo su  $c$ , lasciando le altre inalterate. È come se  $op_c$ ,  $op_f$  e  $op_{\{p\}}$  agissero parallelamente su un documento  $D$ , ma la loro azione fosse limitata alla parte di competenza<sup>5</sup>.

Occorre aggiungere che il nostro approccio sarà quello di considerare  $D_a$  sempre un documento nuovo<sup>6</sup>. Consideriamo come  $D_0$  il solito documento in formato plain text di cui sopra, supponiamo poi che  $D_0$  contenga un insieme di frasi. Se  $op$  è, per esempio, l'aggiunta di una frase,  $D_1$  è il risultato di  $op(D_0) = D_1$ . Anche se fisicamente  $D_1$  è lo stesso file<sup>7</sup>, per il modello proposto  $D_0$  e  $D_1$  sono logicamente due entità distinte.

Possiamo introdurre il concetto di “evoluzione” di un documento. Per evoluzione si intende una struttura a grafo i cui nodi sono i documenti e i cui archi le operazioni, vedi Figura 1.

---

4 Ovvero un insieme con un solo elemento.

5 Questo aspetto è molto importante all'interno delle cosiddette “teorie dei processi”, ma lo è anche all'interno dell'ingegneria del software. Un software disegnato per gestire documenti cartesiani ha aspetti diversi da uno nato per gestire documenti non cartesiani. Anche se i due software hanno lo stesso scopo.

6 Approccio funzionale.

7 Ovvio in caso di file testuali presenti su pc.

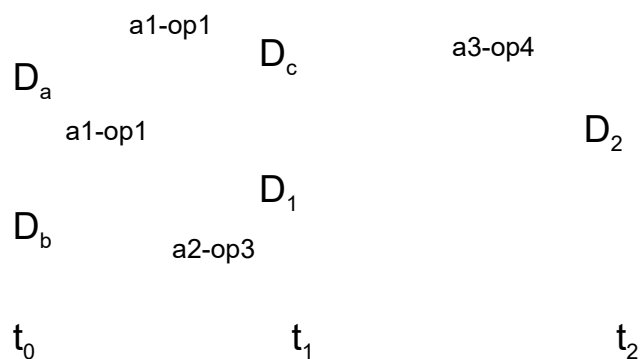


Figura 1. Evoluzione di differenti documenti a diversi istanti temporali

Le operazioni sono azioni astratte che vengono rese possibili attraverso l'intervento di agenti o attori ( $ac$ ). Un attore ( $ac$ ) è un tool automatico o un agente umano che esegue una o più operazioni. In Figura 1 le etichette degli archi stanno esattamente a indicare l'azione congiunta di un dato agente con una certa operazione. A titolo esemplificativo:  $a1-op1$  significa che l'agente ( $a1$ ) ha effettuato  $op1(D_a, D_b)$  per ottenere  $D_c$ .

Un grafo come quello in Figura 1 può essere applicato ad un processo di Linguistica Computazionale, dove da un insieme iniziale di documenti si passa a un output finale attraverso stati intermedi. In questo caso gli attori sono agenti automatici che effettuano operazioni di elaborazione automatica.

Nella stessa figura abbiamo messo una freccia del tempo per enfatizzare come da un gruppo di documenti iniziale il sistema evolva, attraverso documenti intermedi, verso un documento finale. Questa evoluzione ha una certa durata

$$\Delta t = t_2 - t_0 \quad (6)$$

In ottica di ricostruzioni dei rapporti tra le fonti, in Filologia, è molto utile permettere dei *salti all'indietro*, ovvero poter sapere, ad esempio all'istante  $t_1$ , cosa abbia reso possibile ottenere  $D_c$ .

Questo si può affrontare se ipotizziamo di poter effettuare delle "fotografie" del sistema in certi istanti. Sempre dalla Figura 1, abbiamo 3 istanti  $t_0, t_1$  e  $t_2$ . Consideriamo i documenti che abbiamo a disposizione a questi istanti:

$$B(t_0) = \{D_a, D_b\}, B(t_1) = \{D_c, D_1\}, B(t_2) = \{D_2\} \quad (7)$$

Più genericamente, la (7) si può scrivere come:

$$B(\tau) = \{D_1, D_2, \dots, D_n\} \quad (8)$$

La (8) si legge "all'istante  $t = \tau$  ci sono  $n$  documenti  $D_1, D_2, \dots, D_n$  disponibili.

Chiamiamo  $B(\tau)$  spazio di lavoro o base-space a  $t = \tau$ .

Si noti che gli spazi di lavoro nella (7) sono in qualche modo collegati. Infatti,  $D_c$  in  $B(t_1)$  dipende, attraverso  $a1-op1$ , da entrambi i documenti in  $B(t_0)$  mentre  $D_1$  solo da  $D_b$ . Similmente per  $B(t_2)$  e  $B(t_1)$ .

Definiamo lo spazio di evoluzione  $H$  come una collezione di evoluzioni che collegano documenti appartenenti a diversi spazi di lavoro<sup>8</sup>.

8 Ovvero a diversi istanti temporali.



Sempre dalla solita Figura 1, uno spazio di evoluzione tra  $t_0$  e  $t_1$  è il seguente:

$$ev_1 = \{ \dot{b} \}$$

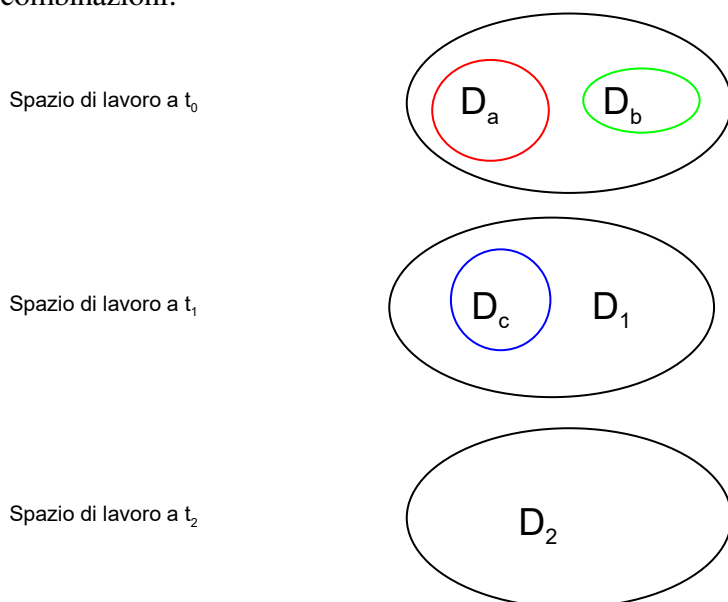
Gli oggetti  $ev$  sono “mappe vincolate” nel senso che di  $ev$  sono noti i valori iniziali e finali, il valore iniziale è un sottoinsieme anche proprio dello spazio di lavoro a un istante  $t=t_0$ , il valore finale è un documento sottoinsieme dello spazio di lavoro a  $t=t_1$ ,  $D_x \subset B(t_1)$ .

$$ev(t_0) = \{ D_0, \dots, D_k \} \subseteq B(t_0); ev(t_1) = D_x \subset B(t_1) \quad (9)$$

La (9) ci dice che se un ipotetico studioso, a  $t=t_1$ , sapesse con certezza come si è arrivati a  $D_c$  partendo da  $D_a, D_b$ , ovvero conoscendo l'operazione a1-op1, sarebbe in grado di replicare il processo che da  $B(\dot{b})$  ha portato a  $D_c \subset B(t_1)$ .

Sfortunatamente, questo è un caso molto raro e spesso si possono solo supporre sia le fonti che le operazioni che da esse hanno portato al documento in esame. Nel caso di  $D_1$ , la Figura riporta un solo arco da  $D_b$  a  $D_1$  attraverso a2-op3. Un lavoro di ricostruzione potrebbe supporre la conoscenza di  $B(\dot{b})$  ma non di ogni singola evoluzione  $ev$  che da  $B(\dot{b})$  ha portato a  $D_1$ , in questo caso il processo è irripetibile e solo speculativo<sup>9</sup>.

La Figura 2 mostra le relazioni tra spazi di lavori e spazi di evoluzione. I diversi colori identificano le differenti combinazioni.



**Figura 2: Rappresentazione grafica delle relazioni tra spazi di lavoro (in nero) e spazi di evoluzione (in diversi colori)**

<sup>9</sup> Ovvero, se a  $t=t_1$ , fosse noto solo lo spazio di lavoro  $B(t_0)$  ma non lo spazio di evoluzione, si potrebbe ipotizzare l'esistenza di una operazione  $op1(D_a, D_b) \rightarrow D_1$ , che è diversa da  $a2-op3(D_b) \rightarrow D_1$ . In questo senso, il processo è speculativo.

Definiamo lo spazio totale  $E$  come prodotto cartesiano tra uno spazio di lavoro e lo spazio delle storie evolutive:

$$E = B \times H \quad (10)$$

Lo spazio totale mette in relazione ogni documento in uno spazio di lavoro con le sue storie evolutive. Gli elementi di  $E$  sono coppie documento-evoluzioni: per esempio la coppia  $\{D_1, e v_3\}$ . Come si evince dalla Figura 2, il processo di “raggruppamento”<sup>10</sup> è iterabile: da  $D_2$  mi collego a  $D_c$  e da  $D_c$  a  $D_a$  e  $D_b$ .

## 4 EVOLUZIONE E DINAMICA

Proviamo a proiettare il modello proposto all’interno di un sistema dinamico. Focalizziamoci su un caso semplice:

$$D_0 \rightarrow D_1(11)_{\square}$$

Come variabile dinamica prendiamo il solo contenuto  $c$ , ovvero si suppone che né  $f$  né  $\{p\}$  cambino. Un’operazione  $op(D_0) \rightarrow D_1$  manda  $c_0$  in  $c_1$ :

$$op(c_0) \rightarrow c_1(12)$$

Dove  $op$  è l’analogo delle forze in un sistema dinamico. Come le forze sono responsabili di “far muovere” un sistema da uno stato ad un altro, così le operazioni “fanno muovere” i documenti. Ci dobbiamo chiedere se sia possibile o meno scrivere delle equazioni (differenziali) che modellino esattamente l’evoluzione di  $c$ . Chiaramente la risposta alla domanda non è unica, dato che dipende da  $op$ . Un esempio semplice: se  $op$  è **togli K parole da c**, e  $c$  contiene 100 parole, sappiamo esattamente calcolare  $c$  dopo l’applicazione di  $op$ .<sup>11</sup>

Estrarre gli eventi da un dato testo è un ulteriore esempio, che può essere modellato come:

$$\{D_0, D_e\} \rightarrow D_1(13)_{\square}$$

Dove  $D_0, D_e$  sono rispettivamente il documento da cui si devono estrarre gli eventi e un documento che contiene la lista degli eventi.  $D_1$  conterrà la lista di eventi estratti. Se l’agente che esegue  $op$  è un agente automatico, è possibile sapere il numero di eventi in  $D_1$ , dati  $D_e$  e  $D_0$ , e, al variare di  $D_e$  e  $D_0$ , si può sempre studiare come varia la lista in  $D_1$ . Ma se l’agente è umano questo non è sempre vero. Un agente umano, per distrazione, può saltare uno o più eventi. Questo semplice esempio mostra che quando l’agente è umano il modello deterministico dinamico non è applicabile. Torniamo alla (11) e supponiamo che  $op$  sia una interpretazione di un testo. Possono verificarsi due situazioni notevoli. La prima è che se due agenti diversi interpretano lo stesso testo, il risultato sarà (molto probabilmente) diverso. La seconda è che se lo stesso agente interpreta lo stesso testo a distanza di tempo, anche in questo caso, il risultato potrebbe essere diverso. Noi riteniamo che un aspetto interessante di questo modello risieda proprio nel descrivere il comportamento di un agente umano, come spieghiamo nelle conclusioni.

---

10 Raggruppamento è la traduzione di bundle. E bundle è la teoria matematica (topologica) da cui questa parte del modello deriva.

11 Per i curiosi l’equazione differenziale è  $dc/dt = -Kc$  che permette di calcolare il numero di parole in entrambe le direzioni temporali.

## 5 CONCLUSIONE E LAVORI FUTURI

In questo articolo abbiamo proposto un modello di gestione dell'evoluzione dei documenti e abbiamo provato a proiettarlo su un sistema dinamico classico, evidenziandone i limiti di applicazione, soprattutto quando l'agente è umano. Questo, crediamo, non è un punto debole, ma piuttosto un di forza del modello proposto. L'idea di base è quella di modellare il più possibile le parti meccaniche di un processo (variazioni di formato, aggiunta di livelli paratestuali, variazioni del contenuto informativo etc.), ma, allo stesso tempo, di predisporre un apparato matematico complesso<sup>1</sup> (e i bundles -cf. nota 10- lo sono) che funga da base e cornice a processi più complicati come quelli che richiedono ricostruzioni, replicabilità etc. Gli spazi di lavoro così come quelli di evoluzione e il loro raggruppamento (bundle) servono a questo scopo. Nell'ultima parte della Sezione 4 abbiamo accennato a due situazioni interessanti, entrambe collegate all'agente umano.

Un'ulteriore linea di ricerca riguarda lo studio del contesto esterno e della conoscenza di un agente umano sul "prodotto" di un processo. Nell'esempio dell'interpretazione di un testo fatta da due agenti diversi a e b, noi postuliamo che la differenza del risultato (la differenza è meccanicamente misurabile) sia una misura dell'impatto del contesto e della conoscenza specifica dei due agenti nel processo interpretativo. Anche nel caso di interpretazioni fatte dallo stesso agente, la differenza delle interpretazioni misura l'impatto del contesto e della conoscenza nel processo interpretativo, ma aggiunge anche una co-evoluzione del processo interpretativo e del rapporto agente-contesto/conoscenza. Una interpretazione cambia perché è cambiato il rapporto dell'agente con il contesto esterno ed è variata la sua conoscenza del fenomeno.

Tutto questo è, a nostro avviso, affascinante e merita ulteriori approfondimenti e sperimentazioni.

## BIBLIOGRAFIA

1. Boschetti, Federico, e Angelo Mario Del Grosso. «TeiCoPhiLib: A library of components for the domain of collaborative philology». A cura di Arianna Ciula e Fabio Ciotti. *Journal of the Text Encoding Initiative*, n. 8 (2014).
2. Carrano, Frank. *Data structures and abstractions with Java*. 3rd ed. USA: Prentice Hall Press, 2011.
3. Coecke, Bob, Edward Grefenstette, e Mehrnoosh Sadrzadeh. «Lambek vs. Lambek: Functorial vector space semantics and string diagrams for Lambek calculus». *Annals of pure and applied logic* 164, n. 11 (2013): 1079–1100.
4. Friesen, Jeff. *Java XML and JSON Document Processing for Java SE*, 2019. <https://doi.org/10.1007/978-1-4842-4330-5>.
5. Lambek, Joachim. *From Word to Sentence: a computational algebraic approach to grammar*. Polimetrica sas, 2008.
6. Liskov, Barbara, e John Guttag. *Program development in JAVA: abstraction, specification, and object-oriented design*. Computer programming. Pearson Education, 2001.
7. Weitzman, Michael. «The analysis of open traditions». *Studies in Bibliography* 38 (1985): 82–120.

# Filosofia e digitale: determinismo e pratiche di lettura sul web

Riccardo Fedriga<sup>1</sup>, Roberto Limonta<sup>2</sup>

<sup>1</sup> Università di Bologna – riccardo.fedriga [at] unibo.it

<sup>2</sup> Università di Bologna – roberto.limonta [at] unibo.it

## PAROLE CHIAVE

snippet, habitus, ontologia sociale, determinismo, storia della lettura

## POSTER

In questo contributo ci proponiamo di indagare i modi del determinismo sul Web tramite analisi della natura dei testi digitali e delle pratiche (*habitus* o abito) di lettura che inducono. In particolare, il riferimento centrale sarà una delle modalità testuali di più comune utilizzo, gli *snippet*, le brevi strisce testuali con cui i motori di ricerca rimandano alle fonti (piattaforme, siti, archivi digitali). Il *case study* della nuova normativa europea sul diritto d'autore (Direttiva UE 2019/790), che all'articolo 15 tocca il nodo della definizione degli oggetti digitali sul Web e dei diritti e limiti che la accompagnano, costituisce il banco di prova per mostrare il fondamentale fraintendimento riguardo a dove si collochi il confine di un oggetto digitale nella comunicazione online: la legge, infatti, colpisce l'azione "parassitaria" dei colossi del Web soltanto quando riproduce testi completi, quindi nella forma più tradizionale, ma non interviene (e anzi legittima) la loro traduzione in *snippet*, mostrando così di ignorare come quest'ultimo costituisca in buona parte l'unità di misura media di un contenuto informativo letto sul Web.

Gli oggetti digitali sul Web hanno la natura di oggetti sociali, secondo la definizione dell'ontologia sociale di John Searle [10] e di Barry Smith [11]. In Rete è l'intenzionalità collettiva degli utenti ad assegnare funzioni di status [10] agli oggetti digitali e ai loro connettivi: ad esempio, determinando che, all'interno di un ipertesto multimediale  $T_I$ , i grafici  $x$ ,  $y$  e  $z$  e il testo  $p$  hanno funzione argomentativa, la sequenza video  $w$  funzione narrativa e i link quella di connettivo logico, retorico o argomentativo. La rete dei rimandi non è solo sintattica ma anche semantica, rappresentando anche e soprattutto un apparato logico-argomentativo: l'ipertesto non soltanto dice  $x$  e  $y$ , ma dice anche che  $x$  e  $y$  sono in una certa relazione tra loro. La struttura dei link funge da regola costitutiva ( $X$  vale per  $Y$  in  $Z$ ) tramite la quale l'intenzionalità collettiva degli internauti costruisce i suoi aggregati assegnando il loro ruolo ai singoli contenuti (testi, immagini, video) che fungono da snodi della rete.

L'ontologia (i confini del testo) e la pragmatica (le pratiche di lettura) che si vengono così determinando producono a loro volta un'epistemologia della lettura digitale, in un intreccio dinamico dove ognuno dei tre ambiti agisce sugli altri e viceversa.

Ma la (apparente) liquidità libertaria degli ipertesti sul Web rischia di nascondere il determinismo digitale innescato dalle pratiche di lettura, guidate dagli algoritmi di ricerca. Possiamo definire il passaggio chiave come uno slittamento tra *status* [10] e *habitus* [9]: con il primo intendiamo le funzioni attribuite a un oggetto (nel nostro caso un testo) al di là della sua struttura “materiale”, e che ne fanno un oggetto sociale (cioè definito e determinato socialmente) e non semplicemente naturale, con *habitus* facciamo riferimento al formarsi di modi e abitudini di lettura. Il testo digitale presenta una struttura a rizoma dove lo *status*, cioè la posizione e la funzione delle sue diverse parti, si profila e si ridefinisce, di volta in volta, in base ai percorsi di navigazione. Questa condizione mobile e all'apparenza infinitamente aperta, tuttavia, crea strati di esperienza e meccanismi di risposta che si fissano e solidificano in un *habitus*, cioè una condizione stabile e funzionale delle nostre facoltà cognitive. Tramite gli *habitus* l'infosfera [6] va ridisegnando la mappa delle nostre funzioni cognitive in direzione di una logica sempre meno ipotattica e consequenziale, e sempre più paratattica, dove la conseguenza di un discorso è prevalentemente empatica e non logica, frutto di accostamenti, analogie, assonanze, similitudini, tutte funzionali al sistema più che ai suoi fruitori.

Alle questioni sopra descritte si lega la seconda parte del nostro contributo. Sempre in relazione alle pratiche di lettura, si partirà dalla seguente domanda: il *modus operandi* degli apparati e dei supporti digitali, ha già cominciato oggi e potrebbe continuare a dettare quello dei suoi utenti?

La risposta alla domanda può essere articolata su differenti piani. Si va da quello metafisico per cui ci si domanda se l'essere umano sia veramente libero, oppure se la nostra idea di libertà non sia semplicemente in qualche modo generata da una sorta di "apparato", come l'anima, del nostro corpo umano, e se questo apparato – cui fanno capo le nostre libere e responsabili azioni e scelte – non si collochi al livello dei fenomeni superiori (credenze, intenzioni, scelte, volizioni, azioni...), rispetto a quelli che l'hanno generato. E quindi un qualsiasi altro apparato, naturale o artificiale che sia, in grado di collocarsi a un livello superiore, è in grado di operare in modo da dettarci un *modus operandi* determinato.

Un altro piano di risposte, potrebbe poi riguardare la cosiddetta tecnologia lato utente. La questione è legata a quella del "tecno-determinismo" di cui gli storici della tecnologia come Hans-Dieter Hellige hanno spesso discusso nei loro testi ai quali intendiamo rifarci in modo critico. Il computer, prima, e la rivoluzione tecnologica che esso ha comportato e generato, in seguito, hanno effettivamente cambiato in modo piuttosto drammatico il modo in cui milioni di esseri umani svolgono oggi il loro lavoro quotidiano. In tali contesti, è già stata sollevata spesso la questione di quanto il computer sia ancora un aiutante dell'umano, o fino a che punto l'umano non sia già diventato una mera appendice della macchina (specialmente nei nuovi proletariati digitali quali quelli del lavoro intellettuale a distanza dettati dalle piattaforme dell'economia dei dati). L'idea del tecno-determinismo è stata confutata sotto due grandi punti di vista: uno è la stupefacente "flessibilità" del computer digitale che, anche se non è onnipotente, “può essere programmato per risolvere una quantità (o "classe") sorprendentemente grande di problemi computabili, così che un auspicabile strumento software utile, che non è ancora nelle nostre mani oggi, potrebbe presto essere disponibile in un futuro non troppo lontano.

In secondo luogo, è molto d'aiuto in questo senso la considerazione di una memoria storica sedimentata secondo pratiche di lettura “a palinsesto” e non solo dettata da un triviale determinismo causale per cui il calcolo delle registrazioni effettuate nel passato si trasferisce necessariamente sulla previsione delle future, bloccando l'apertura tanto del futuro quanto del passato.

Basti pensare al fatto, sottolineato da Stephan Gruner, che il ricercatore nel campo delle scienze umane (per ora ancora digitali e ben presto si spera tout court intese) ha ancora la possibilità di ripiegare su qualsiasi metodo classico di ricerca, decidendo liberamente se e dove un computer digitale non è (o non potrà mai essere) di aiuto in una particolare situazione di ricerca o di lavoro.

## BIBLIOGRAFIA

1. Cardon, Dominique. *Che cosa sognano gli algoritmi. Le nostre vite al tempo dei Big Data*. Tradotto da Chetro De Carolis. Mondadori, 2016.
2. Ciracì, F, R Fedriga, C Gruner, e C Marras. «DH: un dialogo a più voci». *Filosofia (numero monografico)* (forthcoming) (2021).
3. Eco, Umberto. *Dall'albero al labirinto: Studi storici sul segno e l'interpretazione*. La Nave di Teseo Editore spa, 2017.
4. Ferraris, Maurizio. *Documanità*. Bari: Laterza, 2021.
5. Ferraris, Maurizio, e Germano Painsi. *Scienza nuova. Ontologia della trasformazione digitale*. Torino: Rosenberg & Seller, 2018.
6. Floridi, Luciano. *La quarta rivoluzione: come l'infosfera sta trasformando il mondo*. Raffaello Cortina Editore, 2017.
7. Hellige, Hans Dieter. *Geschichten der Informatik: Visionen, Paradigmen, Leitmotive*. Springer-Verlag, 2013.
8. Husserl, Edmund. *Esperienza e giudizio*. Tradotto da F. Costa e L. Samonà. Bompiani, 2007.
9. Sapiro, Gisèle. «Pour une approche sociologique des relations entre littérature et idéologie». *COnTEXTES. Revue de sociologie de la littérature*, n. 2 (2007).
10. Searle, John R. *La costruzione della realtà sociale*. Milano: Einaudi, 2006.
11. Smith, Barry. «Oggetti fiat», 2002.

# Migrations, displacements and relocations: narrative cartography of movement

Tiago Gil  
University of Brasilia – tiagoluisgil [at] gmail.com

## KEYWORDS

narrative maps, motion, moviment, demographic displacement, social networks

## POSTER

This research aims to understand different ways of describing movement in cartography, taking as a starting point ongoing research on the production of cotton fabrics inside Portuguese America at the turn of the 18<sup>th</sup> to 19<sup>th</sup> century. There are many studies on representing motion in cartography. Recently, Cacquard, Pearce, Knowles, and other research has offered new ways of thinking about movement in cartographic terms. Despite the recognized importance of movement and process in historical knowledge, historians have rarely discussed ways to map movement. I will call these maps for historical research purposes narrative maps [2].

So-called narrative maps have been carefully studied and countless perspectives have been seeking ways to represent narratives, especially fictional ones, mainly after Franco Moretti's seminal study and its developments [6]. The results of these explorations are varied in quality and perspective. Nevertheless, they have presented timely reflections that call for a review of cartography itself as a discipline. Historians, as a whole, make slight use of cartography. I do not mean that they do not use maps, or even ignore researchers of the time who were dedicated to studying or creating through cartographic language, far from it. The potential for using this language to represent historical processes is so vast and vital that, however great, its use in research will still be little. One of the most common uses that historians make of cartography is illustrative, presenting maps to introduce the geographical scenario where the plot will take place. A much more elaborate example is to use maps to display quantitative data in specific periods or in time series. This scenario is clearly changing and there are several examples of systematic cartographic language usage [4].

This research uses an approach from economic history, which seeks to explore the universe of spinners, weavers and farmers engaged in the cotton plantation in the American Portuguese colony, present day Brazil, more precisely in the Rio Paraíba valley, in the São Paulo Captaincy. At the end of the 18th century, there were a large number of domestic units spinning cotton in the Captaincy of São Paulo. The four villages that formed what we now call the Paraíba River Valley (Guaratinguetá, Pindamonhangaba, Taubaté and Lorena) were prominent in this craft, with a large concentration of families dedicated to spinning, weaving and sewing. At the same time, practically the entire captaincy cultivated Gossypium.

Considering only the locations where cotton was processed, we may wonder if we are facing some sort of industrial activity, even though one distributed throughout different domestic units, forming a single set in the end.

The productive activity of cotton, from planting to fabric production, was carried out separately in different domestic units. Some made the product in natura, others spun it, others wove it and even probable clothing production is seen in this same region. This production depended on the movement of unfinished and raw products between different domestic units, which simultaneously functioned as productive units. At the same time, it was possible to identify an intense demographic movement in this region during the period, with migration, displacement and relocation at different geographical scales. Were these population movements linked to the production of cotton and fabrics?

Faced with a concept that accounts for the complexity of the phenomenon, we chose to work with the notion of “*industrie familiale*”, as presented by Cottureau [3]. Cottureau’s concept does not presuppose a strict relationship with any other process, demographic or social, that is contemporary, but nor does it exclude any potential relationship. The empirical material already observed for Vale do Paraíba indicates that a strong relationship between land ownership and demographic orientation within the house are very relevant factors in understanding the productive activity of spinning and weaving. Small land size would force investment in non-agricultural activities. At the same time, the phenomenon of widowhood seems to have had a substantial impact on the decision to produce thread or fabric, given the few alternatives. Thus, it seems evident that the domestic production process is related to other processes and events. The problem is that the available and clearer concepts of crafts and proto-industrialization presuppose very specific and predictable relationships between these phenomena. In this sense, the instrumental concept of “*industrie familiale*” will be used in an ad hoc manner, awaiting a more exact definition during the research [5].

The sources of this study are the most traditional in social history: parish registers, public deeds and probate inventories. One of these sources is little known and uncommon for other colonial regions: its nominal lists of inhabitants. It was a kind of municipal census in which someone (probably the captain-general) went from house to house, noting various information about families: the chief, his wife, children, households and slaves, in addition to the economic activity of each house. This source details where people were and their destinations, who got married, died, or just moved. This data indicates, in particular, where these people went or where they came from [1]. From the municipal census, we have produced analyses of movement and discuss ways to represent them through the support of a cartographic series organized specifically for this research. The goal is to create an annual series of demographic movements in space, taking places as nodules of a social network, visible in the map below.



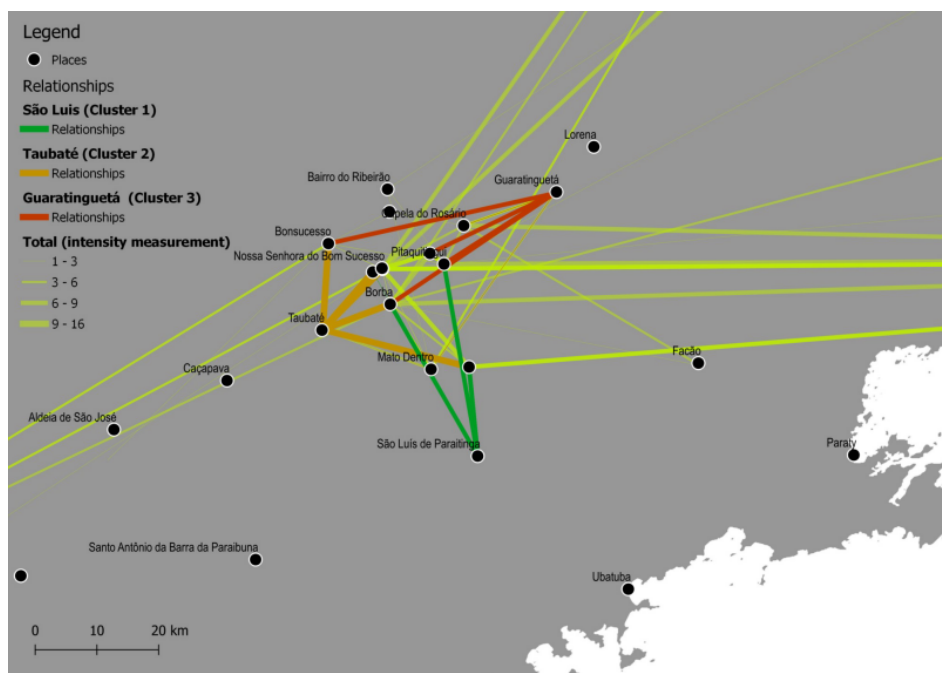


Figure 1: demographic movement of people in the Rio Paraíba Valley, 1802

In this sense, our purpose is to create a series of maps of spatial networks, articulating Social Network Analysis and Geographic Information Systems in History (HGIS). The results will be evaluated in search of straightforward ways of communicating historical processes over time through maps.

## REFERENCES

1. Bacellar, Carlos de Almeida Prado. 'As Listas Nominativas de Habitantes Da Capitania de São Paulo Sob Um Olhar Crítico (1765-1836).' *Anais de Historia de Alem-Mar*, no. 16 (2015).
2. Caquard, Sébastien, and William Cartwright. *Narrative Cartography: From Mapping Stories to the Narrative of Maps and Mapping*. Taylor & Francis, 2014.
3. Cottureau, Alain. 'Sens Du Juste et Usages Du Droit Du Travail: Une Évolution Contrastée Entre La France et La Grande-Bretagne Au XIXe Siècle'. *Revue d'histoire Du XIXe Siècle. Société d'histoire de La Révolution de 1848 et Des Révolutions Du XIXe Siècle*, no. 33 (2006): 101–120.
4. Gabellieri, Nicola, and Massimiliano Grava. 'A Changing Identity: From an Agrarian and Manufacturing Region to a Multi-Functional Territory'. In *Atti Del International Scientific Symposium "Placenames as Intangible Cultural Heritage"*, Florence 25th – 27th March 2015. Firenze: Istituto Geografico Militare, 2016.
5. Mendels, Franklin F. 'Proto-Industrialization: The First Phase of the Industrialization Process'. *The Journal of Economic History* 32, no. 1 (1972): 241–261.
6. Moretti, Franco. *Atlas of the European Novel, 1800-1900*. verso, 1999.

# **“PH-Remix”: un progetto interdisciplinare per la valorizzazione del patrimonio audiovisivo del Festival dei Popoli Festival Internazionale del Film Documentario di Firenze in ambiente digitale**

Giovanni Grasso<sup>1</sup>, Chiara Mannari<sup>2</sup>, Davide Serramazza<sup>3</sup>

1 Università di Pisa – giovanni.grasso [at] fileli.unipi.it

2 Università di Pisa – chiara.mannari [at] fileli.unipi.it

3 Università di Pisa – davide.serramazza [at] fileli.unipi.it

## **SINTESI**

Il seguente contributo mira ad illustrare le finalità previste dal progetto “PH-Remix” (*Public History Remix*) dell’Università di Pisa per la valorizzazione del patrimonio filmico del Festival dei Popoli - Festival Internazionale del Film Documentario di Firenze in collaborazione con la Fondazione Mediateca Toscana (quest’ultima dal 2011 socio FIAF [International Federation of Film Archives]). Un progetto biennale ancora in fase embrionale che ha lo scopo di valorizzare, attraverso la creazione di una piattaforma web basata sull’intelligenza artificiale, l’ampio patrimonio filmico (oltre 25.000 titoli) in diversi formati (pellicole 16 e 35 mm, VHS, U-matic, DVD) raccolto e conservato nel corso dei sessant’anni di attività del Festival dei Popoli, un’eredità culturale in continua espansione grazie allo svolgimento del FdP a cui pervengono annualmente circa 1600 nuove opere.

L’archivio cinematografico del FdP ha cominciato a sedimentarsi sin dalle prime edizioni (la prima è del 1959), quando emerse la consapevolezza tra gli organizzatori di dover conservare un patrimonio culturale cinematografico distante da quelli che erano i circuiti di distribuzione piuttosto tradizionali e convenzionali del tempo. La complessità e la varietà dei titoli presenti nell’archivio riflettono l’evoluzione del cinema documentario negli ultimi sessant’anni per contenuti, stile e linguaggio (in considerazione anche degli oltre settanta paesi partecipanti), che includono le più disparate categorie del cinema documentario da quelle più strettamente antropologiche e sociali a quelle riguardanti la storia contemporanea, la storia dell’arte, la musica, la danza, il costume, l’architettura, i mezzi di comunicazione, l’attualità e l’attualità politica. Fra le varie opere presenti nell’archivio spiccano quelle dei maestri Joris Ivens, Ermanno Olmi, Vittorio De Seta, Ken Loach, Werner Herzog, Marcel Łoziński, Vincent Dieutre, Jean Rouch, Frederick Wiseman e Nagisa Oshima per citarne alcuni

## PAROLE CHIAVE

cinema e storia, remix culture, intelligenza artificiale, archivi audiovisivi, piattaforma web, digital public history

## POSTER

### 1 STUDIO DELLE FONTI AUDIOVISIVE NELLA DIGITAL PUBLIC HISTORY

La salvaguardia e la valorizzazione di questo ingente patrimonio nasce dalla consapevolezza di dover render note le due narrazioni di cui la storia degli audiovisivi (in questo caso dell'archivio del FdP) si compone: da un lato il prodotto filmico inteso come l'intenzionalità degli autori e la sua diffusione in qualità di testimone del mondo contemporaneo e dell'immaginario sociale; dall'altro la storia fisica delle copie che a tali prodotti hanno fornito il supporto materiale, cercando così di trasformare il passato in una risorsa e, al contempo, restituire dignità culturale a queste produzioni.

Per quanto concerne quest'ultimo aspetto risulta necessario un intervento volto ad individuare quelle copie su supporti analogici che necessitano di essere preservate sulla base del valore storico dell'opera, sulla base del rischio di deterioramento della copia fisica e sulle potenzialità del loro riutilizzo in un'ottica di *remix culture* per la creazione di nuovi prodotti audiovisivi in grado di sviluppare nuovi punti di vista, nuove storie, nuovi significati, utilizzando le fonti filmiche primarie.

Per quanto invece riguarda il valore di documento storico, il cinema non deve essere considerato soltanto come spettacolo e/o mero intrattenimento: grazie ad innovativi strumenti in grado di rendere fruibili in rete cospicui archivi audiovisivi sedimentatisi in ambiente digitale, si deve oggi considerare il documento audiovisivo come una nuova risorsa per lo studio e la trasmissione della storia al fianco della memoria ufficiale divulgata attraverso manuali e volumi di storia tradizionali, anche di storia del cinema. Come scrive Pierre Sorlin: «I prodotti audiovisivi non sono, dunque, una documentazione supplementare, un mezzo per conoscere meglio gli eventi, i mestieri, le feste, le cerimonie, ma sono del tutto integrati nella vita, nelle normali attività, formano la trama della quotidianità, né più né meno delle case, il lavoro, le relazioni familiari. [...] pur senza esserne sempre consapevoli, noi siamo integrati in un vasto sistema di rappresentazione: osserviamo il mondo sullo schermo, ma siamo ugualmente quelli stessi che vengono filmati dalla cinepresa» [11].

Partendo da questi presupposti, lo storico che intende utilizzare il documento audiovisivo come nuova risorsa per lo studio della storia deve considerare i rapporti tra cinema e storia secondo tre prospettive principali: il cinema come agente di storia (come creatore di tendenze, immaginari e comportamenti collettivi); il cinema come strumento per raccontare la storia (strumento per raccontare il passato che il film ci racconta e rappresenta sullo schermo); ed infine, il cinema come fonte per la storia riferendosi al presente in cui il film è stato girato (all'interno di un discorso meta-cinematografico che analizza tutti quegli elementi che ci portano dentro la storia del cinema valutando il significato del film nel suo contesto produttivo, socioeconomico e politico).

Ma parallelamente a quest'approccio e alla metodologia di ricerca tradizionale, lo storico che intende servirsi del cinema per lo studio della storia, deve fare i conti con «la rivoluzione digitale [che] ha avuto un profondo impatto sul modo in cui oggi la storia viene studiata, analizzata, condivisa, insegnata oltre che sul modo in cui le fonti che documentano i fenomeni del passato vengono pubblicate, conservate e addirittura prodotte». Lo storico deve necessariamente riflettere sul proprio ruolo sociale nel presente, sull'uso pubblico della storia e del cinema da parte dei soggetti sociali e superare una certa diffidenza «verso il computer e la rete, lo scarso valore attribuito ai prodotti digitali, [...] l'ignoranza diffusa relativa ai principali strumenti del mondo digitale, la loro rapida e continua evoluzione» e abbandonare l'opinione «che l'informatica sia appunto solo uno 'strumento', potenzialmente utile all'umanista solo nella misura in cui lo aiuta a risolvere alcune esigenze di ricerca e non un modo di analizzare, organizzare e visualizzare la conoscenza» [10].

Il progetto nasce proprio dall'esigenza di creare un nuovo strumento per raccontare e rielaborare la storia degli audiovisivi che compongono l'archivio del FdP sotto i due punti di vista poc'anzi presentati e fare in modo che diverse tipologie di pubblico possano usufruire in rete di un patrimonio archivistico tra i più importanti in ambito audiovisivo; ciò è possibile unendo tre ambiti disciplinari differenti (la storiografia cinematografica, l'informatica e l'informatica umanistica), condividendo i metodi e le strategie derivanti dai diversi settori e giungendo ad una consapevolezza delle metodologie delle tre discipline: multidisciplinarietà, o meglio interdisciplinarietà, che è alla base della *digital public history*.

## 2 LA PIATTAFORMA “PH-REMIX”

La piattaforma “PH-Remix” che nasce e si sviluppa all'interno dell'omonimo progetto è lo strumento previsto per consentire l'accesso alle fonti e il remix dei video. Il patrimonio filmico del FdP costituisce il nucleo di partenza che andrà ad alimentare la piattaforma ed è l'oggetto del processo di estrazione automatica presentato in seguito. L'archivio, al cui interno troviamo film documentari provenienti da ogni parte del mondo che affrontano temi disparati con una eterogeneità di stili ed approcci, valorizza autori lontani dal centro del mercato cinematografico e si presta bene allo studio e al riutilizzo dei materiali ai fini della produzione di nuovi contenuti di tema storico.

La piattaforma, tuttora in corso di progettazione, sarà un'applicazione web composta da due componenti essenziali: la prima componente fornirà l'accesso ad un indice molto ampio da cui sarà possibile, attraverso percorsi di ricerca e navigazione, visionare ed infine selezionare i contenuti accessibili sotto forma di clip. Le clip consistono in frammenti video di durata variabile contenenti scene o momenti significativi estratti, oltre che attraverso una schedatura tradizionale basata su temi e parole chiave, anche tramite metodi innovativi di estrazione automatica dell'informazione ottenuti mediante l'uso di tecniche di intelligenza artificiale; la seconda componente si occuperà del montaggio del nuovo video attraverso l'unione di più clip, l'aggiunta di immagini, testi, tracce audio e l'esportazione finale.

Una piattaforma di questo tipo si inserisce nel contesto della *remix culture*, una pratica di produzione artistica basata sul riutilizzo di contenuti provenienti da svariate fonti che ha preso piede negli anni Sessanta nel panorama dei remix musicali per poi diffondersi anche in campo audiovisivo.

Con l'invenzione del web la *remix culture* ha conquistato terreno e sono stati aperti numerosi siti web e spazi virtuali dove condividere progetti di remix<sup>1</sup>. Nell'era di Instagram Reels<sup>2</sup>, delle coreografie su Tik Tok<sup>3</sup> e dei film interattivi<sup>4</sup>, in una realtà in cui le tradizionali modalità di accesso ai contenuti sono stravolte e gli utenti sono diventati allo stesso tempo fruitori e creatori, la *remix culture* è penetrata al punto da far perdere la percezione di sé: *everything is a Remix*<sup>5</sup>.

La natura interdisciplinare del progetto "PH-Remix" permette di applicare buone pratiche ed elementi innovativi provenienti dalle discipline coinvolte facendo emergere i seguenti punti di forza:

- Sperimentare la pratica del remix nell'ambito della *digital public history* in quanto disciplina che ha come obiettivo primario quello di avvicinare storia e pubblico attraverso lo sviluppo di strumenti al passo con l'evoluzione tecnologica e di nuove modalità di interazione;
- Rendere accessibile agli studiosi fonti storiche, la cui fruizione risultava più complicata, grazie all'accesso a un ingente patrimonio filmico digitalizzato, schedato e indicizzato in maniera sistematica;
- Aprire ad una serie di scenari interessanti per la didattica e la produzione artistica;
- Servirsi del contributo dell'intelligenza artificiale, delle tecniche di estrazione automatica e in particolar modo del *machine learning* per la valorizzazione del patrimonio storico e culturale grazie alla possibilità di processare, in tempi brevi, quantità di dati impossibili da trattare con schedature manuali limitate nell'azione da vincoli temporali.

Le problematiche emerse fin dalle fasi iniziali del progetto riguardano soprattutto due aspetti:

- Il primo consiste nella tutela giuridica dei materiali d'archivio e lo sviluppo della piattaforma nel rispetto del diritto d'autore. In questa direzione si sta tentando, in accordo con i vari partner di progetto, di impiegare un sistema di licenze di utilizzo che si basi sul modello delle *Creative Commons*<sup>6</sup>;
- Il secondo riguarda alcune difficoltà nella definizione delle specifiche della piattaforma dipendenti sia dalle tempistiche interne al progetto "PH-Remix", in particolare in merito ai tempi necessari all'analisi dell'archivio FdP e alla sperimentazione delle tecniche di estrazione dell'informazione, che dalle tempistiche del progetto parallelo di Fondazione Mediateca Toscana che prevede l'acquisizione dell'archivio FdP all'interno della mediateca e la digitalizzazione dell'intero patrimonio archivistico (i cui materiali sono al momento conservati in copia unica su supporto analogico).

---

1 Rip! A Remix manifesto

<https://www.brettgaylor.com/#/rip-a-remix-manifesto/>

2 <https://about.instagram.com/blog/announcements/introducing-instagram-reels-announcement>

3 <https://www.nytimes.com/2019/03/10/style/what-is-tik-tok.html>

4 <https://medium.com/swlh/films-that-watch-you-while-you-watch-them-a-brief-history-of-interactive-movies-820fb5e2ee25>

5 Cfr. Ferguson K., *Everything is a Remix*

<https://www.everythingsaremix.info/watch-the-series>

6 <https://creativecommons.org/>

Per far fronte a queste difficoltà, emerge quindi la necessità di creare un sistema flessibile in cui le modifiche di progetto siano gestite in maniera ottimale anche sul lungo termine. La soluzione per soddisfare questa esigenza è quella di realizzare un'architettura software basata su microservizi in cui tutte le componenti siano separate e implementabili anche in momenti successivi. Il sistema verrà progettato affinché risulti possibile in qualsiasi momento aggiungere agevolmente nuove funzionalità come ad esempio nuovi criteri di indicizzazione e di ricerca, oppure sottomettere e analizzare altri film in maniera piuttosto semplice anche tramite un'interfaccia web utilizzabile da utenti non esperti. L'architettura prevista permette inoltre di avere maggiore flessibilità per quanto riguarda lo strumento di remix: l'indice e tutti i moduli che verranno implementati potranno essere riutilizzati all'interno di nuove interfacce, come ad esempio nel caso in cui si decida di affiancare all'applicazione web in corso di sviluppo una o più app mobile.

### 3 IL CONTRIBUTO DELL'INTELLIGENZA ARTIFICIALE

Per tale tipo di patrimonio composto da materiale audiovisivo, la piattaforma sin qui descritta può essere realizzata mediante l'uso di tecniche d'intelligenza artificiale da affiancare a un processo più 'classico' di catalogazione manuale.

In letteratura scientifica sono presenti diversi modelli pre-allenati per etichettare parti di video: questi modelli, da qui in poi chiamati anche *processor*, possono essere assoggettati agli scopi del progetto per suddividere i documentari in varie clip, ovvero in frammenti contigui e omogenei all'interno del film. Le clip così estratte saranno inoltre **indicizzate**, assegnando ad ognuna di esse una **label** che andrà ad indicare l'informazione ricavata sulla porzione di video dai modelli adoperati.

Per entrare nel dettaglio dell'intero processo composto da estrazione, indicizzazione ed etichettamento, poniamo un esempio: possibili *task* svolti dai *processor* consistono nell'identificazione di oggetti nelle clip come in *Looking fast and slow* [6], di azioni compiute come nell'algoritmo di Kalfaoglu [4] o ancora nell'estrazione di testo dal video (in questo caso i sottotitoli) come nel lavoro di Wang [14]. I *processor* vengono utilizzati per suddividere un singolo film in più clip basandosi sui *task* precedentemente menzionati. Inoltre, è da sottolineare come da un singolo film vengano estratti **più insieme di clip**, uno per ciascun *processor* utilizzato; fra clip estratte dallo stesso documentario possono potenzialmente verificarsi delle sovrapposizioni. Infatti, una clip in cui appare l'oggetto A, si può sovrapporre, anche solo parzialmente, ad un'altra clip in cui è compiuta l'azione B, ma anche ad un'altra clip in cui compare un altro oggetto C.

Una volta isolate dal resto del documentario, le clip riceveranno una *label* corrispondente all'informazione estratta dal modello utilizzato: le *label* assegnate coincideranno con le parole chiave utilizzate dagli utenti della piattaforma di remix per **ricercare** una particolare clip (ad esempio tutte quelle contenenti un particolare tipo di oggetto, di azione, ecc.). Tale processo andrà a costruire l'indice a cui si è accennato nella sezione precedente, che sarà composto dalle varie clip estratte dai film a nostra disposizione e dalle *label* ad esse associate. Gli esempi riportati precedentemente di *processor* non sono da ritenersi esaustivi: il numero, le categorie alle quali apparterranno e i modelli specifici che verranno adoperati, saranno decisi durante lo svolgimento del progetto.

Ai modelli pre-allenati saranno affiancati altri tipi di modelli sviluppati *ex novo*: verrà dunque interamente progettata l'architettura e saranno successivamente allenati usando i documentari come dati di allenamento. Quest'ultimo tipo di modelli risulta ancora più interessante dal punto di vista della ricerca scientifica.

Un esempio di tali funzionalità riguarda la possibilità da parte dell'utente finale, una volta scelte due clip A e C, di poter ottenere dal sistema un suggerimento di una terza clip B come intermezzo adeguato: in tal modo si potrebbero scoprire relazioni non banali tra clip provenienti da due film con tematiche non apparentemente collegate tra loro.

Tecnicamente ciò potrebbe essere realizzato tramite tecniche di estrazione di *features* a partire dalle clip e tramite l'utilizzo di uno spazio comune di *embedding*: le varie clip precedentemente estratte saranno dunque mappate in uno **spazio semantico comune** nel quale sarà possibile calcolare la distanza fra i vari elementi. A tal proposito il nostro intento è quello di applicare al video l'utilizzo degli *embedding* neurali per modellare ed estrarre nuove relazioni (già investigate nel lavoro di Xu [15]) e sfruttarli per quantificare la similarità tra le varie clip e dunque scoprire nuove relazioni fra esse.

Un altro esempio di funzionalità sviluppata *ex novo* sarà il supporto fornito agli esperti del settore nella catalogazione dei film: il modello previsto, analizzando aspetti quali il contenuto delle immagini, delle musiche di sottofondo e dei dialoghi di una determinata sequenza, predirà le tematiche trattate. La problematica maggiore risulta essere l'identificazione dell'insieme delle tematiche che tale modello può predire: se i documentari hanno per la maggior parte un'indicazione delle tematiche trattate verranno utilizzate tali categorie per comporre il nostro insieme, altrimenti il modello prevederà l'uso di tecniche *unsupervised* che non richiedono la presenza di dati etichettati.

Quanto descritto in questa sezione, ovvero sia i *processor* che i nuovi modelli che ci si propone di sviluppare, verranno tutti implementati come eseguibili separati nell'ottica di costruire un'architettura a microservizi come precedentemente accennato. I vantaggi di tale scelta sono principalmente due: ogni funzionalità è indipendente dalle altre, ovvero l'aggiunta/rimozione/modifica di una singola funzionalità sarà molto più agevole rispetto a un corrispettivo monolitico e in secondo luogo viene favorita la scalabilità del sistema in maniera tale da poter sfruttare tutte le risorse computazionali in nostro possesso.

I risultati di tale sezione del progetto, oltre a essere indispensabili per la piattaforma "PH-Remix" alla quale verranno fornite le clip ricercabili dagli utenti, sono di per sé utili anche ad altri tipi di studi e analisi da svolgere su di essi. Un esempio su tutti riguarda la trascrizione dei dialoghi presenti nel testo: da esse è possibile ottenere dei corpora etichettati con le varie informazioni ricavabili dai documentari, sia in maniera automatica con gli strumenti descritti in precedenza, che manualmente con l'ausilio di un esperto. Questi corpora costituirebbero una risorsa preziosa per alcuni studiosi, come ad esempio storici o linguisti, che potranno ulteriormente analizzare tale corpora autonomamente in maniera del tutto indipendente da tale progetto, utilizzando altre metodologie più consone alle loro discipline.

## BIBLIOGRAFIA

1. Bell, Desmond. «Documentary film and the poetics of history». *Journal of Media Practice* 12, n. 1 (2011): 3–25.
2. De Luna, Giovanni. «Le nuove frontiere della storia. Il cinema come documento storico». *Le fonti - audiovisive per la storia e la didattica*, Annali AAMOD, n. 16 (2013).
3. Gallagher, Owen. *Reclaiming Critical Remix Video: The Role of Sampling in Transformative Works*. Routledge, 2018.
4. Kalfaoglu, M., Sinan Kalkan, e A. Aydin Alatan. «Late temporal modeling in 3d cnn architectures with bert for action recognition». In *arXiv preprint arXiv:2008.01232*, 2020.
5. Lagny, Michèle. «Il cinema come fonte di storia». A cura di Gian Piero Runetta. *Storia del cinema mondial* 5 (2009): 99–131.
6. Liu, Mason, Menglong Zhu, Marie White, Yinxiao Li, e Dmitry Kalenichenko. «Looking fast and slow: Memory-guided mobile video object detection». *arXiv preprint arXiv:1903.10172*, 2019.
7. Navas, Eduardo. *Remix theory: The aesthetics of sampling*. Springer, 2012.
8. Pang, Natalie, Kai Khiun Liew, e Brenda Chan. «Participatory archives in a world of ubiquitous media». *Archives and Manuscripts* 42, n. 1 (2014): 1–4.
9. Rosenstone, Robert A. *Revisioning history: Film and the construction of a new past*. Vol. 5. Princeton University Press, 1995.
10. Salvatori, Enrica. «Digital (Public) History: la nuova strada di una antica disciplina». A cura di Giovanni Sini. *RiMe. Rivista dell'Istituto di Storia dell'Europa Mediterranea*, 2017, 57–94.
11. Sorlin, Pierre. «L'immagine e l'evento. L'uso storico delle fonti audiovisive», 1999.
12. ———. «Una premessa. Scuola, storia, mezzi audiovisivi nell'era digitale». *Le fonti - audiovisive per la storia e la didattica*, n. 16 (2013).
13. Taviani, Ermanno. «Il cinema e gli storici». *Le fonti - audiovisive per la storia e la didattica*, n. 16 (2013).
14. Wang, Tao. «End-to-end text recognition with convolutional neural networks». In *Proceedings of the 21st international conference on pattern recognition (ICPR2012)*, 3304–3308. IEEE, 2012.
15. Xu, Peng, e Denilson Barbosa. «Investigations on knowledge base embedding for relation prediction and extraction». *arXiv preprint arXiv:1802.02114*, 2018.



# Un Historical GIS per lo studio della geografia medica storica

Massimiliano Grava  
Università di Pisa – massimiliano.grava [at] unipi.it

## SINTESI

In questo contributo illustreremo come l'uso di tecnologie GIS (Geographic Information System) consenta, mettendo mano a fonti storiche impiegate sino ad oggi con finalità quantitative, di realizzare, grazie all'incrocio tra le serie di dati create utilizzando questi dati e la cartografia digitale odierna, studi qualitativi sulla distribuzione dei fenomeni sanitari storici. La fonte archivistica impiegata in questa ricerca per la costruzione del geodatabase è una serie documentaria di straordinario interesse per la rilevante qualità delle informazioni in essa contenuta. Questa fonte, prodotta dall'Ufficio di Sanità in occasione della peste che contagiò la città e il contado di Pisa nel 1630 (calendario pisano), è difatti particolarmente adatta per elaborazioni con questo tipo di applicativi. I documenti prodotti dall'Ufficio di Sanità di Pisa, una dozzina di buste, illustrano le misure adottate per contenere il morbo indicando, in modo più dettagliato per la città rispetto ai vicariati ad essa confinanti, la struttura dei nuclei familiari, il numero dei contagiati trasferiti nei lazzeretti, dei morti e le pratiche adottate per contrastare "il morbo".

Una banca dati geografica che, in epoche di pandemia quale quella che attualmente stiamo attraversando, può essere un utile strumento per prevedere, a parità di condizioni e analoghi contesti, le odierne dinamiche epidemiologiche che attraversano la nostra società.

## PAROLE CHIAVE

GIS (Geographic Information System), peste, ufficio di sanità

## POSTER

### 1 INTRODUZIONE

I GIS sono oggi impiegati in tutte le attività connesse con la ricerca e lo studio di qualsivoglia fenomeno che sia in relazione con una superficie; sia questa il pianeta terra piuttosto che la più microscopica cellula epiteliale.

Nati e sviluppati tra Canada e Stati Uniti negli anni '60 del secolo scorso, ove furono impiegati rispettivamente per inventari agroforestali e con finalità militari, nei decenni successivi i GIS hanno conosciuto una straordinaria diffusione. Artefice primo della "popolarizzazione" dei dati geografici "digitali" e

quindi dei GIS con cui questi sono elaborati è stata indubbiamente Google. La società fondata da Larry Page e Sergey Brin a pochi anni dalla fondazione (avvenuta nel 1998) ha letteralmente rivoluzionato il mondo dell'informazione geografica con la pubblicazione dei primi «maps service» il servizio «a bird's-eye view» della società californiana, è infatti riuscito a ricoprire di foto satellitari e di informazioni commerciali il territorio di USA, Regno Unito, Francia, Germania, Italia, Spagna, Australia; per poi “avvolgere” di immagini georeferenziate scattate dal satellite la superficie dell'intero pianeta<sup>1</sup>.

Tra i settori ove con più proficuo sono stati impiegati i sistemi informativi geografici troviamo innegabilmente quello della Medicina. I Medical-GIS sono infatti ampiamente diffusi tanto a livello europeo quanto mondiale e le capacità dei programmi GIS di analisi e studio delle incidenze delle malattie (dati), rispetto al territorio (geografia) è evidente quanto realmente utile.<sup>1</sup> Meno diffuso l'impiego di questi applicativi per ricerche di geo-medicina storica. Il ritardo di ricerche geostorico-mediche (sic.), è causato da uno scarso impiego di uno “strumento” che è invece, come tenteremo di mettere in luce in questo contributo, fondamentale per l'elaborazione di costrutti storici che considerino finalmente prospettive molto più ampie di quelle “ortodosse” e di impiegare ricerche umanistiche con scopi che non solo quelli letterari.

## 2 LUNGHEZZA DELLE PROPOSTE

Le fonti d'archivio utilizzate sono essenzialmente quelle prodotte dall'Ufficio di Sanità nel periodo compreso tra il 1630 e il 1633 (calendario pisano)<sup>2</sup>, cui vanno aggiunte le informazioni estratte da due manoscritti: le Croniche di Pisa di Jacopo Arrostiti (1655), e le Memorie in tempo di mal contagio di Paolo Tronci (GRAVA 2014, pp. 269-277). Questa documentazione, una dozzina di buste per oltre 10.000 carte, è conservata nei fondi di Commissariato (dove sono concentrate il maggior numero di unità archivistiche), Comune D ed F e Manoscritti di libera proprietà. In questi documenti si trovano quindi trattate, in tutti i loro aspetti, le vicende relative alla crisi epidemica pisana (città e contado). I corpus documentari di maggior rilievo sono quelli relativi all'amministrazione sanitaria e della giustizia di una Pisa che, assediata dal morbo, si trovava del tutto isolata, si poteva infatti entrare o uscire dalla città solo mostrando alle guardie poste alle porte della cinta muraria autorizzazioni (Bullette) concesse dall'Ufficio di Sanità e se si fosse trasgredito agli “Ordini” si sarebbe rischiato arresto e sanzioni economiche. Una popolazione costretta quindi a più riprese alla quarantena (alcune strade vennero “sbarrate” in più occasioni) e comunque obbligata, come ci ricordano gli stessi cronisti dell'epoca, a ordinamenti eccezionalmente stringenti le libertà personali.

---

1 <http://cdrewu.eagle-i.net/sweet/provider?uri=http://cdrewu.eagle-i.net/i/00000135-8c2b-f30d-1a88-e81c80000000>  
<http://www.cdc.gov/nchs/data/gis/atmappne.pdf>

2 Il Calendario Pisano, o stile dell'Incarnazione o anche stile pisano, era il calendario utilizzato a Pisa e faceva iniziare l'anno il 25 di marzo (festa dell'annunciazione della vergine Maria). Questo come altri calendari in uso in Toscana fu definitivamente soppresso con Motuproprio dal Granduca il 20 novembre 1749.

“Non veniva in Pisa più mercanzie di fuori, e di Pisa non usciva nulla essendo per tutto bandito che si pativa una gran carestia; e di molti poveri dallo stento morivano che propriamente pareva un assedio, a tal che la povera gente pativa di fame e di peste. Quando si andava alle botteghe a comprare qualcosa, non si entrava dentro ma di fuori si diceva il suo bisogno: dato che il bottegaio aveva la robba pigliava i denari in sur una tavoletta per vedere se il conto stava bene, e poi gli buttava in una catinella piena di aceto”<sup>3</sup>.

La prima morte causata dal bubbone fu attestata il 30 settembre 1630 in via Santa Viviana (nel quartiere settentrionale della città rispetto al fiume Arno di Sant'Antonio), alla quale seguì, nei giorni successivi, il decesso di un intero nucleo familiare nel quartiere meridionale di San Martino. Giulio del Mosca, Magistrato di Sanità, decise allora di suddividere Pisa in sestieri cui vennero assegnati dei Deputati (principalmente reclutati tra le famiglie nobiliari pisane) che avevano, quale obbligo principale, quello di visitare e porre in quarantena o nei lazzaretti (ne furono creati tre), gli infetti e i loro familiari. Durante le loro visite nei sestieri di competenza questi funzionari annotarono, in registri arrivati in buona parte sino a noi, struttura e composizione di tutti i nuclei familiari (scopo principale di questo censimento fu quello di individuare le famiglie “che non possono andare a lavorare [...] e per i quali invece bisognerà farli le spese a tutta la famiglia”)<sup>4</sup>. Strada per strada, casa per casa (particolarmente accurato per i sestieri meridionali della città) i Deputati segnarono tutti i familiari, compresi morti o infetti trasferiti nei lazzaretti, attività lavorativa del capofamiglia (da cui si poteva evincere se vi era la necessità o meno di fornire generi alimentari durante la quarantena) e il numero dei componenti il nucleo (bocche). Analogamente a quanto avveniva in città abbiamo poi per Vicariati e Potesterie limitrofe a Pisa degli elenchi con il numero dei morti di ogni parrocchia. Se per Pisa è stato quindi possibile georeferenziare con un margine di errore piuttosto contenuto i dati raccolti dai Deputati di Sanità, il dato extraurbano, ancorato invece alle parrocchie è certamente, dal punto di vista di un'analisi geomedica, molto più interessante (figura 1).

Le fonti digitali utilizzate per georeferenziare i dati archivistici sono quelle prodotte dal SITA (Sistema Informativo Territoriale e Ambientale) della Regione Toscana e sono dati vettoriali e raster.<sup>4</sup> I layer vettoriali impiegati sono i limiti amministrativi comunali, le curve di livello -da cui si è ottenuto un DTM (Digital Terrain Model)- la viabilità; mentre i file raster sono quelli relativi alle Sezioni della CTR in scala 1:10.000 che coprono il centro storico di Pisa e che sono stati caricati nel nostro GIS, abbiamo utilizzato il software open source Quantum GIS, tramite protocolli WMS (Web Map Service).

---

3 ASP, Miscellanea Manoscritti di Proprietà Libera, 8. I. Arrosti, Croniche di Pisa.

4 ASP, Commissariato Appendice, 73.

1631

Scacco di San Paolo a Ripa d'Arno

|    |                                    | Borghi | Terza | Cap. C. C. |
|----|------------------------------------|--------|-------|------------|
| 1. | Borghi di San Paolo                | 3.     | 1.    |            |
|    | Borghi a Borghi                    |        |       |            |
| B  | 2. Casa della Spedale di San Paolo | 1.     | 2.    |            |
|    | 3. Cap. C. C.                      |        |       |            |
|    | 4. Farnel d'ortolano               | 5.     | 3.    |            |
|    | 5. Farnel di Borghi d'ortolano     | 2.     | 1.    | 2.         |
| B  | 6. Farnel di Borghi d'ortolano     | 2.     | 2.    | 4.         |
|    | 7. Farnel di Borghi d'ortolano     |        |       |            |
| B  | 8. Farnel di Borghi d'ortolano     | 2.     | 1.    | 2.         |
|    | 9. Farnel di Borghi d'ortolano     | 2.     | 1.    | 2.         |
| B  | 10. Farnel di Borghi d'ortolano    | 1.     | 1.    | 1.         |
|    | 11. Farnel di Borghi d'ortolano    | 1.     | 1.    | 2.         |
| B  | 12. Farnel di Borghi d'ortolano    | 2.     | 1.    |            |
|    | 13. Farnel di Borghi d'ortolano    | 1.     | 1.    | 1.         |
| B  | 14. Farnel di Borghi d'ortolano    | 1.     | 3.    | 1.         |
|    | 15. Farnel di Borghi d'ortolano    | 5.     | 2.    | 2.         |
|    | 16. Farnel di Borghi d'ortolano    |        | 1.    |            |
| B  | 17. Farnel di Borghi d'ortolano    | 2.     |       |            |
|    | 18. Farnel di Borghi d'ortolano    | 2.     | 1.    | 2.         |
| B  | 19. Farnel di Borghi d'ortolano    | 3.     | 4.    | 4.         |
|    | 20. Farnel di Borghi d'ortolano    | 3.     | 3.    | 1.         |
| B  | 21. Farnel di Borghi d'ortolano    | 1.     | 2.    |            |
|    | 22. Farnel di Borghi d'ortolano    | 1.     | 2.    |            |
| B  | 23. Farnel di Borghi d'ortolano    | 1.     | 3.    |            |
|    | 24. Farnel di Borghi d'ortolano    | 1.     | 1.    |            |
| B  | 25. Farnel di Borghi d'ortolano    | 1.     |       | 2.         |
|    | 26. Farnel di Borghi d'ortolano    | 3.     | 2.    | 2.         |
|    |                                    | 46.    | 42.   | 21.        |
|    |                                    |        |       | 4.         |

Figura 1: elenco delle persone messe in quarantena in San Paolo a Ripa d'Arno.

Queste diverse serie di dati hanno pertanto dato origine a due distinti geodatabase. Un primo layer (entrambi gli shapefile creati sono file puntuali) in cui ogni nucleo familiare è stato georeferenziato rispetto a un edificio della Pisa del XVII secolo, e un secondo strato informativo in cui il numero dei decessi registrati fuori città è stato agganciato ad una parrocchia (edificio). Se nel primo caso l'informazione è molto più ricca e si presta anche, e forse soprattutto, ad elaborazioni più propriamente demografico-economico, la fonte extraurbana è più interessante sotto il profilo della geografia medica<sup>5</sup>.

Ciò che emerge con chiarezza analizzando questo secondo geodatabase è che i picchi di mortalità causati dalla peste nel pisano settentrionale, ciò vale per il 1630 come forse per altre epoche del passato, si concentrò nelle valli montane ove la popolazione era tutta concentrata nei borghi e lungo le principali vie di comunicazione (figura 2).

5 Trooskin S, Hadler J, St. Louis T, and Navarro V (2005). Geospatial analysis of hepatitis C in Connecticut: a novel application of a public health tool. *Public Health*, 119(11), 1042-7; Rushton G, Elmes G, McMaster R (2003). Considerations for improving geographic information research in public health. *URISA Journal*, 12(2), 31-49; O'Carroll, P.W. (2003). Introduction to public health informatics. In P.W. O'Carroll, Y.A. Yasnoff, M.E. Ward, L.H. Ripp, and E.L. Martin (Ed.), *Public Health Informatics* (pp. 1-15). New York, NY: Springer; Hanchette, C.L. (2003). Geographic Information Systems. In P.W. O'Carroll, Y.A. Yasnoff, M.E. Ward, L.H. Ripp, and E.L. Martin (Ed.), *Public Health Informatics* (pp. 431-466). New York, NY: Springer; Yasnoff, W.A. and Miller, P.L. (2003). Decision Support and Expert Systems in Public Health. In P.W. O'Carroll, Y.A. Yasnoff, M.E. Ward, L.H. Ripp, and E.L. Martin (Ed.), *Public Health Informatics* (pp. 494-512). New York, NY: Springer; E. Brianti, J. B. Malone, J. C. McCarroll, M. Bernardi, M. Drigo, K. Gruszynski, Minimum medical GIS database (MMDb) per l'Europa. *Parassitologia*. 2004 June; 46(1-2): 67-70.

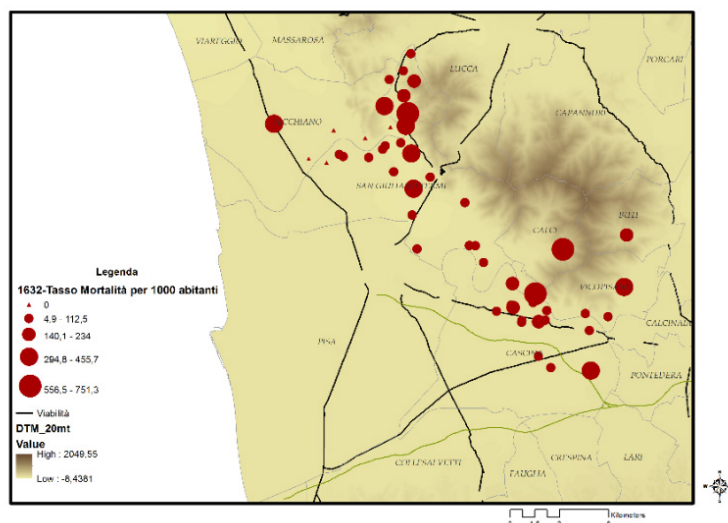


Figura 2: tasso di mortalità

Interrogando la nostra banca dati geografica emerge poi come i tassi di mortalità siano molto più bassi, in taluni casi nulli, in quei luoghi caratterizzati da una prevalenza di case sparse - che è uno degli elementi distintivi dalla mezzadria poderale Toscana.

Le ragioni di queste concentrazioni dei tassi di mortalità in aree rurali sembrano pertanto essere ragionevolmente connesse con la concentrazione di case (e quindi di popolazione) e con la viabilità principale e quella periferica che termina in zone isolate. Più complesso il tema della “misura” dei decessi causati dalla peste per il centro cittadino per i quali non ci sono dati aggregati realizzati dall’Ufficio di Sanità.

### 3 INTRODUZIONE

L’utilizzo di software GIS in ricerche storico-mediche è oggi un imprescindibile strumento di fondamentale necessità per studiare i comportamenti e ipotizzare le dinamiche comportamentali di fenomeni sanitari. Se poi, come dimostrato in questo studio, al generico teorema dello studio della storia quale strumento per conoscere e comprendere il presente, e finanche secondo alcuni il futuro (ciclicità vi-chiana della storia), si affianca uno “strumento di misura” che poggia sullo studio e la rappresentazione della Terra, ecco che lo studio di malattie epidemiche del passato (come di tutte quelle che hanno lasciato traccia nei documenti), permette di costruire e indagare modelli di diffusione e magari, studiando il concatenamento di elementi storico-geografico-medico, di elaborare efficaci strategie di contenimento nella diffusione di infezioni.

Il caso studio che qui abbiamo presentato è in tal senso estremamente esplicativo di come utilizzando fonti storico-archivistico e GIS si possano elaborare modelli tanto urbani che extraurbani e di come sia possibile far interloquire proficuamente tra loro discipline geografiche, storico-medico con sistemi informativi ai fini di produrre conoscenze complesse imprescindibili per settori strategici quali studi di geo-medicina.

## BIBLIOGRAFIA

1. A. Feroci, *La Peste Bubbonica in Pisa nel Medio Evo e nel 1630*, Tipografia Vannucchi, Pisa, 1892, pp. 48-196.
2. Archivio di Stato di Pisa (ASP), P. Tronci, *Memorie in tempo di mal contagio*, Manoscritti 8, c. 8.
3. G. Tanti, *La peste a Pisa nel 1630*, in *Strutture sanitarie a Pisa. Contributi alla storia di una città. Secoli XIII-XIX*, Composit, Pisa, 1986.
4. ASP, *Commissariato Appendice*, 73.
5. ASP, *Commissariato Appendice*, 72.
6. ASP, *Miscellanea Manoscritti di Proprietà Libera*, 8. I. Arrostiti, *Croniche di Pisa*.
7. M. Grava, *L'anno che era la peste a Pisa: 1630*, in E. Salvatori (a cura di), *Studi di Storia degli Insediamenti in onore di Gabriella Garzella*, Pisa, 2014.
8. P. Mogorovich, P. Mussio, *Automazione del Sistema Informativo territoriale. Elaborazione Automatica dei Dati Geografici*, Masson 1988, vol. 2, pp. 503-508.

# **La rete delle Biblioteche depositarie delle Nazioni Unite e la sua evoluzione in “Open Community”**

Deborah Grbac

Biblioteca d'Ateneo, Università Cattolica del Sacro Cuore (Milano) – [deborah.grbac \[at\] unicatt.it](mailto:deborah.grbac@unicatt.it)

## **SINTESI**

La recente evoluzione del Sistema delle Biblioteche depositarie delle Nazioni Unite e del relativo programma a sostegno, anche in seguito alla pandemia Covid-19, ha trasformato il Sistema in una vera e propria comunità on-line volta alla diffusione, scambio e riuso di informazioni, documenti e materiali in formato aperto.

Le biblioteche depositarie si sono ritrovate ad essere, non solo le custodi dell'eredità culturale rappresentata dalla documentazione delle Nazioni Unite ricevuta negli anni, ma anche le dispensatrici della conoscenza sulle Nazioni Unite e dei prodotti della proprietà intellettuale dell'organizzazione internazionale. In un'era in cui l'attività di selezione dell'informazione disponibile in formato digitale è diventata sempre più importante e strumento di lotta alla disinformazione, le biblioteche depositarie possono garantire tale attività in modo professionale, proprio a partire dall'eredità culturale, che negli anni hanno costruito e conservato, sulla quale i bibliotecari si sono formati ed hanno acquisito competenze che mettono a disposizione delle comunità locali di studiosi, ricercatori, decisori politici e cittadini.

## **PAROLE CHIAVE**

documentazione delle Nazioni Unite, biblioteca Dag Hammarskjöld di New York, sistema delle biblioteche depositarie delle Nazioni Unite.

## **INTERVENTO**

### **1 INTRODUZIONE**

Il Sistema delle Biblioteche depositarie delle Nazioni Unite è stato ufficialmente istituito nel 1947 e la sua evoluzione accompagna i settantacinque anni di storia dell'organizzazione internazionale. L'intento del Programma era quello di creare una rete mondiale di biblioteche volta a ricevere documenti ufficiali e pubblicazioni delle Nazioni Unite, con l'obbligo statutario di raccolta, catalogazione, conservazione e messa a disposizione al pubblico (deposito cartaceo).

Dal momento della revisione del programma nel 2015 e in particolare in seguito alla crisi mondiale causata dall'epidemia del Covid-19, la rete delle biblioteche depositarie (ad oggi 351 biblioteche sparse in 135 Paesi e nello Stato di Palestina) si è trasformata di fatto in una "Open Community" volta alla conservazione e trasmissione dell'eredità culturale comune (la conoscenza specialistica sulle Nazioni Unite e la produzione intellettuale delle Nazioni Unite) e alla costante condivisione on-line di informazioni e di contenuti sull'organizzazione e la sua attività, disponibili, usufruibili, riutilizzabili e condivisibili in regime di dominio pubblico per i documenti ufficiali e con preghiera di citazione della fonte per il materiale promozionale e i contenuti prodotti dalla DHL, Dag Hammarskjöld Library, la Biblioteca delle Nazioni Unite di New York che coordina il Sistema delle biblioteche depositarie.

## **2 L'EVOLUZIONE NEL TEMPO DEL PROGRAMMA DELLE BIBLIOTECHE DEPOSITARIE DELLA NAZIONI UNITE**

Come accennato sopra, il lancio ufficiale del programma risale al 1947, ma, come ci ricorda il bibliotecario e studioso James Church, già nei documenti della Società delle Nazioni si faceva menzione all'esistenza di un programma di biblioteche depositarie a partire dagli anni venti che garantiva la distribuzione di documenti e pubblicazioni della Società presso biblioteche dislocate in diverse Nazioni appartenenti all'organizzazione [2].

Il programma delle biblioteche depositarie si era posto come obiettivo di creare una sorta di copia conforme della collezione dei documenti e pubblicazioni delle Nazioni Unite presente a New York e Ginevra, sedi delle principali biblioteche del sistema, distribuita in diverse parti del mondo, in modo da garantirne la conservazione e il duplicato della documentazione da preservare, ad esempio come si può leggere in un documento interno al Programma emesso dopo l'11 settembre 2001, in caso di un attentato terroristico alla collezione della Biblioteca DHL con sede presso il quartiere generale delle Nazioni Unite a New York.

Per decenni questa collezione in copia presso diverse biblioteche nel mondo ha avuto forma cartacea; la spedizione avveniva per posta e per l'accesso occorreva recarsi presso una delle biblioteche depositarie presenti nella Nazione o nel Continente geografico di appartenenza. In qualche modo la condivisione delle informazioni era chiusa ad una cerchia limitata di utenti. Con l'avvento di internet, ecco che le Nazioni Unite cominciano a dotarsi di data-base in formato digitale. Il primo è stato ODS, Official Documents System creato nel 1993 per i documenti ufficiali delle Nazioni Unite e dal 2005 in libero accesso<sup>6</sup> e il sito web della Sezione Trattati dell'Ufficio Legale delle Nazioni Unite chiamato United Nations Treaty Collection che consente di accedere ai testi della pubblicazione a pagamento "Treaty series", attraverso la consultazione di un data-base, interrogabile tramite query<sup>7</sup>.

Il passaggio dalla collezione cartacea a quella digitale è avvenuto anche in seguito ad un disastro naturale: quando l'uragano Sandy nel 2012 ha distrutto la stamperia delle Nazioni Unite a New York, nel 2013 il Consiglio delle Pubblicazioni ha deciso il passaggio dalla distribuzione cartacea del materiale a quella digitale anche per ridurre la produzione di carta e le spese di spedizione.

---

6 Sito ufficiale:  
<https://documents.un.org/prod/ods.nsf/home.xsp>

7 Sito ufficiale:  
<https://treaties.un.org/>



Sono seguiti nel tempo diversi progetti che hanno portato alla creazione di strumenti per la condivisione di documenti e di pubblicazioni in versione digitale: oltre alla banca dati ODS per l'accesso aperto ai documenti delle Nazioni Unite nati in formato digitale, ovvero dal 1993, dal 2016 esistono anche l'archivio istituzionale per le pubblicazioni a pagamento Un-i-Library, disponibile ad accesso aperto in sola lettura<sup>8</sup> e la Digital Library, lanciata nel maggio del 2017. La Digital Library ha sostituito il precedente catalogo online che raccoglieva i dati bibliografici di tutta la documentazione delle Nazioni Unite: UNBISnet. Essa contiene i documenti ufficiali (presenti anche in ODS), le pubblicazioni ad accesso aperto e i documenti e le pubblicazioni oggetto di progetti di digitalizzazione (ovvero pubblicati prima del 1993 e scansionati, i documenti scansionati sono disponibili anche via ODS), il data-base dei discorsi tenuti nelle riunioni dei principali organi delle Nazioni Unite, i dati delle votazioni di tutte le risoluzioni adottate dall'Assemblea generale e dal Consiglio di Sicurezza e le carte geografiche allegate ai trattati internazionali<sup>9</sup>.

Anche i compiti delle biblioteche depositarie sono cambiati nel tempo. Al momento dell'istituzione del sistema gli obblighi erano quelli di mantenere la collezione ricevuta in ordine, attraverso la catalogazione, la conservazione e l'accessibilità al pubblico. Tutto questo competeva a personale bibliotecario qualificato, che doveva assicurare l'accesso gratuito a ricercatori e cittadini in orari adeguati<sup>10</sup>.

Con il passaggio dalla collezione cartacea a quella digitale nel 2014 si è aggiunto anche il compito di garantire un sostegno alla ricerca di informazioni e documentazione sulle Nazioni Unite da parte di personale competente appositamente formato (United Nations specialists). Nel 2016 all'interno del concetto di "deposito digitale primario" è stata sottolineata l'importanza della professionalità locale al fine di rendere la conoscenza specialistica sulle Nazioni Unite e la produzione intellettuale delle Nazioni Unite comprensibile e moltiplicabile attraverso la produzione scientifica locale [1].

Inoltre, in settantacinque anni di storia delle Nazioni Unite, l'eredità culturale costituita dalla collezione cartacea conservata in diversi luoghi sparsi nel mondo continua a mantenere, anche nel passaggio al formato digitale, la sua funzione di controllo dell'autenticità dei documenti ufficiali [3], e funge da copia del patrimonio non ancora disponibile in versione digitale (i progetti di digitalizzazione sono tutt'ora in corso).

---

8 Sito ufficiale (necessaria la sottoscrizione per l'accesso):

<https://www.un-ilibrary.org/>

9 Sito ufficiale e guida alla ricerca:

<https://digitallibrary.un.org/>

<https://research.un.org/en/digitallibrary>

10 L'ultima revisione ufficiale degli obblighi e delle istruzioni per le biblioteche depositarie è del 1995

<https://digitallibrary.un.org/record/190585>

<https://digitallibrary.un.org/record/179201>

### 3 LO STATO ATTUALE DEL PROGRAMMA DELLE BIBLIOTECHE DEPOSITARIE

Un programma che nel 2013 sembrava essere in crisi a causa del complesso passaggio al digitale, si è visto rilanciato grazie al processo di revisione avvenuto nel 2015-2016 e trasformato a partire dalla crisi mondiale del Covid-19. Poiché, a causa della pandemia, le persone sono state costrette a rimanere a casa, si è rivelato necessario riformulare le relazioni degli appartenenti al Sistema attraverso lo scambio di informazioni e materiali via mezzi telematici. A sei mesi dall'inizio della crisi si può dire che il Sistema delle biblioteche depositarie si è trasformato in una vera e propria comunità aperta di condivisione della conoscenza sulle e del patrimonio intellettuale delle Nazioni Unite.

La crisi ha di fatto sancito il duplice passaggio al riconoscimento del “deposito elettronico” in primis della collezione, rispetto al tradizionale deposito cartaceo, comunque sempre in vigore, e all'accessibilità aperta e di pubblico dominio ai documenti e alle pubblicazioni delle Nazioni Unite, rispetto alle risorse sottoscritte. La Biblioteca DHL, dove operano i responsabili del programma delle Biblioteche depositarie, ha fornito sempre più strumenti digitale ad accesso aperto, ovvero tutti i contenuti della Digital Library, le guide di ricerca dei contenuti e il materiale promozionale ed informativo relativo ad eventi celebrativi e/o alle attività delle Nazioni Unite.

In questi mesi la comunità delle biblioteche depositarie è stata costantemente rifornita di:

- informazione e contenuti trasmessi via email e twitter su iniziative ed anniversari di celebrazione delle Nazioni Unite; fra l'altro durante l'emergenza Covid-19 cadeva l'anniversario dei 75 anni della firma della Carta delle Nazioni Unite;
- materiale di formazione per i bibliotecari specializzati delle biblioteche depositarie (webinars registrati durante il periodo della crisi), attraverso la creazione di una apposita sezione di formazione nel sito del Programma<sup>11</sup>;
- guide alla ricerca per gli utenti delle biblioteche depositarie costantemente aggiornate su tematiche di interesse e di attualità delle Nazioni Unite e presentate attraverso i più nuovi ritrovati delle tecnologie di visualizzazione e presentazione;
- costante aggiornamento sui progetti di digitalizzazione, con la produzione di apposite guide alla ricerca;
- informazione, pubblicizzazione e condivisione delle attività delle diverse biblioteche depositarie<sup>12</sup>;
- continuo supporto informativo e di sostegno alla valorizzazione delle attività delle biblioteche depositarie (ad esempio, fornendo documentazione per eventi organizzati presso le biblioteche depositarie come anniversari celebrativi locali);

---

11 <https://library.un.org/content/training-resources-depository-libraries>

12 <https://findery.com/UNDepLibs/notemaps/un-depository-libraries>

- colloqui via videoconferenze che hanno permesso una maggiore vicinanza, nonostante la lontananza, con la Biblioteca di New York;
- Webinar live con possibilità di porre domande come quello sul tema del ruolo delle biblioteche nel combattere la disinformazione, tenutosi il 13 agosto 2020;
- l'annuncio in prospettiva di ulteriori corsi di formazione online sulla tematica della lotta alla disinformazione.

Mai come in questo periodo il Sistema delle biblioteche depositarie ha fatto rete attorno alla Biblioteca coordinatrice di New York e i bibliotecari che lavorano presso le biblioteche depositarie si sono sentiti parte di una “comunità aperta”, liberi di accedere, utilizzare, condividere la documentazione in pubblico dominio e i materiali di supporto alla ricerca prodotti dietro semplice citazione della fonte. Senza dimenticare la possibilità di accedere alla loro collezione conservata sul posto e pronti ad offrire tutto questo alle comunità locali di studiosi, ricercatori, a coloro che scrivono o devono rendere conto in materia di questioni globali o che attivamente operano per rendere possibile la collaborazione globale.

## 4 RINGRAZIAMENTI

Si ringraziano la Coordinatrice delle relazioni esterne del Programma delle Biblioteche depositarie delle Nazioni Unite, la Sig.ra Ramona Kohrs e il Bibliotecario senior e Assistente al Programma delle Biblioteche depositarie delle Nazioni Unite, il Sig. Armando da Silva.

## BIBLIOGRAFIA

1. Aldis, Sherri, e Ramona Kohrs. «Access to UN Information in the Digital Era: Reengineering the UN Depository Libraries Programme». Columbus (Ohio), 2016. <http://library.ifla.org/id/eprint/1429>.
2. Church, James. «Documents without borders. The Once and Future UN Depository». *A Quarterly Journal Of Government Information Practice and Perspective* 43, n. 3 (2015). <https://escholarship.org/uc/item/5fm3k568>.
3. Church, Jim. «The Embattled UN Depository». *Documents to the People* 44, n. 3 (2016). <https://escholarship.org/uc/item/4xd8k0p9>.

# Analisi del sentiment delle Confessioni di Sant'Agostino

Angelica Lo Duca<sup>1</sup>, Andrea Marchetti<sup>1</sup>  
IIT CNR – angelica.loduca [at] iit.cnr.it  
IIT CNR – andrea.marchetti [at] iit.cnr.it

## SINTESI

Le Confessioni di Sant'Agostino costituiscono uno dei capolavori della letteratura cristiana di ogni tempo. In quest'opera, costituita da 13 libri, il celebre scrittore descrive il proprio percorso di conversione a Dio. Il presente articolo applica tecniche di sentiment analysis a ciascun libro, attraverso due tecniche, una basata sull'Apprendimento Supervisionato e l'altra sull'Apprendimento Non Supervisionato, al fine di capire lo stato d'animo dell'autore man mano che il percorso di conversione arriva a compimento. Come risultato, si può evincere che mentre nei primi libri i sentimenti positivi sono comparabili a quelli negativi, negli ultimi libri, i sentimenti negativi tendono totalmente a scomparire.

## PAROLE CHIAVE

sentiment analysis, digital humanities, Sant'Agostino, Confessioni.

## INTERVENTO

### 1 INTRODUZIONE

Le Confessioni di Sant'Agostino sono uno dei maggiori capolavori della letteratura cristiana. Opera scritta tra il 397 e il 400 d.C., le Confessioni riflettono la vita dell'autore. Esse si suddividono in 13 libri, nei quali l'autore racconta il proprio cammino di conversione a Dio, descrivendo gli eventi che lo hanno portato a diventare cristiano, a partire dall'infanzia. Mentre i primi nove libri sono totalmente autobiografici, gli ultimi quattro riguardano diverse concezioni filosofiche, come l'origine dell'uomo e l'essenza del tempo. L'evento della conversione avviene propriamente nel nono libro.

L'obiettivo di questo lavoro consiste nell'analisi del sentiment nell'ambito dei 13 libri che costituiscono le Confessioni di Sant'Agostino, al fine di stabilire se esista o meno una relazione con il percorso della conversione<sup>13</sup>.

---

13 Il software implementato per eseguire le analisi può essere scaricato dal seguente repository Github <https://github.com/alod83/papers/tree/master/aiucd2021>

Nello specifico, sono descritti due esperimenti di analisi del sentiment, uno basato su tecniche di Apprendimento Supervisionato e l'altro su tecniche di Apprendimento Non Supervisionato. Sebbene i risultati delle due tecniche siano leggermente diversi, si conferma il trend che negli ultimi due libri, prevale il sentiment positivo. Ciò è in linea con il processo di conversione descritto nel corso dei libri.

## 2 STATO DELL'ARTE

Il problema della Sentiment Analysis di un testo può essere affrontato sia utilizzando tecniche di Apprendimento Supervisionato sia tecniche di Apprendimento Non Supervisionato [7]. In generale le prime hanno prestazioni migliori rispetto alle seconde [12]. L'Apprendimento Supervisionato (AS) è una tecnica di Machine Learning (ML) che si basa sull'annotazione iniziale di un dataset di riferimento, al fine di identificare un insieme finito di classi di uscita. Nell'ambito della Sentiment Analysis, le tecniche di AS maggiormente utilizzate sono basate su algoritmi classici di ML, come Support Vector Machines, Naive Bayes [1, 3, 12] e Deep Convolutional Neural Networks [6, 13].

L'Apprendimento Non Supervisionato (ANS) è una tecnica di Machine Learning che cerca di identificare le caratteristiche comuni dei dati forniti in ingresso, al fine di classificare i dati futuri forniti in ingresso. Non è prevista alcuna annotazione iniziale dei dati, in quanto le classi di uscita sono calcolate automaticamente dall'algoritmo. Nell'ambito della Sentiment Analysis, le tecniche di ANS si basano su k-means [9] e LDA [8].

In letteratura esistono altre tecniche per effettuare la Sentiment Analysis di un testo, basate sull'uso di lessici standard, come ad esempio come SentiWordNet [2], AFINN [11], o lessici emotivi, come ad esempio NRC/Emolex [10] e Vader [5], senza contare senza contare i dizionari creati tramite word embeddings [4].

## 3 METODOLOGIA

Al fine di definire il sentiment di ogni libro, è stata considerata la versione in inglese del testo, fornita dal progetto Gutenberg<sup>14</sup>. Ogni libro è stato analizzato separatamente e diviso in frasi. La Tabella 1 mostra alcune statistiche per ogni libro, mentre la Fig. 1 mostra la word cloud relativa alle parole maggiormente utilizzate nell'intera opera, da cui sono state eliminate le stopwords.

| Libro      | 1     | 2     | 3     | 4     | 5     | 6     | 7     | 8     | 9     | 10    | 11    | 12    | 13    |
|------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| #frasi     | 186   | 98    | 119   | 182   | 153   | 178   | 185   | 208   | 207   | 455   | 258   | 176   | 303   |
| #caratteri | 37735 | 20500 | 30011 | 37217 | 36606 | 42798 | 45414 | 43764 | 47943 | 86515 | 49733 | 55939 | 68395 |
| #parole    | 7055  | 3781  | 5466  | 6975  | 6695  | 7838  | 8409  | 8123  | 8819  | 16315 | 9501  | 10205 | 12598 |

Tabella 1: statistiche per ogni libro

14 La versione in inglese del testo è disponibile al seguente link

<https://www.gutenberg.org/files/3296/3296-h/3296-h.htm>



A partire dallo score di ogni frase, sono stati calcolati gli indici di negatività, positività e neutralità. La Fig. 2 mostra il risultato del processo di analisi.

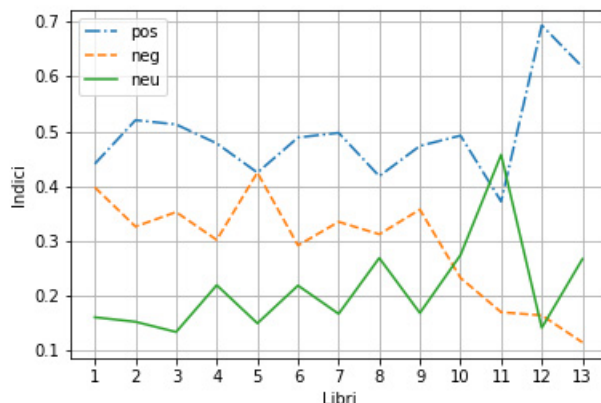


Figura 2: Indici di positività, negatività e neutralità nei vari libri, utilizzando il lessico AFINN

Nell'ambito dell'AS, è stato utilizzato il seguente procedimento: ogni frase del primo libro è stata annotata, attribuendo un sentiment positivo, nullo, o negativo. In totale, sono state ottenute 82 frasi con sentiment positivo, 74 con sentiment negativo e 30 con sentiment nullo. Del totale, sono state considerate solo le frasi con sentiment positivo e negativo, in modo da costruire il dataset di riferimento (156 campioni), utilizzato per allenare l'algoritmo di AS. La Tabella 2 mostra alcuni campioni del dataset di riferimento: la frase costituisce l'input e il sentiment l'output dell'algoritmo.

| Frases   | Sentiment |
|--|-----------|
| Great art Thou, O Lord, and greatly to be praised; great is Thy power, and Thy wisdom infinite.  | 1         |
| And Thee would man praise; man, but a particle of Thy creation; man, that bears about him his mortality, the witness of his sin, the witness that Thou resistest the proud: yet would man praise Thee; he, but a particle of Thy creation. | 1         |
| For if I go down into hell, Thou art there.  | 0         |
| I could not be then, O my God, could not be at all, wert Thou not in me; or, rather, unless I were in Thee, of whom are all things, by whom are all things, in whom are all things? Even so, Lord, even so.                                | 1         |

Tabella 2: Alcuni campioni del dataset usato per allenare l'algoritmo di AS.

Il dataset di riferimento è stato diviso in due parti, training set, contenente il 75% dei campioni e il test set, contenente il rimanente 25%. Ciascuna frase è stata divisa in token al fine di costruire la matrice dei token. Sono stati considerati solo i token con frequenza minima pari a 5 per cui il numero dei nomi contenuti nella matrice dei token è pari a 189. Sono stati condotti due esperimenti, uno senza gli n-grammi, l'altro che considera n-grammi di dimensione 2. Come algoritmo di AS è stato considerato il Linear Support Vector Machines. La Fig. 3 mostra i risultati per l'AS, senza e con l'uso di n-grammi di dimensione 2. Nei due casi i risultati sono abbastanza simili, con una differenza nel valore dello Score AUC, pari a 0.69 nel primo caso e 0.82 nel secondo.

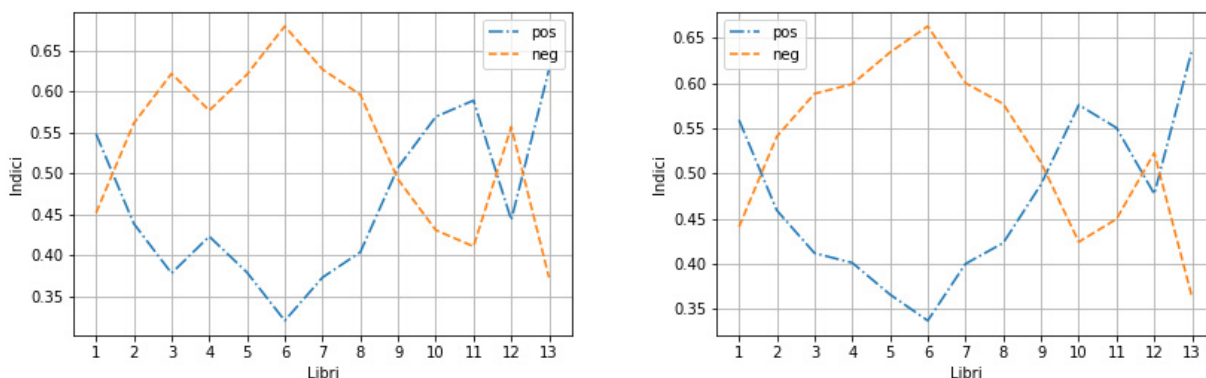


Figura 3: Indici di positività e negatività nei vari libri, utilizzando l'algoritmo Linear Support Vector Machines, senza (a) e con l'uso di n-grammi (b)

## 5 DISCUSSIONE E SVILUPPI FUTURI

Ad una prima analisi, è interessante notare come i due esperimenti diano risultati abbastanza diversi. Infatti, mentre nell'esperimento basato su ANS il sentiment positivo è sempre maggiore di quello negativo, nell'esperimento basato su AS, ciò è vero solo nei capitoli 9, 10, 11 e 13. In realtà, entrambi gli esperimenti mostrano come nei primi libri dell'opera il sentiment negativo sia molto alto, mentre negli ultimi libri vada scemando, per lasciare il posto ad un sentiment positivo molto alto. In conclusione, si può dire che entrambi gli esperimenti riflettono a grandi linee il processo di conversione dell'autore, che passa da uno stato d'animo negativo (all'inizio del libro) ad uno stato d'animo positivo (alla fine del libro).

Il presente lavoro costituisce un punto di partenza nell'analisi di questa splendida opera letteraria. Come sviluppo futuro, si potrebbe pensare di analizzare il collegamento tra il sentiment descritto in ogni libro e il contenuto del libro stesso, con un focus particolare alle fasi della vita dell'autore.



## BIBLIOGRAFIA

1. Al-Hadhrami, Suheer, Norah Al-Fassam, e Hafida Benhidour. «Sentiment Analysis Of English Tweets: A Comparative Study of Supervised and Unsupervised Approaches». In *2019 2nd International Conference on Computer Applications & Information Security (ICCAIS)*, 1–5. IEEE, 2019.
2. Baccianella, Stefano, Andrea Esuli, e Fabrizio Sebastiani. «Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining.» In *Lrec*, 10:2200–2204, 2010.
3. Durant, Kathleen T., e Michael D. Smith. «Mining sentiment classification from political web logs». In *Proceedings of Workshop on Web Mining and Web Usage Analysis of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (WebKDD-2006)*, Philadelphia, PA, 10, 2006.
4. Giatsoglou, Maria, Manolis G. Vozalis, Konstantinos Diamantaras, Athena Vakali, George Sarrigiannidis, e Konstantinos Ch Chatzisavvas. «Sentiment analysis leveraging emotions and word embeddings». *Expert Systems with Applications* 69 (2017): 214–224.
5. Hutto, Clayton, e Eric Gilbert. «Vader: A parsimonious rule-based model for sentiment analysis of social media text». In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 8, 2014. <http://comp.social.gatech.edu/papers/icwsm14>.
6. Jianqiang, Zhao, Gui Xiaolin, e Zhang Xuejun. «Deep convolution neural networks for twitter sentiment analysis». *IEEE Access* 6 (2018): 23253–23260. <https://doi.org/10.1109/ACCESS.2017.2776930>.
7. Kim, Evgeny, e Roman Klinger. «A survey on sentiment and emotion analysis for computational literary studies». *arXiv preprint arXiv:1808.03137*, 2018.
8. Li, Fangtao, Minlie Huang, e Xiaoyan Zhu. «Sentiment analysis with global topics and local dependency». In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 24, 2010. <https://ojs.aaai.org/index.php/AAAI/article/view/7523>.
9. Li, Nan, e Desheng Dash Wu. «Using text mining and sentiment analysis for online forums hotspot detection and forecast». *Decision support systems* 48, n. 2 (2010): 354–368.
10. Mohammad, Saif M., e Peter D. Turney. «Crowdsourcing a word–emotion association lexicon». *Computational Intelligence* 29, n. 3 (2013): 436–465. <https://doi.org/10.1111/j.1467-8640.2012.00460.x>.
11. Nielsen, Finn \AArup. «afinn project», 2017.
12. Pang, Bo, Lillian Lee, e Shivakumar Vaithyanathan. «Thumbs up? Sentiment classification using machine learning techniques». *arXiv preprint cs/0205070* 10 (2002): 79–86.
13. Severyn, Aliaksei, e Alessandro Moschitti. «Unitn: Training deep convolutional neural network for twitter sentiment classification». In *Proceedings of the 9th international workshop on semantic evaluation (SemEval 2015)*, 464–469, 2015.

# Hashtags as an information source. Analyzing tweets to map La Terra dei Fuochi

author

## ABSTRACT

In this paper, we present the UNIOR Eye (Earth your estate) corpus made up of 228,412 tweets dealing with environmental crimes. The core of this research is the extraction of information from tweets according to hashtags and their frequency. Then, in an unsupervised scenario, the K-means clustering algorithm is used to automatically cluster the different contents of the tweets by keeping the hashtags hidden. The analysis of each cluster at the lexical level allows us to identify i) alert tweets, ii) political background tweets and iii) personal opinion tweets. Once the alerts tweets about La Terra dei Fuochi are identified (precisely 36,207), we first extract the exact location (preceded by a hashtag) where the crime took place and then establish the type of crime. This information is displayed on a map with the use of Carto<sup>1</sup>.

## KEYWORDS

natural language processing, corpus linguistics, environmental crimes

## POSTER

### 1 INTRODUCTION

The aim of our research is to study the frequency and co-occurrence of hashtags in order to gather information related to the contents of tweets and then to monitor natural human-caused disasters. More specifically, we focus on *La Terra dei Fuochi* (literally The Land of Fires), a large area between Naples and Caserta (in the north of the Campania region, Italy) which has been plagued for about fifty years by illegal toxic wastes routinely dumped by criminal organizations to make space for new ones [10]. Over the years, the public awareness about what happens in *La Terra dei Fuochi* has increased and common people have started playing a role in the fight against illegal and toxic waste disposal. In particular, they have started using social media to share real-time information [1] and report the presence of illegally disposed toxic wastes.

---

1 <https://carto.com/>

Twitter, especially, is one of the most popular microblogging and represents a fundamental channel for sharing news and opinions related to a specific topic thanks to hashtags, namely words or unspaced phrases preceded by #, that are a valuable resource for identifying and aggregating messages and posts related to a specific subject [2]. Moreover, the fact that tweets contain information shared by ordinary people helps to assess the real situation of events [4]. Therefore, by monitoring tweets about criminal events in a specific place, some crimes could be detected immediately [5]. For this purpose, we present an unsupervised approach that hides hashtags from the tweets, in order to classify the textual contents of tweets thanks to the K-means clustering algorithm and so to discriminate the clusters related to i) alert tweets, ii) political background tweets and iii) personal opinion tweets. This research is carried out in the framework of the C4E (Crowd for the Environment) project. The paper is organized as follows: in Section 2 we discuss Related Work, in Section 3 we present the *UNIOR Eye* corpus, in Section 4 we show the case study and the map of environmental crime places detected in the corpus. Conclusions are in Section 5 along with Future Work.

## 2 RELATED WORK

Many researchers propose event detection approaches that monitor Twitter data and determine whether special events, such as accidents, extreme weather conditions, earthquakes or crimes, occur by analyzing hashtags. [12] propose an event detection approach based on hashtags in tweets to collect information related to the Paris attack from 13 to 17 November 2015. [11] examine how the co-occurrence of specific hashtags leads to the virality of information during the Occupy Wall Street movement in 2011. Through network analysis, the authors identify popular hashtag types and examine co-occurrence patterns during the two days of the movement. This study shows that participants in the self-organised movement make a strategic use of hashtags to better spread their message and drive the viral movement. [9] propose a method for data mining on Twitter to retrieve messages about floods and tropical cyclones. The scholars describe an automated process for collecting hashtags which are strongly related to a specific event and demonstrate that hashtags are good markers for separating similar and simultaneous events. The method uses disaster databases to find the location of an event and to estimate the area of impact.

[3] construct a dataset of more than 67,000 tweets related to natural disasters of various types. The corpus was then annotated with the hashtags from the tweets, where present. By using this large dataset, the scholars further investigate a deep learning model for keyphrase extraction from general tweets, namely a joint-layer Long-Short Term Memory network trained using Multi-Task Learning (LSTM-MTL) and its variants that incorporate informal writing styles, in order to evaluate its performance capability for hashtag extraction from tweets related to different natural and environmental disasters.

### 3 THE UNIOR EYE CORPUS

The C4E - Crowd for the Environment project<sup>2</sup> involves several partners dealing with artificial intelligence and is partly supported by the PON 2014-20 Ricerca e Innovazione fund. The C4E project aims at developing an innovative framework for the detection and monitoring of illegal spills, such as illegal landfills, micro-dumps or illegal releases in surface waters, and the organization of subsequent monitoring actions on site. For this purpose, we compiled the UNIOR Eye corpus [Manna et al., 2020], a corpus composed of tweets related to such environmental crimes posted over the time span 1 January 2013 - 6 August 2020. The corpus is made up of 228,412 tweets, 22,780,746 tokens, and 569,905 types and with a type/token ratio (TTR) of 0.025. The creation of the corpus was preceded by the compilation of a glossary<sup>3</sup> containing 43 terms often used in relation to environmental crimes. The glossary has allowed us to create the corpus as, with the use of Twitter API, it was possible to download all tweets where these words, preceded by hashtags, were present. Therefore, hashtags were essential elements to gather the information needed to detect crimes against the environment. As regards the corpus configuration, it is internally divided into four semantic areas, each one concerning a specific environmental crime: rifiuti e Terra dei fuochi (waste and Terra dei fuochi) composed of 142,174 tweets; reati contro le acque (water-related crimes) composed of 37,473 tweets; materiali e sostanze pericolose (hazardous substances and materials) composed of 13,536 tweets; incendi e roghi ambientali (environmental fires) composed of 35,229 tweets. These sets are further divided into more specific subsets, e.g. the folder incendi e roghi ambientali (environmental fires) contains the subset incendi boschivi (forest fires) consisting of 1,434 tweets and incendi dolosi (arsons) consisting of 23,130 tweets.

### 4 CASE STUDY

This section outlines the methodology adopted to analyze the information conveyed by hashtags. In particular, our aim is to show that:

- the information conveyed by pairs (e.g. #terradeifuochi - #stopbiocidio) and trios (e.g. #rifiuti - #terradeifuochi - #m5s) of hashtags is shared by the corresponding tweets [Hong and Davidson, 2010].
- different textual contents can be discerned through the information affixed by the user via hashtags [7]. More specifically, the main purpose is to separate alert tweets from tweets containing political and personal opinions in a dataset of tweets concerning La Terra dei Fuochi.

We start selecting from the semantic area Terra dei Fuochi of our corpus 77,205 tweets, previously extracted using the hashtag #terradeifuochi and its spelling variants (e.g. #Terradeifuochi, #TerradeiFuochi, #terredeifuochi).

---

2 <https://sites.google.com/view/c4e-crowdfortheenvironment/home-page>

3 [https://drive.google.com/file/d/1H\\_b\\_iH0sEA6V7C9bVfVI3YYXEq1U0rr/view](https://drive.google.com/file/d/1H_b_iH0sEA6V7C9bVfVI3YYXEq1U0rr/view)

Then, we perform the following preprocessing steps: lowercasing the textual content of the tweets and the hashtags, thus avoiding the spelling variations of the same information; removing punctuation and special characters from the tweet; removing a custom set of Italian stopwords (i.e. determiners, conjunctions and prepositions) from tweets. All preprocessing steps are performed using the NLTK package in Python. Once these steps are completed, we focus on analyzing the co-occurrences (n-grams) of hashtags. The calculated average of token-hashtags in our corpus corresponds to 2,425 tokens-hashtags per tweet. Therefore, we extract, visualize and analyze the textual content of bigrams and trigrams, both shown in Figure 1, from the hashtags following each tweet in the dataset. Considering the hashtags bigrams and trigrams shown below, we qualitatively identify three types of textual information related to Terra dei Fuochi: political parties, criminal organizations and alerts. In particular, the textual link between Terra dei Fuochi and politics is reported in bigrams (`#terradeifuochi` and `#m5s`, `#renzi`, `#napolitano` and so on), while texts related to opinions about criminal organizations are more present in trigrams (e.g. `#terradeifuochi #caserta #lupi`). Then, keeping the hashtags hidden, we cluster the tweets using the K-means algorithm in an unsupervised way. We evaluate the number of clusters using the Elbow method, shown in Figure 2 on the left side, and silhouette score with a peak at  $k = 6$  with value of 0.78. Therefore, six appropriate clusters are identified to describe the dataset. These clusters are presented in Figure 2 on the right. By inspecting the words contained in each cluster, we identify two alert clusters containing words such as: `rifiuti`, `tossici`, `vicino`, `località`, etc. Then the cluster labels are assigned to the alert tweets. Therefore, we are able to identify the most relevant hashtags for alert tweets shown in Figure 3.

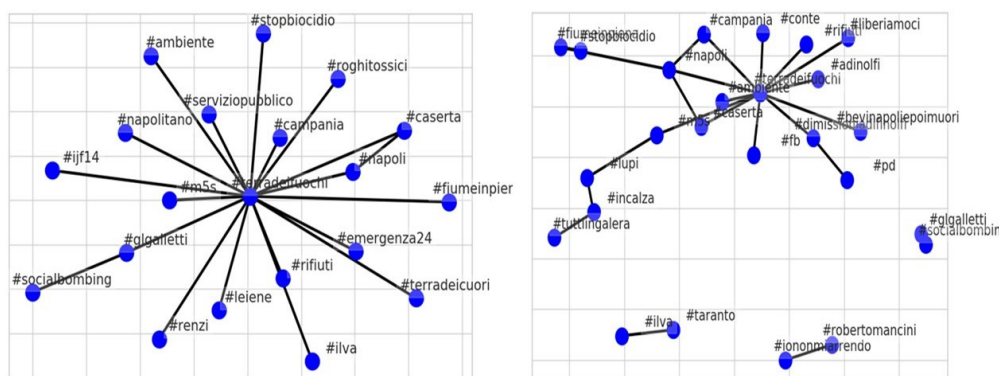


Figure 1: hashtags bigrams and trigrams

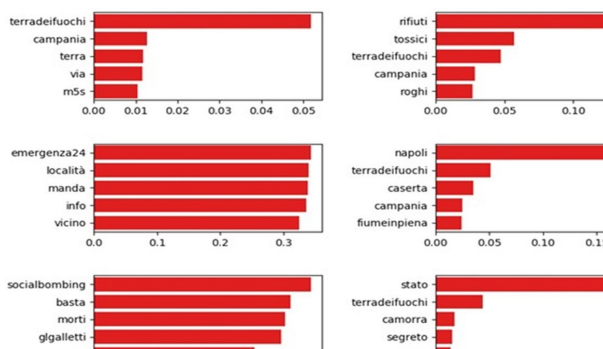
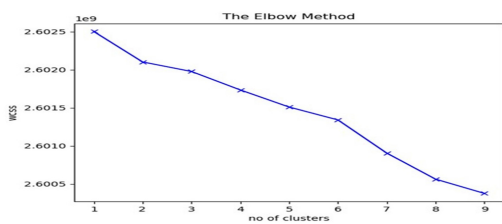


Figure 2: Elbow method on the left and six clusters on the right side

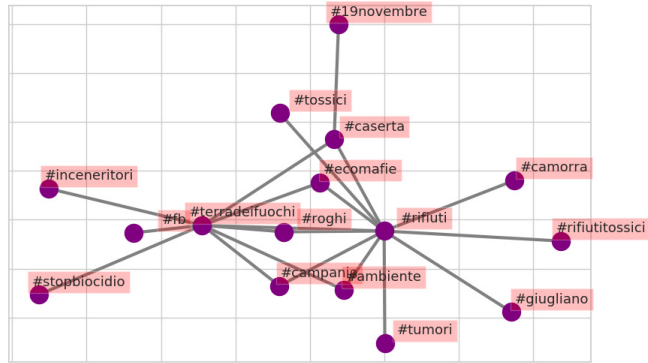


Figure 3: relevant hashtags for the alert tweets

Once the alerts tweets about *La Terra dei Fuochi* are identified (precisely 36,207), we extract the exact place (preceded by a hashtag) where the crime took place and establish the type of crime, with the view to representing this information on a map. Using a link<sup>4</sup> it is possible to visualize the interactive Hashtag frequency map, namely the map of places preceded by a hashtag reported in the alert tweets. It is also possible to verify the instances of the alert tweets preceded by a hashtag for the single place by placing the mouse cursor on them. In Figure 4 we show a screenshot of the map.

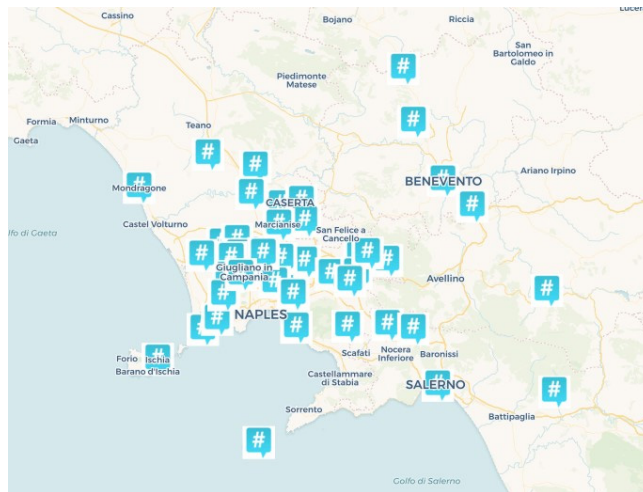


Figure 4: a screenshot of the Hashtag frequency map

4 <https://uniornlp.carto.com/builder/4e072331-54a8-43bd-aa59-55302cafe923/embed>

## **5 CONCLUSIONS AND FUTURE WORK**

We presented a case study concerning the UNIOR Eye corpus, a corpus made up of 228,412 tweets related to environmental crimes, aimed at investigating the frequency of co-occurrence of hashtags to monitor environmental crimes, particularly related to La Terra dei Fuochi. The research is carried out in the framework of the C4E - Crowd for the Environment project, partly supported by the PON Ricerca e Innovazione fund, which aims at developing an innovative framework for the detection and monitoring of illegal spills. Our method is based on the extraction of hashtags contained in the tweets gathered in the dataset, and then on clustering these tweets in order to identify the most relevant hashtags for each labeled alert tweet. Future work will focus on extending our hashtag-based analysis to cluster and map the entire corpus in order to provide computational methods to monitor the environment through social media.

## **6 ACKNOWLEDGEMENTS**

This research has been conducted within the framework of two Innovative Industrial PhD projects supported by the PON Ricerca e Innovazione 2014/20 and the POR Campania FSE 2014/2020 funds and two research grants supported by the PON Ricerca e Innovazione 2014/20 in the context of the C4E project. Special thanks to Annarita Magliacane for helping us during this research. We are grateful to Prof. Johanna Monti for supervising the research.

## REFERENCES

1. Ampofo, Lawrence. 'The Social Life of Real-Time Social Media Monitoring'. *Participations* 8, no. 1 (2011): 21–47.
2. Austin, Lucinda L., and Yan Jin. *Social Media and Crisis Communication*. Taylor & Francis, 2017.
3. Chowdhury, Jishnu Ray, Cornelia Caragea, and Doina Caragea. 'On Identifying Hashtags in Disaster Twitter Data'. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 34:498–506, 2020.
4. Cobo, Alfredo, Denis Parra, and Jaime Navón. 'Identifying Relevant Messages in a Twitter-Based Citizen Channel for Natural Disaster Situations'. In *Proceedings of the 24th International Conference on World Wide Web*, 1189–1194, 2015.
5. Crowe, Adam. *Disasters 2.0: The Application of Social Media Systems for Modern Emergency Management*. CRC press, 2012.
6. Hong, Liangjie, and Brian D. Davison. 'Empirical Study of Topic Modeling in Twitter'. In *Proceedings of the First Workshop on Social Media Analytics*, 80–88, 2010.
7. Lim, Kar Wai, and Wray Buntine. 'Twitter Opinion Topic Model: Extracting Product Opinions from Tweets by Leveraging Hashtags and Sentiment Lexicon'. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, 1319–1328, 2014.
8. Manna, Raffaele, Antonio Pascucci, Wanda Punzi Zarino, Vincenzo Simoniello, and Johanna Monti. 'Monitoring Social Media to Identify Environmental Crimes through NLP A Preliminary Study', 2020.
9. Murzintcev, Nikita, and Changxiu Cheng. 'Disaster Hashtags in Social Media'. *ISPRS International Journal of Geo-Information* 6, no. 7 (2017): 204.
10. Peluso, Pasquale. 'Dalla Terra Dei Fuochi Alle Terre Avvelenate: Lo Smaltimento Illecito Dei Rifiuti in Italia'. *Rivista Di Criminologia, Vittimologia e Sicurezza* 9, no. 2 (2015): 13–30.
11. Wang, Rong, Wenlin Liu, and Shuyang Gao. 'Hashtags and Information Virality in Networked Social Movement'. *Online Information Review*, 2016.
12. Yang, Shih-Feng, and Julia Taylor Rayz. 'An Event Detection Approach Based on Twitter Hashtags'. *ArXiv Preprint ArXiv:1804.11243*, 2018.



# Languages and Cultures of Ancient Italy. Historical Linguistics and Digital Models

A. Marinetti<sup>1</sup>, F. Murano<sup>2</sup>, V. Quochi<sup>3</sup>, M. Ballerini<sup>2</sup>, F. Boschetti<sup>3</sup>, A.M. Del Grosso<sup>3</sup>,  
S. Piccini<sup>3</sup>, L. Rigobianco<sup>1</sup>, P. Solinas<sup>1</sup>

<sup>1</sup> Università “Ca’Foscari” di Venezia - <linda;solinas;luca.rigobianco> [at] unive.it

<sup>2</sup> Università di Firenze - name.surname [at] unifi.it

<sup>3</sup> ILC-CNR - name.surname [at] ilc.cnr.it

## ABSTRACT

The poster presents a newly started project about languages and cultures of Ancient Italy, which brings together competences from Historical Linguistics, Computational Lexicography and Digital Humanities. The main objective of the project is to investigate the cultures of ancient Italy on the basis of their linguistic documentation (7th - 1st c. B.C.) by means of digital tools specifically tailored for their peculiarities.

## KEYWORDS

restsprachen, ancient Italy, epigraphic text corpora, computational lexica, semantic web

## POSTER

### 1 INTRODUCTION AND BACKGROUND

During the period dating from the appearance of writing (8th c. B.C.) until complete linguistic Romanisation (1st c. B.C. to 1st c. A.D.), ancient Italy is characterised by the presence of numerous linguistic varieties, which are both Indo-European and non-Indo-European. These varieties are documented by *corpora* that are not homogenous in their quantity, quality, geographical distribution, and chronology. Apart from this lack of homogeneity, with the exception of the Latin of Rome they can be qualified as *Restsprachen*, i.e. fragmentarily attested languages, inasmuch as their linguistic documentation consists almost exclusively of epigraphic texts, mostly short and repetitive. For this reason, *Restsprachen* often present specific problems concerning the reading of the inscriptions, the segmentation into words, their linguistic analysis, and their interpretation. The linguistic framework as *Restsprachen* of the languages involved in this project – Venetic, Oscan, Faliscan, and Celtic – thus requires the adoption of specific methods and approaches for the analysis and edition of their texts[10; 17; 20].

To achieve this challenging goal, we will combine the traditional method proper to Historical Linguistics and its products (study, edition, and the linguistic, historical, and cultural commentary on the texts) with the setting up of digital technologies that will facilitate scholars in their activities.

## 2 OBJECTIVES

The main goal of the project is to investigate the cultures of ancient Italy on the basis of the relevant linguistic documentation in order to show the forms of linguistic variability in Italy before Romanisation. The study and valorisation of this cultural heritage is to be achieved through interdisciplinary work that brings together methods and practices from the “traditional” study of such materials, computational lexicology and lexicography, semantic web and various digital humanities technology, such as digital archiving, textual corpora, contextualisation and visualisation methods. In particular, we aim at digitising key materials and deploying electronic tools and resources that will permit their easy consultation and revision, which in turn will both facilitate scholars’ work and allow greater disclosure of the acquired scientific knowledge. Given the experimental nature of the project, we will focus on the Venetic, Oscan, Faliscan, and Celtic *corpora*. Although ambitious and challenging, given the main goal of the project, i.e. to develop a new methodology that leverages specifically tailored digital technologies to facilitate scholars’ work and collaboration while at the same time ensuring long term preservation of the acquired knowledge, the risks are calculated. To minimise them we build upon our own direct experiences in developing tools for the digital humanities and in modeling language data, as well as upon other similar national and European endeavours. At the end of the project, it is expected that both the methodology as well as the digital toolkit and technology set up will be ready to be extended to the other *Restsprachen*.

## 3 METHODOLOGIES AND EXPECTED OUTCOMES

The project plans to create a digital set of interrelated resources: a *corpus* of epigraphic texts, a computational lexicon for the languages involved, a dataset of bibliographic references, and an experimental semantic dataset of interpretations. The *corpus* of the texts will be managed and exploited in a digital archive containing the formal representation of the texts leveraging the TEI/EpiDoc encoding schema [9]; it may be necessary to create an *ad-hoc* schema for the peculiarities presented by languages of fragmentary attestation, the EpiDoc model having been used so far only for the creation of annotated ‘major’ corpora, especially of Greek and Latin inscriptions. For example, *Restsprachen* may require the identification of standards suitable for the encoding of unusual and odd writing ductus, punctuation marks with particular values (e.g. syllabic punctuation), characters whose linguistic nature is uncertain, multiple hypotheses of word segmentation, etc. The application of the EpiDoc model to *Restsprachen* is, thus, a complete novelty in the field. Each text in the archive will be enriched with shared and standard metadata allowing for an accurate description, both as a linguistic object (text: language, alphabet, date, etc.) and as a material object (support: chronology, data of discovery, material, etc.). Accordingly, such a description will facilitate the retrieval of relevant information.

The project additionally plans to experiment with the use of CRMtex[11; 12] and CRMinf [4] extensions of CIDOC CRM, the *de facto* standard ontology in the Digital Humanities for the representation of the texts and their scientific interpretation in a semantic format [5]. As regards the lexicon, the project will investigate the specific requirements for the design of an efficient computational lexical model specifically dedicated to languages of fragmentary attestation in order to produce a multilingual (Venetic, Oscan, Faliscan, and Celtic) computational lexicon. We will adopt Semantic Web standards and vocabularies for providing a structured and formal representation of the lexical items and their related information as well as for allowing for a sophisticated semantic access to the *corpus* of epigraphies [1] [16]. The challenges to be faced in lexical modeling are numerous, since we are dealing with Restsprachen, and they range from lemmatisation issues, as in many cases the relationship between words and lemmas is not certain for various reasons (different graphic standards, difficult linguistic analysis, incomplete paradigms, etc.), to sense representation, given that meanings are often only partially and hypothetically reconstructible, and smart reference or linking to attestations. Fortunately, recent efforts within modeling and standardisation initiatives are tackling many of these critical issues, and they will provide us with inspiration and guidance. We refer in particular to the works on the extension of existing representation models such as TEI and Ontolex-lemon for the account of attestations[3] and etymologies[2; 14; 13]. Models and methods for interlinking the datasets will also be designed as the backbone for a digital toolkit that will allow for both the creation/revision of some of the materials and their online fruition, see [15; 8; 19; 18]. Specifically, the heart of the toolkit will be a lexicon editing and consultation service specifically tailored to Restsprachen, based on inhouse previous experiences [1]; this will be enriched for allowing the interlinking among the different datasets, thus acting as a sort of hub that will primarily integrate lexicon and epigraphy transcriptions, together with contextual metadata, bibliography, and experimentally the hermeneutic positions. It is also planned to experiment with Domain-Specific Languages to deploy a system that can assist scholars in the creation of the textual digital resources and ensure compatibility with the aforementioned standards [19]. The tools and resources produced and developed within the project will finally be made available through relevant European-wide Research Infrastructures such as CLARIN [6] and DARIAH [7], the two currently most relevant infrastructures for the e-Humanities and (immaterial) Cultural Heritage. This will ensure both a long-term preservation of the resources produced and a high valorisation of this heritage.

## 4 TITLE

The project is supported by the Italian Ministry of Research [PRIN 2017XJLE8J]

## REFERENCES

1. Bellandi, Andrea, Emiliano Giovannetti, Silvia Piccini, and Anja Weingart. 'Developing LexO: A Collaborative Editor of Multilingual Lexica and Terminology-Ontological Resources in the Humanities'. In *Proceedings of Language, Ontology, Terminology and Knowledge Structures Workshop (LOTKS 2017)*, 2017. <https://www.aclweb.org/anthology/W17-7010.pdf>.
2. Bowers, Jack, and Laurent Romary. 'Deep Encoding of Etymological Information in TEI'. *Journal of the Text Encoding Initiative*, no. 10 (2016). <https://doi.org/https://doi.org/10.4000/jtei.1643>.
3. Chiarcos, Christian, Maxim Ionov, Jesse de Does, Katrien Depuydt, Fahad Khan, Sander Stolk, Thierry Declerck, and John Philip McCrae. 'Modelling Frequency and Attestations for OntoLex-Lemon'. In *Proceedings of the 2020 Globalex Workshop on Linked Lexicography*, 1–9, 2020.
4. CIDOC. 'Home | CIDOC CRM', 2020. <http://www.cidoc-crm.org/>.
5. ———. 'Home | CRMInf', 2020. <http://www.cidoc-crm.org/crminf/>.
6. CLARIN-it. 'CLARIN-IT | the Italian Common Language Resources and Technology Infrastructure', 2020. <http://www.clarin-it.it/>.
7. DARIAH. 'Dariah - Progetto Science and Technology Digital Library', 2020. <http://stdl.cnr.it/it/dariah>.
8. Del Grosso, Angelo Mario, Andrea Bellandi, Emiliano Giovannetti, Simone Marchi, and Ouafae Nahli. 'Scanning Is Just the Beginning: Exploiting Text and Language Technologies to Enhance the Value of Historical Manuscripts'. In *Proceedings of the IEEE 5th International Congress on Information Science and Technology (CiSt)*, 214–219. Marrakech, Morocco: IEEE, 2018.
9. EpiDoc. 'EpiDoc: Epigraphic Documents in TEI XML', 2020. <https://epidoc.stoa.org/gl/latest/toc-it.htm>.
10. Farney, Gary D., and Guy Bradley. *The Peoples of Ancient Italy*. Walter de Gruyter GmbH & Co KG, 2017.
11. Felicetti, Achille, and Francesca Murano. 'Scripta Manent: A CIDOC CRM Semiotic Reading of Ancient Texts'. *International Journal on Digital Libraries* 18, no. 4 (2017): 263–270.
12. Felicetti, Achille, Francesca Murano, Paola Ronzino, and Franco Niccolucci. 'CIDOC CRM and Epigraphy: A Hermeneutic Challenge.' In *Extending, Mapping and Focusing the CIDOC CRM. CRMEX 2015 Workshop, 19th International Conference on Theory and Practice of Digital Libraries*, 55–68, 2015. <http://ceur-ws.org/Vol-1656/paper5.pdf>.
13. Khan, Anas Fahad. 'Towards the Representation of Etymological Data on the Semantic Web'. In *Information*, 9:304, 2018. <https://hal.inria.fr/hal-02618067/document>.
14. Khan, Anas Fahad, and Federico Boschetti. 'Towards a Representation of Citations in Linked Data Lexical Resources'. In *Proceedings of the XVIII EURALEX International Congress: Lexicography in Global Contexts*, 137–147, 2018.

15. Khan, Fahad, Andrea Bellandi, and Monica Monachini. 'Tools and Instruments for Building and Querying Diachronic Computational Lexica'. In *Proceedings of the Workshop on Language Technology Resources and Tools for Digital Humanities (LT4DH)*, 164–171, 2016.
16. Khan, Fahad, Laurent Romary, Ana Salgado, Jack Bowers, Mohamed Khemakhem, and Toma Tasovac. 'Modelling Etymology in LMF/TEI: The Grande Dicionário Houaiss Da Língua Portuguesa Dictionary as a Use Case'. In *Proceedings of The 12th Language Resources and Evaluation Conference*, 3172–3180. Marseille, France, 2020. <https://hal.inria.fr/hal-02618067/document>.
17. Klein, Jared, Joseph Brian, Matthias Fritz, and Mark Wenthe. *Handbook of Comparative and Historical Indo-European Linguistics*. Berlin; Boston: Mouton De Gruyter, 2018.
18. Mambrini, Francesco, Flavio Massimo Cecchini, Greta Franzini, Eleonora Litta, Marco Carlo Passarotti, and Paolo Ruffolo. 'LiLa: Linking Latin. Risorse Linguistiche per Il Latino Nel Semantic Web'. *Umanistica Digitale* 4, no. 8 (2020). <https://doi.org/10.6092/issn.2532-8816/9975>.
19. Mugelli, Gloria, Federico Boschetti, Riccardo Del Gratta, Angelo Mario Del Grosso, Fahad Khan, and Andrea Taddei. 'A USER-CENTRED DESIGN TO ANNOTATE RITUAL FACTS IN ANCIENT GREEK TRAGEDIES'. *Bulletin of the Institute of Classical Studies* 59, no. 2 (2016): 103–120.
20. Passarotti, Marco, Francesco Mambrini, Greta Franzini, Flavio Massimiliano Cecchini, Eleonora Litta, Giovanni Moretti, Paolo Ruffolo, and Rachele Sprugnoli. 'Interlinking through Lemmas. The Lexical Collection of the LiLa Knowledge Base of Linguistic Resources for Latin'. *Studi e Saggi Linguistici* 58, no. 1 (2020): 177–212. <https://doi.org/10.4454/ssl.v58i1.277>.
21. Prodocimi, Aldo Luigi. *Lingue e Dialetti Dell'Italia Antica*. Vol. 6. Popoli e Civiltà dell'Italia Antica. Roma;Padova: Biblioteca di Storia Patria, 1978.

# Digital preservation “FAIRness” and “TRUSTworthiness”: i principi FAIR e TRUST nei contesti di conservazione digitale

Federica Marti  
Università di Macerata – f.marti [at] unimc.it

## SINTESI

La presente relazione introduce i *FAIR Guiding Principles for Scientific Data Management and Stewardship* e i *TRUST Principles for Digital Repositories* e ha come scopo il vaglio di alcune esperienze di applicazione di tali principi in ambito internazionale, al fine di fornire una valutazione riguardo l'efficacia del loro impiego nel contesto della conservazione digitale e la coerenza rispetto all'adozione degli standard di conservazione.

## PAROLE CHIAVE

FAIR, TRUST, conservazione digitale, accessibilità, repositories, contesto internazionale.

## INTERVENTO

### 1 INTRODUZIONE: “BUILD TRUST TO BE FAIR”<sup>1</sup>

*Findable, Accessible, Interoperable, Reusable; Transparency, Responsibility, User Focus, Sustainability, Tecnology*: queste le parole chiave in contesto internazionale in relazione alla creazione, diffusione e valorizzazione dei dati di ricerca e per l'implementazione di *repositories* in grado di conservare a lungo termine questi dati. Concepite come linee guida generali e applicabili a contesti diversificati<sup>2</sup>, i *FAIR Guiding principles for scientific data management and stewardship* e i *TRUST Principles for digital repositories* auspicano la condivisione da parte degli enti di ricerca a livello internazionale, da un lato al fine di agevolare e incrementare la condivisione e la consultabilità di dati e informazioni, dall'altro di

---

1 Si cita, per il titolo di questo paragrafo, la tredicesima Sessione plenaria della Research Data Alliance (RDA) “RDA/WDS Certification of Digital Repositories: Build TRUST to be FAIR - Emerging Needs of Certification in Life Sciences, Geosciences and Humanities”: nel corso di questo evento, tenutosi a Philadelphia il 4 aprile 2019, i TRUST Principles vengono formalmente proposti e riconosciuti come community discussion.

2 Ci si riferisce, ad esempio, a istituti di ricerca in ambito fisico-matematico, astronomico, biologico e sanitario; o, ancora, a enti preposti alla curatela del patrimonio culturale come biblioteche e archivi.

conservarli a lungo termine. Nello specifico, i *FAIR Guiding principles* forniscono indicazioni sulle modalità di predisposizione di dati, metadati e infrastrutture, perché siano, accessibili, interoperabili, ricercabili e riusabili<sup>3</sup>.

| Box 2   The FAIR Guiding Principles  |  |
|--|--|
| <b>To be Findable:</b>   |  |
| F1. (meta)data are assigned a globally unique and persistent identifier  |  |
| F2. data are described with rich metadata (defined by R1 below)  |  |
| F3. metadata clearly and explicitly include the identifier of the data it describes                            |  |
| F4. (meta)data are registered or indexed in a searchable resource  |  |
| <b>To be Accessible:</b>   |  |
| A1. (meta)data are retrievable by their identifier using a standardized communications protocol                |  |
| A1.1 the protocol is open, free, and universally implementable   |  |
| A1.2 the protocol allows for an authentication and authorization procedure, where necessary                    |  |
| A2. metadata are accessible, even when the data are no longer available  |  |
| <b>To be Interoperable:</b>  |  |
| I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation. |  |
| I2. (meta)data use vocabularies that follow FAIR principles  |  |
| I3. (meta)data include qualified references to other (meta)data  |  |
| <b>To be Reusable:</b>   |  |
| R1. meta(data) are richly described with a plurality of accurate and relevant attributes                       |  |
| R1.1. (meta)data are released with a clear and accessible data usage license                                   |  |
| R1.2. (meta)data are associated with detailed provenance   |  |
| R1.3. (meta)data meet domain-relevant community standards  |  |

Fig. 1. I FAIR Guiding principles (Wilkinson et al., 2016)

I recentissimi *TRUST Principles*, invece, si propongono come linee guida per la costruzione di un *repository*, che consenta il persistere delle proprietà di affidabilità, accessibilità e utilizzabilità dei dati in essi riposti<sup>4</sup>.

| Box 1 The TRUST Principles |  |
|----------------------------|--|
| Principle                  | Guidance for repositories  |
| Transparency               | To be transparent about specific repository services and data holdings that are verifiable by publicly accessible evidence.            |
| Responsibility             | To be responsible for ensuring the authenticity and integrity of data holdings and for the reliability and persistence of its service. |
| User Focus                 | To ensure that the data management norms and expectations of target user communities are met.  |
| Sustainability             | To sustain services and preserve data holdings for the long-term.  |
| Technology                 | To provide infrastructure and capabilities to support secure, persistent, and reliable services.                                       |

Fig. 2. I TRUST Principles (Dawei Lin et al., 2020)

- 3 I *FAIR Guiding principles* sono stati formulati nell'ambito di un gruppo di lavoro comprendente ricercatori, editori accademici, professionisti, esperti commerciali ed enti di finanziamento e pubblicati nel 2016 sulla rivista *Scientific Data* [7]. Si appoggiano a diverse iniziative, che mirano sia al tracciamento di metriche di dettaglio per la compliance a FAIR (si vedano, a proposito, FAIRsFAIR al link <https://www.fairsfair.eu/> e [8]) sia all'utilizzo su larga scala di questi principi (si veda, ad esempio, GoFair (<https://www.go-fair.org/> - ultima consultazione: 12/09/2020).
- 4 Rilasciati il 14 maggio 2020, anch'essi su *Scientific Data* e risultanti da mesi di consultazione e discussione della comunità di ricerca nell'ambito della Research Data Alliance (RDA), i principi TRUST sono stati sviluppati al fine di facilitare l'adozione di best practices per la gestione dei repositories. Il primo giugno 2020 l'Open Preservation Foundation (OPF), organismo tra i più attivi in contesto europeo per l'elaborazione di strumenti e linee guida per la conservazione digitale, ha accolto questo set di principi, uno dei cui punti di forza risiede nell'aver acquisito l'endorsement di diverse istituzioni su scala mondiale (il cui elenco è consultabile al link <https://www.rd-alliance.org/trust-principles-rda-community-effort> - ultima consultazione 12/09/2020) e, dunque, nel rappresentare un documento ampiamente condiviso [4].

Il presupposto della riflessione che segue è la considerazione dei principi TRUST come una “conseguente integrazione” di FAIR<sup>5</sup>; questi ultimi, infatti, non essendo elaborati esplicitamente ai fini della conservazione digitale, possono costituire sia il presupposto, sia la conseguenza dei TRUST: Lin Dawei (National Institute of Health – NIH), nel corso dell’evento “TRUST Principles Mini Symposium: The Future of Digital Repositories”<sup>6</sup>, ha illustrato come i principi di TRUST siano stati progettati anche allo scopo di influire positivamente sulla *FAIRness*<sup>7</sup> dei dati all’interno dei *repository*. Infatti, gestire un *repository* in maniera conforme ai principi TRUST non solo permette di implementare la *FAIRness* nei dati per cui originariamente manca, ma consente anche di garantire il perdurare di questa caratteristica (ad esempio, nel caso in cui venga meno a causa di un cambiamento negli standard di metadattazione).



Fig. 3. Rappresentazione grafica della relazione tra i principi TRUST e FAIR (Dawei Lin, 2020)

## 2 ADAPTABILITY: I PRINCIPI FAIR E TRUST NEI CONTESTI DI CONSERVAZIONE DIGITALE

L’adattabilità di questi principi rispetto ai modelli e agli standard della conservazione digitale e, dunque, la loro efficacia in ambito applicativo è rilevabile attraverso dei contesti d’uso.

In primo luogo, un esempio di conferma della compatibilità dei principi FAIR con lo standard OAIS, riferimento internazionale per l’implementazione di sistemi di conservazione, si trova nel progetto ARCHIVER, iniziativa europea finalizzata all’implementazione di un servizio archiviazione e conservazione per i dati di ricerca, i cui *stakeholders* sono molti tra i maggiori istituti di ricerca internazionali<sup>8</sup>: si richiede, a bando, la presenza integrata di entrambi gli elementi tra i requisiti funzionali, al fine di po-

5 A questo proposito, si sottolinea che, chiaramente, alcune raccomandazioni presenti in questi due documenti si sovrappongono: si veda il caso dei metadati, per descrivere i quali è raccomandato l’utilizzo di standard impiegati dalla vasta *community* di riferimento.

6 Nel corso di questo “minisimposio”, promosso da Research Data Canada, Research Data Alliance, and World Data Systems e svoltosi il 7 luglio 2020, sono stati discussi, secondo diverse prospettive e casi d’uso, i *TRUST Principles*. A proposito, si veda il resoconto di Robyn Nicholson (Research Data Alliance) al link <https://www.rdc-drc.ca/trust-principles-mini-symposium-the-future-of-digital-repositories/>.

7 Per *FAIRness* si vuole intendere, avvalendosi del gioco di parole tra il sostantivo “correttezza” e l’acronimo dei principi, la conformità dei dati ai principi FAIR.

8 Il *lead partner* del progetto è il CERN e il consorzio di *buyers* è formato dall’European Molecular Biology Laboratory (EMBL), dal Deutsches Elektronen-Synchrotron (DESY) e dal Port d’Informació Científica (PIC); tra gli *early adopters* si trovano National Research and Education Network in the United Kingdom, Friedrich Miescher Institute for Biomedical Research (Svizzera), Stockholm University Library (Svezia), European Institute of Oncology (Italia), National Research and Education Network in the Netherlands, Australia’s Academic and Research Network (Australia). <https://www.archiver-project.eu/>.



ter disporre di strumenti software di base interoperabili entro cui si possano rintracciare, monitorare e riutilizzare dati predisposti dalla creazione come FAIR.

In secondo luogo, riguardo alla posizione dei principi FAIR rispetto alla certificazione CoreTrustSeal, la cui metrica di valutazione per certificare i *repository* si basa su ISO 16363 ed è ampiamente utilizzata in ambito internazionale, gli studiosi Mustapha Mokrane (Data Archiving and Networked Services - DANS) e Jonas Recker (GESIS Data Archive for the Social Sciences) hanno valutato che i requisiti CoreTrustSeal incorporano e riferiscono, di fatto, i focus dei principi FAIR, ma li completano visualizzando, appunto, gli aspetti connessi alla conservazione dei dati. Per esemplificare, se si considera il requisito FAIR “*F1. (meta)data are assigned a globally unique and persistent identifier*”, si può notare che la concreta persistenza dell’identificativo consegua alla conservazione di questo dato all’interno di un *trusted digital repository* [6].

Per quanto riguarda i principi TRUST, questi sono stati elaborati concependo come propria implementazione gli standard di conservazione e le certificazioni CoreTrustSeal e ISO 16363: David Giaretta (Primary Trustworthy Digital Repository Authorisation Body - PTAB) ha esaminato l’intersezione dei criteri di *auditing* ISO 16363 - sottolineando come questi rappresentino la componente “testabile” di OAIS - e ISO 16919 con i principi TRUST, definendo questi ultimi come un punto di partenza e affermando che sia, comunque, imprescindibile la verifica dell’effettiva *trustworthiness* dei *repositories* nei dettagli, attraverso i processi di audit<sup>9</sup>.

Vi è, infine, la complementarità tra gli stessi principi FAIR e TRUST: Ingrid Dillo (DANS) ha evidenziato l’importanza dell’ecosistema in cui i dati sono riposti (secondo la stessa logica esposta per CoreTrustSeal) e asserito come l’*endorsement* TRUST di un *repository* costituisca la base per l’affermazione della propria affidabilità e, di conseguenza, per la predisposizione FAIR dei dati.

### 3 WHAT ABOUT THE FITNESS?

Emerge, dai casi presentati, che la duttilità che c’è alla base della formulazione dei principi FAIR e TRUST è particolarmente funzionale, in quanto questi consentono una scalabilità in contesti diversi che cerca di salvaguardare le diverse esigenze proprie della conservazione digitale. Pertanto, la genericità *by design* con cui sono stati concepiti non costituisce un limite, ma, al contrario, un vantaggio: il fatto che non vengano indicati standard e requisiti tecnologici specifici (si veda, per FAIR, “*I.1.: (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation*”) consente ai professionisti della conservazione digitale di adottare i più adeguati al proprio contesto di riferimento, mentre le definizioni piuttosto indeterminate (si veda, per TRUST, “*Transparency. To be transparent about specific repository services and data holdings that are verifiable by publicly accessible evidence*”) permettono ai *digital preservation practitioners* di accostarle a quelle del proprio ambito nazionale e, in particolare, del proprio quadro legislativo. Non si dimentichi che FAIR e TRUST sono stati elaborati in un ambito estremamente multidisciplinare, che comprende istituti di ricerca che vanno dall’ambito sanitario all’ambito archivistico, dall’ambito matematico a quello bibliotecono-

---

<sup>9</sup> Per questo caso di studio e per il seguente si fa riferimento al già citato “TRUST Principles Mini Symposium: The Future of Digital Repositories” (v. nota 6). In particolare, si fa riferimento agli interventi “Implementation Pathway: TRUST and ISO 16363/ISO 16919 Intersections” di David Giaretta e “TRUST and FAIR: Complementarity of the Principles” di Ingrid Dillo.

mico, collocati in tutto il mondo: si tratta di *principi* scaturiti da un impegno collettivo, che riflettono necessità che possono essere molto distanti<sup>10</sup>, il cui intento è giungere all’obiettivo comune di gestire *repositories* funzionali e durevoli nel tempo, senza contrastare con il contesto tecnologico e normativo specifico in cui si agisce<sup>11</sup>.

Dunque, non si vede la necessità di enunciazioni teoriche o regole precise: queste linee guida non sono formulate a scopo introduttivo rispetto alla gestione dei dati e alla costruzione dei *repositories*, né si propongono come modelli circoscritti e “certificanti”, ma sono destinati a un pubblico di professionisti ed esperti che vi possano trovare un *framework* di riferimento per l’elaborazione delle proprie strategie di gestione e conservazione di dati e documenti e l’implementazione dei propri sistemi. A tal proposito, infatti, a completamento di FAIR sono state elaborate metriche e griglie di valutazione che consentano di misurare l’effettiva *FAIRness* delle soluzioni utilizzate<sup>12</sup>, mentre, riguardo TRUST, il processo può dirsi inverso: più che introdurre concetti innovativi, vogliono sistematizzare quelli già esistenti e, in parte, già delimitati da standard, sulla base dei quali si innesta la formulazione dei principi stessi. Su questa scia, l’elaborazione dei principi TRUST può essere considerata come espediente per sopperire alla mancanza, nei principi FAIR, di una prospettiva sul mantenimento a lungo termine della reperibilità, dell’accessibilità e della riusabilità dei dati, ricercata in precedenza con l’integrazione di questi principi con standard e certificazioni di conservazione<sup>13</sup>.

In conclusione, è da sottolineare che la discussione riguardo alle potenzialità e alle criticità connesse all’utilizzo di queste *guidelines*, data anche la loro recente formulazione, è attuale e aperta nell’ambito internazionale della conservazione digitale e, di conseguenza, sensibile a sempre nuovi input e soggetta a prospettive diverse.

---

10 Si pensi, ad esempio, all’utilizzo degli open data per la diffusione degli articoli scientifici di matrice divulgativa da un lato, e alla riservatezza dei dati gestiti in ambito sanitario dall’altro.

11 Si consideri il già citato caso della certificazione dei repositories: si può scegliere di vincolarla a ISO 16363, a Core-TrustSeal o a griglie valutative customizzate su altri standard, come quelli rilasciati dall’European Telecommunications Standards Institute (ETSI).

12 Si veda, a proposito, la nota 3.

13 Questa “lacuna” è evidenziata nel rapporto della Commissione Europea «Turning FAIR Data into Reality: Final Report and Action Plan from the European Commission Expert Group on FAIR Data» [3].

## BIBLIOGRAFIA

1. Bartoli, Margherita, Fabiana Guernaccini, e Giovanni Michetti. «Preservation of Linked Open Data». *JLIS. it* 11, n. 2 (2020): 20–44. <https://doi.org/10.4403/jlis.it-12633>.
2. Biro, Timea, Natalie Harrower, Beat Immenhauser, Gehrard Lauer, Maciej Maryl, Tito Orlandi, Bernard Rentier, e Eveline Wandl-Vogt. «Sustainable and FAIR Data Sharing in the Humanities». ALLEA-All European Academies, 2020.
3. European Commission. «Turning FAIR into Reality : Final Report and Action Plan from the European Commission Expert Group on FAIR Data.» Website, 26 novembre 2018. <http://op.europa.eu/en/publication-detail/-/publication/7769a148-f1f6-11e8-9982-01aa75ed71a1/language-en>.
4. Lin, Dawei, Jonathan Crabtree, Ingrid Dillo, Robert R. Downs, Rorie Edmunds, David Giaretta, Marisa De Giusti, Hervé L'Hours, Wim Hugo, e Reyna Jenkyns. «The TRUST Principles for digital repositories». *Scientific Data* 7, n. 1 (2020): 1–5. <https://doi.org/https://doi.org/10.1038/s41597-020-0486-7>.
5. McMurry, Julie A., Nick Juty, Niklas Blomberg, Tony Burdett, Tom Conlin, Nathalie Conte, Mélanie Courtot, John Deck, Michel Dumontier, e Donal K. Fellows. «Identifiers for the 21st century: How to design, provision, and reuse persistent identifiers to maximize utility and impact of life science data». *PLoS biology* 15, n. 6 (2017): e2001414. <https://doi.org/https://doi.org/10.1101/117812>.
6. Mokrane, Mustapha, e Jonas Recker. «CORETRUSTSEAL-CERTIFIED REPOSITORIES. Enabling Findable, Accessible, Interoperable, and Reusable (FAIR) Data», 2019. <https://doi.org/https://doi.org/10.1038/sdata.2016.18>.
7. Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, e Philip E. Bourne. «The FAIR Guiding Principles for scientific data management and stewardship». *Scientific data* 3, n. 1 (2016): 1–9.
8. Wilkinson, Mark D., Susanna-Assunta Sansone, Erik Schultes, Peter Doorn, Luiz Olavo Bonino da Silva Santos, e Michel Dumontier. «A design framework and exemplar metrics for FAIRness». *Scientific data* 5 (2018). <https://doi.org/10.1038/sdata.2018.118>.

# Proposta di modello per realizzare edizioni scientifiche digitali

Chiara Martignano  
Università di Firenze – [chiaramartignano \[at\] gmail.com](mailto:chiaramartignano@gmail.com)

## SINTESI

Attualmente nel campo dello *scholarly digital editing* mancano modelli forti di riferimento per sviluppare e progettare edizioni scientifiche digitali (*Digital Scholarly Editions*, DSE). In questo poster si illustra un progetto di modello concettuale per la realizzazione di DSE, che faciliti lo sviluppo di DSE e di *tool* e piattaforme per la pubblicazione di edizioni digitali. Il modello sarà utilizzabile anche come riferimento per la valutazione della qualità scientifica delle DSE, contribuendo al consolidamento delle pratiche adottate negli ultimi anni nella disciplina dello *scholarly digital editing*. Il modello, che comprenderà la formalizzazione degli elementi che costituiscono un'edizione e a indicazioni pratiche per la progettazione dell'interfaccia di una DSE, sarà sviluppato a partire da una ricognizione dello stato dell'arte e, successivamente, testato sviluppando un'applicazione per la pubblicazione di edizioni digitali.

## PAROLE CHIAVE

edizione scientifica digitale, scholarly digital editing, filologia digitale, modello, interfaccia, standard

## INTERVENTO

### 1 INTRODUZIONE

Nel campo della filologia digitale o, più in generale, dello *scholarly digital editing*, negli ultimi anni si sono consolidate delle pratiche che possono essere definite “*soft models*” [7], come, per esempio, l'utilizzo di XML-TEI come tecnologia di base. Al tempo stesso diversi filologi digitali hanno elaborato linee guida e raccomandazioni per la realizzazione di DSE (cfr. [9] e [5], pp. 89-97). Tuttavia la comunità scientifica ancora non ha proposto né sviluppato esplicitamente dei modelli “forti”, degli standard di riferimento per la realizzazione di edizioni digitali.

Un modello forte, che, a seconda della tipologia di edizione digitale che si desidera realizzare, indichi come l'edizione debba funzionare ed essere consultabile dal lettore, faciliterebbe la produzione di DSE interoperabili e di qualità. Grazie a uno standard condiviso sarebbe possibile sviluppare *tool* e piattaforme generaliste, che consentano a un pubblico più ampio di studiosi di pubblicare edizioni digitali a basso costo e senza dover disporre di particolari competenze tecniche o di un supporto informatico.

Il modello, inoltre, rappresenterebbe uno strumento imprescindibile per valutare la qualità scientifica di una DSE, e contribuirebbe al definitivo consolidamento delle sperimentazioni e delle *best practice* attuate finora nel campo dello *scholarly digital editing*.

## 2 CARATTERISTICHE DEL MODELLO

Negli anni alle tipologie tradizionali di edizione (es. edizione critica, diplomatica, genetica, ecc.), si sono affiancate nuove tipologie *born digital*, tra cui per esempio l’*“edizione digitale documentaria”* (cfr. [5], pp 56-59). Ogni tipologia di edizione riflette un determinato approccio ecdotico ed è costituita da particolari elementi (testuali e paratestuali). In un’edizione critica, per esempio, è fondamentale la presenza dell’apparato critico. Per questo motivo il modello è suddiviso in due parti, una concettuale e l’altra prescrittiva. La parte concettuale individua, per ciascuna tipologia di edizione, i componenti astratti e le relative funzionalità, necessarie alla consultazione da parte del lettore/utente finale. La parte prescrittiva offre layout e soluzioni grafiche per presentare i diversi componenti all’interno dell’interfaccia. Infine il modello dovrà essere facilmente traducibile in template o altri tipi di risorse da poter offrire agli sviluppatori informatici.

## 3 METODOLOGIA

Dato che il modello mira a rappresentare le diverse tipologie o, più precisamente, i modelli di edizione, lo sviluppo del modello sarà avviato e testato in primis sulle edizioni critiche. Una volta constatata l’efficacia della metodologia e l’espressività del modello ottenuto, con la stessa metodologia sarà possibile espandere il modello anche alle altre tipologie di edizione.

Lo sviluppo del modello, come anticipato nella sezione precedente, sarà avviato con l’analisi dei componenti astratti, ovvero gli elementi testuali e paratestuali che solitamente costituiscono un’edizione critica. Una volta individuati i componenti, lo sviluppo proseguirà in un’approfondita analisi dello stato dell’arte. Tale analisi ha lo scopo di individuare all’interno di DSE esistenti delle soluzioni di design efficaci e facilmente riproducibili per ciascun componente, da inserire nel modello insieme a delle *best practice* consolidate sia nell’ambito dello *scholarly digital editing* sia del *web design*. Una volta costituito il modello, ne verrà testata l’adeguatezza mediante lo sviluppo di un’applicazione realizzata per pubblicare testi poetici in italiano francese e latino come edizioni critiche digitali nell’ambito del progetto ERC European Ars Nova<sup>14</sup>. L’applicazione sarà integrata nel software GattoWeb<sup>15</sup>, sviluppato dall’Opera del Vocabolario Italiano (OVI). In questo modo, il lavoro sperimentale sarà condotto in sinergia con filologi e linguisti e l’applicazione potrà essere testata anche sui corpora testuali elaborati dall’OVI.

---

14 <https://www.europeanarsnova.eu/it/>

15 <https://www.cnr.it/it/focus/062-4/il-sistema-software-gatto-gattoweb>

## 4 OBIETTIVI

I principali obiettivi della presente ricerca sono: alimentare all'interno della comunità scientifica un dibattito sulla necessità di un modello forte, di uno standard per la realizzazione di edizioni digitali; proporre una possibile metodologia per sviluppare tale standard, mediante dei modelli di DSE; offrire uno strumento che renda accessibile a un maggior numero di filologi la pubblicazione delle proprie edizioni in digitale.

## BIBLIOGRAFIA

1. Buzzoni, Marina. «A protocol for Scholarly Digital Editions? The Italian point of view». *Digital scholarly editing: Theories and practices*, 2016, 59–82.
2. Eide, Øyvind. «Ontologies, data modeling, and TEI». *Journal of the Text Encoding Initiative*, n. 8 (2014). <https://doi.org/https://doi.org/10.4000/jtei.1191>.
3. Franzini, Greta, Melissa Terras, e Simon Mahony. «Digital Editions of Text: Surveying User Requirements in the Digital Humanities». *Journal on Computing and Cultural Heritage (JOC-CH)* 12, n. 1 (2019): 1–23. <https://doi.org/https://doi.org/10.1145/3230671>.
4. Gonzalez-Perez, Cesar, e Patricia Martin-Rodilla. «Metainformation scenarios in Digital Humanities: Characterization and conceptual modelling strategies». *Revista de Humanidades Digitale* 1, n. 408 (2017): 29–48. <https://doi.org/https://doi.org/10.5944/rhd.vol.1.2017.16128>.
5. Mancinelli, Tiziana, e Elena Pierazzo. *Che cos'è un'edizione scientifica digitale*. Roma: Carocci, 2020.
6. Pierazzo, Elena. *Digital scholarly editing: Theories, models and methods*. Routledge, 2014.
7. ———. «What future for digital scholarly editions? From Haute Couture to Prêt-à-Porter». *International Journal of Digital Humanities* 1, n. 2 (2019): 209–220. <https://doi.org/https://doi.org/10.1007/s42803-019-00019-3>.
8. Sahle, Patrick. «What is a scholarly digital edition?» In *Digital scholarly editing: Theories and practices*, 19–39. Cambridge: OpenBook Publishers, 2016.

# Acquisizioni metodologiche per un'edizione critica digitale: il caso dell'Ars Breviata

Francesca Michelone  
Università del Piemonte Orientale – francesca.michelone [at] gmail.com

## SINTESI

Il poster presenta riflessioni sulla gestione di fonti testuali attraverso `xi:include` e `@resp`, sull'utilizzo delle immagini dei manoscritti e sulla relazione segni grafici-glifi Junicode, emerse nel corso del lavoro preparatorio per l'edizione critica digitale di un testo latino tardoantico.

## PAROLE CHIAVE

edizioni digitali, edizioni critiche, EVT, manoscritti, annotazione XML-TEI, `xi:Include`, `@resp`

## POSTER

### 1 L'ARS BREVIATA DIGITALE

Come segnalato da diversi studiosi ([6], [9], [12]) le edizioni critiche digitali di testi appartenenti al mondo classico e tardoantico sono poche. Tra queste si possono citare il Lucullus a cura di Malaspina [9], gli Scholia di Euripide a cura di Mastronarde [11] e progetti come Catullus online [8], DEFrag-Tragedy [3], Vindolanda Tablets online [17]. Attraverso lo studio di queste opere e una riflessione teorica basata principalmente su Franzini [4], Sahle [15], Pierazzo e Mancinelli ([10], [13]) si è consolidata l'idea di creare per l'Ars breviata attribuita ad Agostino un'edizione che sia scholarly secondo la definizione di Pierazzo e che diventi critica in senso lachmanniano [9], senza perdere di vista il paradigma digitale [15].

Dall'elaborazione di un prototipo di tale edizione che utilizza EVT [14] nascono diversi nuclei tematici di riflessione e approfondimento: in particolare la gestione di più documenti XML tramite `xi:include`; l'utilizzo delle immagini nella costruzione dell'edizione digitale; la relazione segni grafici-glifi.

## 2 GESTIONE DEI MOLTEPLICI LAYER DELL'OPERA: IMMAGINE, TRASCRIZIONE, TESTO NORMALIZZATO

L'Ars breviata è tramandata da tre testimoni, per ognuno di essi è stata prodotta una trascrizione alla quale è seguita una versione normalizzata contenente i primi interventi editoriali su ogni codice. Ai fini dell'edizione ogni testimone è costituito da un file annotato in XML-TEI [16] che contiene sia il testo trascritto, sia il testo normalizzato arricchito da una codifica strutturale (fogli, righe, capitoli, paragrafi) e una codifica specifica per i glifi, le abbreviazioni, gli interventi editoriali e il collegamento alle immagini. Ogni file/testimone è stato gestito separatamente e in un secondo momento è entrato a far parte del file/opera tramite `xi:include`. Il file/opera, (o file master) contiene il TEI-header generale completo e i marcatori `xi:include` collegati al `<text>` di ogni testimone. La codifica descritta si presta molto bene anche alla gestione di opere in più libri facilitando la produzione di edizioni collaborative in cui gli elementi inclusi non sono i diversi testimoni ma i singoli libri dell'opera. Specialmente in quest'ottica rimane fondamentale l'uso dell'attributo `@resp` per riconoscere la paternità degli interventi degli editori nel testo. Nell'Ars breviata digitale `@resp` è stato inserito nei tag `<corr>` e `<supplied>`. Si aprono così nuove prospettive di lavoro su opere complesse che coinvolgono più studiosi competenti su parti specifiche dell'insieme.

L'inserimento delle immagini nell'edizione digitale e il loro accostamento al testo reso possibile da EVT permette un livello di trasparenza delle scelte filologiche che nelle edizioni a stampa non è possibile raggiungere. L'idea-guida è che un'edizione critica renda disponibili agli studiosi-lettori non soltanto il testo e l'apparato critico, ma anche le fonti primarie sulle quali si è basato. Un'edizione critica digitale impostata in questo modo aprirebbe nuove prospettive non solo di studio ma anche didattiche, sia in ambito filologico, mostrando in modo esplicito le scelte operate dal filologo in relazione alla tradizione manoscritta, sia paleografico affiancando ogni trascrizione all'immagine del relativo manoscritto. In EVT è infatti possibile collegare l'immagine di un punto poco leggibile di un manoscritto con la trascrizione corrispondente operata dallo studioso. La codifica utilizzata permette anche di indicare il tipo di danno e la bassa leggibilità tramite gli elementi `<damage>` e `<unclear>` e ai loro attributi (`@type`, `@quantity`, `@unit`, `@reason`). Allo stato attuale del progetto sull'Ars breviata le immagini sono inserite da file locali e per ogni riga del testo sono indicate le sue coordinate nell'immagine tramite l'annotazione con i tag `<facsimile>` e `<surface>`. Uno degli obiettivi futuri è inserire le immagini in IIIF direttamente dai siti delle biblioteche che hanno fornito le immagini ([1][5][7]).

## 3 GESTIONE DEI GLIFI

Un altro aspetto che merge in evidenza dall'analisi dei manoscritti è la casistica complessa di relazioni tra i segni grafici e i glifi disponibili per la loro rappresentazione con Junicode, il font open source utilizzato per la trascrizione. Tale analisi non solo ha sollecitato riflessioni sulla codifica dei glifi e sul suo significato, ma ha anche mostrato l'importanza di uno studio approfondito delle abbreviazioni [2] e della loro normalizzazione in una fase in cui la lingua latina tardoantica è nel vivo della sua evoluzione.



## 4 CONCLUSIONI

La scelta di utilizzare strumenti open source (XML-TEI, EVT) e fonti online (le immagini fornite dalle biblioteche dei manoscritti) colloca questo lavoro in quell'ambito di progetti orientati verso l'accesso aperto con la consapevolezza che più verranno utilizzati linguaggi e software aperti, più ci sarà il modo di far circolare i dati, aumentando la loro possibilità di sopravvivenza sull'asse diacronico.

## BIBLIOGRAFIA

1. Biblioteca Apostolica Vaticana. «“OVL - VATICAN LIBRARY”. Manuscripts», s.d. <https://opac.vatlib.it/mss/>.
2. Cappelli, Adriano. *Lexicon abbreviaturarum. Dizionario di abbreviature latine ed italiane*. 4. edizione (anastatica) corredata con 9 tavole fuori testo. Milano: Hoepli, 1913.
3. DEFrAG-Tragedy. «DEFrAG-Tragedy», s.d. <https://marvelapp.com/63adb3j/screen/21255005>.
4. Franzini, Greta. «Dig-Ed-Cat Catalogue Digital Editions», 2019. <https://dig-ed-cat.ac-dh.oeaw.ac.at/>.
5. Gallica. «Gallica», s.d. <https://gallica.bnf.fr/accueil/it/content/accueil-it?mode=desktop>.
6. Italia, Paola, e Francesca Tomasi. «Filologia digitale. Fra teoria, metodologia e tecnica». *Ecdotica* 11, n. 1 (2014): 112–130.
7. KBR. «KBR», s.d. <https://www.kbr.be/en/>.
8. Kiss, Daniel. «CatullusOnline», 2017. [http://www.catullusonline.org/CatullusOnline/index.php?dir=edited\\_pages&pageID=5](http://www.catullusonline.org/CatullusOnline/index.php?dir=edited_pages&pageID=5).
9. Malaspina, Ermanno. «Il Futuro Dell'edizione Critica (Cioè Lachmanniana) Più o Meno Digitale. Riflessioni (in)Attuali». *Storie e Linguaggi* 5, n. 1 (2019): 35–60.
10. Mancinelli, Tiziana, e Elena Pierazzo. *Che cos'è un'edizione scientifica digitale*. Roma: Carocci, 2020.
11. Mastronarde, Donald J. «Euripides Scholia: Home», s.d. <https://euripidesscholia.org/>.
12. Monella, Paolo. «Why are there no comprehensively digital scholarly editions of classical texts?» In *Digital Philology: New Thoughts on Old Questions*, a cura di Adele Cipolla, 141–59. Padova: Libreriauniversitaria, 2018. <https://iris.unipa.it/handle/10447/294132#.XnH0a27cltk>.
13. Pierazzo, Elena. «Digital Scholarly Editing. Theories, Models and Methods». *Digital Scholarship in the Humanities* 31, n. 3 (2016): 513–16. <https://doi.org/10.1093/lc/fqw033>.
14. Rosselli Del Turco, Roberto. «Designing an advanced software tool for Digital Scholarly Editions»: *Textual Cultures* 12, n. 2 (29 agosto 2019): 91–111. <https://doi.org/10.14434/textual.v12i2.27690>.
15. Sahle, Patrick. «2. What is a Scholarly Digital Edition?» In *Digital Scholarly Editing: Theories and Practices*, a cura di Matthew James Driscoll e Elena Pierazzo, 19–40. Open Book Publishers, 2016. <https://doi.org/10.11647/OBP.0095.02>.
16. TEI Consortium. *The TEI Guidelines. TEI P5*. Text Encoding Initiative Consortium, 2020.
17. Vindolanda. «Vindolanda Tablets Online | Welcome», s.d. <http://vindolanda.csad.ox.ac.uk/index.shtml>.

# Language Disparity in the Interaction with Chatbots for the Administrative Domain

Martina Miliani<sup>1</sup>, Marina Benedetti<sup>1</sup>, Alessandro Lenci<sup>3</sup>

<sup>1</sup> Università per Stranieri di Siena, CoLing Lab Università di Pisa – m.miliani [at] studenti.unistrasi.it

<sup>2</sup> Università per Stranieri di Siena – benedetti [at] studenti.unistrasi.it

<sup>3</sup> Università di Pisa – alessandro.lenci [at] unipi.it

## ABSTRACT

The high impact of the Internet on citizens' daily life and the widespread use of mobile devices has led the Italian Public Administrations to communicate through the Web and digital media. Chatbots are one of the most recent technologies adopted by public institutions. This work focuses on the interaction of citizens with a chatbot able to answer questions about the administrative domain. In particular, the main objective is to identify the relevant variables involved in the reading comprehension process of texts written in the Italian administrative language. A key element of this research is represented by the target population (i.e., Italian second-language learners, elderly Italians, and Italians with a low-literacy level) to ease the access to administrative texts by people with a lack of reading skills.

## KEYWORDS

chatbot, language disparity, second-language learners, low-literacy, elderly, public administration

## POSTER

### 1 INTRODUCTION

The complexity of the Italian administrative language is well-known. The writing style adopted in administrative texts is often artificial and obscure [15], to the point of misleading the readers [12]. The Italian community shows several language disparities: Its members in fact have different levels of language proficiency [9]. Since the early Nineties, Italian institutions have encouraged Public Administrations (PAs) to adopt plain language in writing official acts and communications [11], in order to fulfill constitutional obligations concerning substantive equality of citizens [10]. Witnessing the broad access to the Internet and the extensive use of mobile devices by citizens, PAs have been increasingly choosing to communicate through innovative digital technologies, like chatbots. Chatbots, or chatterbots, are artificial conversational agents that interact with humans and answer their questions using natural language dialogues, thanks to Artificial Intelligence (AI) (mostly neural) techniques [13].

In the interaction between PA chatbots and citizens, language disparity could impede full access to administrative texts. The main goal of this work is to detect which factors affect the comprehension process of citizens with a lack of reading skills when reading administrative texts, in the interaction with AI systems. The long-term aim of this study is to increase the legibility of these texts and facilitate access information related to the administrative domain.

## 2 METHODOLOGY

According to [10], text comprehension can be seen as the result of the interaction among a subject with his own socio-cultural and sociolinguistic identity, the communication context, and the key element of the communicative event, i.e., the text. Text simplification based on complexity linguistic features can improve the legibility of texts. However, it may not be sufficient to guarantee text legibility in the AI-human interaction on digital devices, especially if the readers are not fully proficient in the Italian language. Reading comprehension tests represent an essential tool to assess legibility when designed to observe and measure all the variables affecting text comprehension. In fact, text complexity can be defined as a function of the linguistic features of a text (i.e., the administrative language), the variables related to specific targets (i.e., members of a community with language disparities), and the reading context, which includes the purpose of reading and the medium used (i.e., new digital technologies).

Our experiment goal is to assess the effectiveness of text simplification based on linguistic traits within this context. To this purpose, informants will be asked to read administrative texts in their original or simplified versions. Such texts consist of administrative acts and city halls' web pages content. Text simplification is evaluated through readability assessment tools based on Natural Language Processing (NLP) techniques [4]. The result of this process is a parallel corpus, where administrative texts are split into sentences and coupled with their simplified version. Readers will be also asked to answer a small set of questions related to each text. The readability level of test and items rubrics will be assessed using NLP tools as well.

### 2.1 Italian Administrative Language

The Italian administrative language is defined by [10] as a sectorial variety of Italian, along with other language varieties *for special purposes*. A fundamental step for the simplification of administrative texts is to detect the linguistic features of the administrative language. The use of *pseudo-technicisms* (e.g., “balneazione”, “fattispecie”) and formal terms (e.g., “ovvero”, “allorché”, “suddetto”), the predominance of hypotaxis over parataxis, and a writing style focused on the writer rather than on the reader can be considered among such linguistic traits [4]. Finally, it is necessary to distinguish the features that can be automatically extracted with computational linguistics techniques from those that can be detected only by human experts.

## 2.2 Language Disparity

We selected four groups of informants: Italian second-language learners, elderly Italians, Italians with a low-literacy level, and a control group of Italians with medium-high-literacy level. For what concerns the first group, supporting the access in reading administrative texts could be especially useful for learners with A2 proficiency level<sup>1</sup> [14]. The A2 level is a basic requirement to obtain a long stay permit in Italy [2], whereas B1 learners are already independent in second language communication [20]. Italian low-literacy readers are those subjects who did not obtain a high school certificate. Among them, there are also *functional illiterates*, who have lost or never developed the ability to write or read a text about ordinary events or problems of social interest [9]. Finally, we consider an elderly a person who is at least 65 years old. In 2019, 72.66% of elderly people in Italy had early secondary school education at most<sup>2</sup>. Besides the decline in cognitive resources [7], elderly's reading skills are influenced by the reader's previous knowledge and even more by their reading habits [16].

## 2.3 New Technologies

Since the test will be taken remotely on mobile devices, variables concerning the text physical support will be considered as well. Digital media affect the reading comprehension process, especially in relation to the text length, the readers' age, and electronic device features, such as the screen brightness [18]. Finally, we want to examine how readers' comprehension is affected while reading administrative texts returned by a chatbot based on AI. Interfaces design has become a central issue in Digital Humanities. Design choices affect Human-Computer Interaction since they can enhance or damage any digital experience [5]. To improve the usability of digital tools, designers decided to build them adopting metaphors from real life: Macintosh organized the screen as a desktop, whereas Facebook based its usage on the concept of friendship [8]. Chatbots ease access to information by imitating the users' interaction with real operators, although such technology still shows some limitations. According to recent studies users are rather reluctant to interact with conversational agents, which are not able to establish a satisfactory relation since they lack the sensibility of human operators [19]. The first chatbot, ELIZA, was created to simulate a dialogue with a Rogerian Psychologist. It was implemented in 1966 and based on a pattern-matching rule-based system [21]. Nowadays, chatbots are defined as non-task-oriented dialogue systems, usually implemented adopting either generative methods or retrieval-based techniques [6]. The formers are usually based on AI techniques: neural networks are fed with dialogue data and return answers generated through statistical computations. In the letters, retrieval-based techniques match the user's message with the correct stored answer. State-of-art systems adopt neural architectures to label the correct answer to each message [6].

---

1 According to CEFR (Common European Framework of Reference for Languages)

2 Source: Istat (Italian Institute of Statistics).

### 3 EXPECTED RESULTS AND FUTURE WORK

The results obtained from this reading comprehension test can be used to make predictions on the actual level of informants' comprehension, by applying a generalization process [1]. If readers find the simplified texts still hard to read, this test will be a useful tool to detect features affecting their comprehension process, other than complexity linguistic traits.

Future work will also include the design of suitable simplification strategies which will involve specific features related to each group of readers, the medium used, and, more in general, the examined communication context. Neural architectures already employed in machine-translation tasks [3] and models exploiting contextual word embeddings [17] will be evaluated for this purpose.

An interesting development of this work could consider a shifting along the diamesic axis. People with visual impairments, like some elderly people, may prefer to ask for such information vocally. For this reason, comprehension texts based on speech recognition tasks could ensure the access to this part of the population too.

### 4 CONCLUSION

PAs employ new digital technologies for communicating with citizens, in order to reach a higher number of users. This research aims at identifying the variables involved in the comprehension of administrative texts by people not fully proficient in the Italian language using digital media, to enhance their digital and democratic inclusion within the Italian citizenry.

### REFERENCES

1. Alderson J., Charles. *Assessing reading*, 2000.
2. Borri, Alessandro. *Insegnare Italiano L2 ad apprendenti analfabeti: contesto e indicazioni didattiche*. Loescher, 2014.
3. Botarleanu, Robert-Mihai, Mihai Dascalu, Scott Andrew Crossley, e Danielle S. McNamara. «Sequence-to-sequence models for automated text simplification». In *International Conference on Artificial Intelligence in Education*, 31–36. Springer, 2020.
4. Brunato, Dominique. «A study on linguistic complexity from a computational linguistics perspective. A corpus-based investigation of italian bureaucratic texts.» Ph.D. Thesis, University of Siena, 2015.
5. Burdick, Anne, Johanna Drucker, Peter Lunenfeld, Todd Presner, e Jeffrey Schnapp. *Digital Humanities*. Mit Press, 2012.
6. Chen, Hongshen, Xiaorui Liu, Dawei Yin, e Jiliang Tang. «A survey on dialogue systems: Recent advances and new frontiers». *Acm Sigkdd Explorations Newsletter* 19, n. 2 (2017): 25–35.
7. Davies, Rob AI, Ruth Arnell, Julia MH Birchenough, Debbie Grimmond, e Sam Houlson. «Reading through the life span: Individual differences in psycholinguistic effects.» *Journal of Experimental Psychology: Learning, Memory, and Cognition* 43, n. 8 (2017): 1298.
8. De Biase, Luca. *I media civici*. Feltrinelli, 2013.
9. De Mauro, Tullio. *L'educazione linguistica democratica*. Gius. Laterza & Figli Spa, 2018.

10. De Mauro, Tullio, e Massimo Vedovelli. *Dante, il gendarme e la bolletta: la comunicazione pubblica in Italia e la nuova bolletta Enel*. Laterza, 1999.
11. Fortis, Daniele. «Il linguaggio amministrativo italiano». *Revista de Llengua i Dret* 43 (2005): 47–116.
12. Franceschini, Francesco, e Sara Gigli. *Manuale di scrittura amministrativa*. Agenzia delle entrate, 2003.
13. Hill, Jennifer, W. Randolph Ford, e Ingrid G. Farreras. «Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations». *Computers in human behavior* 49 (2015): 245–250.
14. Lombardi, Giulia. «Capire i documenti in L2: dall’analisi della comprensibilità di un corpus di testi istituzionali per stranieri alla sperimentazione di soluzioni efficaci», 2016. <https://doi.org/10.13130/2037-3597/7566>.
15. Lubello, Sergio. *Il linguaggio burocratico*. Carocci, 2014.
16. Miller, Lisa M. Soederberg, e Elizabeth AL Stine-Morrow. «Aging and the effects of knowledge on on-line reading strategies». *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences* 53, n. 4 (1998): P223–P233.
17. Qiang, Jipeng, Yun Li, Yi Zhu, Yunhao Yuan, e Xindong Wu. «A simple BERT-based approach for lexical simplification». *arXiv preprint arXiv:1907.06226*, 2019.
18. Singer, Lauren M., e Patricia A. Alexander. «Reading on paper and digitally: What the past decades of empirical research reveal». *Review of educational research* 87, n. 6 (2017): 1007–1041.
19. Van Pinxteren, Michelle ME, Mark Pluymaekers, e Jos GAM Lemmink. «Human-like communication in conversational agents: a literature review and research agenda». *Journal of Service Management*, 2020.
20. Vedovelli, Massimo. «20 L’italiano degli stranieri; l’italiano fuori d’Italia (dall’Unità)». *Manuale di linguistica italiana* 13 (2016): 459.
21. Weizenbaum, Joseph. «ELIZA—a computer program for the study of natural language communication between man and machine». *Communications of the ACM* 9, n. 1 (1966): 36–45.

# Capturing Political Polarization of Reddit Submissions in the Trump Era

Virginia Morini<sup>1</sup>, Laura Pollacci<sup>2</sup>, Giulio Rossetti<sup>3</sup>

<sup>1</sup>ISTI-CNR – virginia.morini [at] isti.cnr.it

<sup>2</sup>Università di Pisa – laura.pollacci [at] di.unipi.it

<sup>3</sup>ISTI-CNR – giulio.rosselli [at] isti.cnr.it

## ABSTRACT

The American political situation of the last years, combined with the incredible growth of Social Networks, led to the spreading of political polarization phenomenon online. Our work presents a model that attempts to measure Reddit political polarization of submissions during the first half of Donald Trump's presidency. For such a purpose, we design a text classification task in which the political polarization of submissions is assessed by quantifying their degree of alignment with pro-Trump ideologies and vice versa. First, we build our ground truth by picking submissions from subreddits known to be strongly polarized. Then, we perform a model selection by testing different hyper-parameters of a Neural Network with word embeddings and a Long Short-Term Memory layer. Finally, we assess model performances on both the test set and three less polarized corpora.

## KEYWORDS

political polarization, classification, text analysis

## POSTER

### 1 INTRODUCTION

During the last decade, the rise of social networks has drastically changed how people interact and communicate. The number of opinions shared publicly among users is increasing, and particular attention turned on the diffusion of political discourse online and its implications. The sharing of political beliefs, combined with the contemporary political situation, leads to political polarization. This phenomenon refers to the increasing gap between two different political ideologies. During Donald Trump's presidency, online polarization found its fertile ground in the debate between Trump supporters and anti-Trump citizens [1].

This paper proposes a model to quantify Reddit submissions political polarization in the Trump Era. We model the issue as a text classification task. Given new submissions, we assess their political polarization by quantifying their degree of alignment with pro-Trump or anti-Trump ideologies. To assess the existence of a strongly polarized environment, we build a model to measure its components polar-

ization. As a case study, we choose Reddit because of its internal structure which is composed of thousands of subreddits. This makes it easier to identify homogenous communities concerning a specific topic. Additionally, since users can write anonymously and posts are not limited in length, this platform is particularly active in political discussions [11].

The rest of the paper is organized as follows. Section 2 discusses the literature on text classification in general and with respect to the political domain. Section 3 describes the data collection and preparation to build the dataset. Section 4 proposes the model selection, while Section 5 concludes the paper and sets future research directions.

## 2 RELATED WORKS

With the growth of Social Networks, researchers focused on studying methods to extract information from a large quantity of unstructured data. For such a purpose, text preprocessing plays a key role in data cleaning, allowing to remarkably improve text classifier performances [4,16]. Furthermore, with classification tasks, we also need to take into account the suitable type of word representation. Since traditional ones, i.e., bag-of-words, encode words as discrete symbols not directly comparable to others [8], they are not fully able to model semantic relations between words. Word embedding (e.g., Word2vec [10] and Glove [12]), instead, map words to a continuously-valued low dimensional space, capturing their semantic and syntactic features. Also, their structure makes them suitable to be deployed with Deep Learning models, fruitfully used to address NLP-related classification tasks. Specifically, as shown in this survey [9], Recurrent Neural Networks (RNNs) have been proven to be extremely successful for sequence learning. Among them, Long Short-Term Memory (LSTM) network can maintain long-term dependencies through an elaborate gates mechanism, overcoming the vanishing gradient problem of standard RNNs [15].

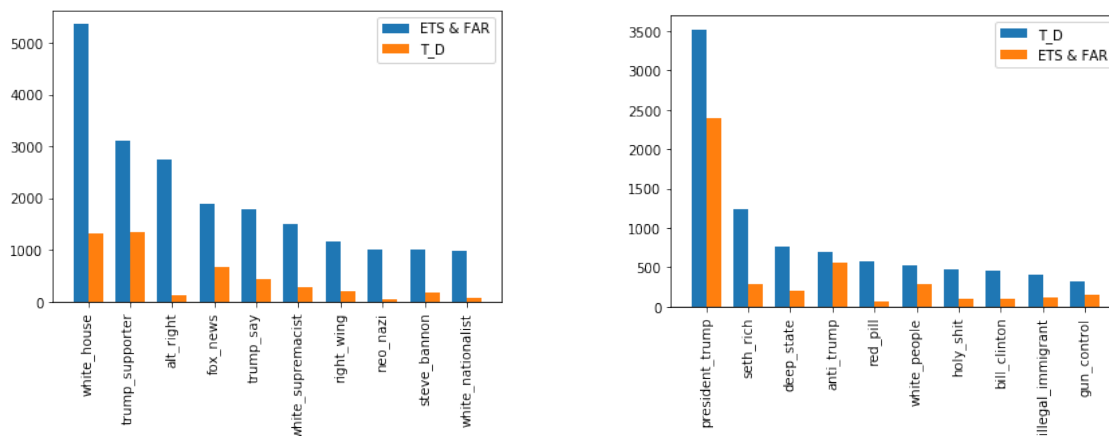


Figure 1: top 10 most frequent bigrams of subreddits and their frequencies in opposite political ideology.

Finally, concerning political leaning classification on Social Network textual data, in the last few years have been presented several encouraging works. For example, Chang et al. [5] propose a model to predict the political affiliation of Facebook posts in Taiwan.



They build two models, one using the K-NN algorithm and the other one using AdaBoost combined with Naive Bayes classifier. Instead, Rao et al. [14] use word embeddings and LSTM to predict if a Twitter post deals with Republican or Democratic beliefs. To the best of our knowledge, research about political classification on Reddit is sparse or nonexistent.

### 3 DATA DESCRIPTION AND PREPARATION

We collect submissions belonging to known pro-Trump and anti-Trump subreddits in the first two years and a half of Donald Trump's presidency (Jan. 2017 - May 2019) by using the Pushshift API [2]. We choose r/TheDonald (151,395 posts) as pro-Trump, and r/Fuckthealtright (78,200 posts) and r/EnoughTrumpSpam (73,168 posts) for anti-Trump, to obtain a balanced dataset. For each submission, we collect the fields id, selftext, and title, respectively, the identifier, the content, and the title of the submission. We merge the last two fields in a unique one, because the selftext of a post may be empty or just a reference to the title itself. Then, we label submissions with 1 if pro-Trump, 0 otherwise. Furthermore, we remove noise due to free form writing by applying a standard preprocessing pipeline to give clean data to LSTM. Specifically, we convert text to lowercase, then remove punctuation and number, as well as stop words. Lastly, to check the validity of our initial choice of polarized subreddits, we identify some of the most frequent bigrams of each subreddit to analyze their frequencies in the opposite one. As shown in Figure 1, these words are discriminant and semantically related to their belonging subreddit.

### 4 MEASURING THE POLITICAL POLARIZATION

To assess the best model, we preprocess input text sequences, we perform model selection, and we test it on new instances.

**Neural Network Preprocessing.** We vectorized submissions into sequences of integers to pass them as model inputs. Thus, we build a lexicon index based on word frequency (0 is for padding), and we replace submissions words with their index in the lexicon (e.g., "president trump says" becomes [5,1,39]). Also, we pad sequences to the same length  $l$  to optimize the batch matrix operations, based on the mean length of all submissions.

**Neural Network Architecture.** Since our training data consists of sequences, standard Neural Networks (NNs) are not suitable for our task. Thus, we choose the LSTM network [6] for its ability to model a sentence meaning by considering its paradigmatic structure.

| Emb. type      | LSTM units | Training Acc. | Validation Acc. |
|----------------|------------|---------------|-----------------|
| <i>Learned</i> | 32         | 0.832         | 0.816           |
|                | 64         | 0.837         | 0.816           |
|                | 128        | 0.828         | 0.816           |
| <i>Glove</i>   | 32         | 0.831         | 0.814           |
|                | 64         | 0.839         | 0.819           |
|                | 128        | <b>0.846</b>  | <b>0.829</b>    |

Table 1. Hyper-parameters tuning for models with learned and Glove embeddings.

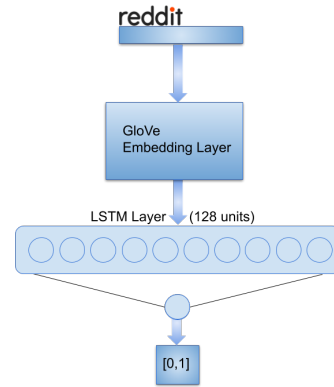


Figure 2. High-level architecture of our LSTM model.

Figure 2 shows the high-level architecture of our model, composed of three layers:

1. The Embedding layer takes as input the sequences of integers  $w_1, \dots, w_t$  and turns them into 100-dimensional dense vectors  $x_t$ , where  $t$  is the length of the sequence. Specifically, we experience both pre-trained and learned word embeddings (i.e., embeddings learned directly from our corpora).
2. The LSTM layer consists of multiple LSTM units, each maintaining a memory cell. Cells encode the information of the inputs observed up to that step through the gates mechanism. The input gate controls whether the memory cell is updated. The forget gate controls if the cell is zero; lastly, the output gate controls whether the cell state information is visible. To avoid overfitting, we add a dropout regularization of 0.3.
3. The Output layer is a fully connected layer which outputs a single neuron to perform binary predictions. We use the Sigmoid activation function to have a probability output between 0 and 1. Consequently, we label submissions with a probability score of  $\geq 0.5$  as pro-Trump and as anti-Trump otherwise. As a loss function, we use Binary cross-entropy and as optimizer Adam.

**Model Selection.** We use a training set of 242,763 instances, balanced among the two political ideologies. We perform a 3-fold Cross-Validation trying different values of LSTM units (32,64,128) and embeddings (learned and pre-trained), with a fixed 100 embedding dimension. As detailed in Table 1, the model with Glove pre-trained embeddings and 128 LSTM units achieves the best accuracy in training and validation sets (84,6% and 83%, respectively).

**Model Evaluation.** On the test set, the best model reaches an encouraging accuracy of 84,3% (Table 2). We further assess model performances on three less polarized sociopolitical topics (Jan. 2017 - Dec. 2019), i.e., gun control, minority discrimination, political discussion. Given the absence of ground truth, we validate our model through polarized users. We compute users' polarization scores, and we select the most polarized ones, i.e.,  $\leq 0.2$ , and  $\geq 0.8$ , to label submissions.

Finally, we assess model performances on the three sets to evaluate its ability to generalize. Despite differences in size, the model always reaches an accuracy of over 72%.

## 5 CONCLUSION AND FUTURE WORKS

This study proposes an approach to measure the degree of political polarization of Reddit submissions in the Trump Era. We design a NN with word embeddings and LSTM layer to quantify submissions political alignment. The model with Glove embeddings achieves 83% and 84,3% accuracy, respectively on validation and test set. Such results are in line with previous works on different Social Networks, i.e., 86,3% on Facebook [5] and 85% on Twitter [14]. Also, we assess model performances on other three less polarized topics to evaluate its ability to generalize. We achieve quite good results with accuracy ranging from 72% to 82%. As a future research direction, it would be interesting to extend our case study, including data related to other polarizing public figures, in such a way to give a broader contribution to the literature.

| Dataset                        | #post  | Accuracy |
|--------------------------------|--------|----------|
| <i>Test set</i>                | 60,000 | 0.843    |
| <i>Gun control</i>             | 2,411  | 0.712    |
| <i>Minority discrimination</i> | 4,839  | 0.732    |
| <i>Political discussion</i>    | 46,339 | 0.721    |

Table 2: model performances on the test set and the less polarized topics

## REFERENCES

1. Bartels, Larry M. ‘Partisanship in the Trump Era’. *The Journal of Politics* 80, no. 4 (2018): 1483–1494.
2. Baumgartner, Jason, Savvas Zannettou, Brian Keegan, Megan Squire, and Jeremy Blackburn. ‘The Pushshift Reddit Dataset’. *ArXiv:2001.08435 [Cs]*, 23 January 2020. <http://arxiv.org/abs/2001.08435>.
3. Boulianne, Shelley. ‘Social Media Use and Participation: A Meta-Analysis of Current Research’. *Information, Communication & Society* 18, no. 5 (2015): 524–538.
4. Camacho-Collados, Jose, and Mohammad Taher Pilehvar. ‘On the Role of Text Preprocessing in Neural Network Architectures: An Evaluation Study on Text Categorization and Sentiment Analysis’. *ArXiv:1707.01780 [Cs]*, 23 August 2018. <http://arxiv.org/abs/1707.01780>.
5. Chang, Che-Chia, Shu-I. Chiu, and Kuo-Wei Hsu. ‘Predicting Political Affiliation of Posts on Facebook’. In *Proceedings of the 11th International Conference on Ubiquitous Information Management and Communication*, 1–8, 2017.
6. Hochreiter, Sepp, and Jürgen Schmidhuber. ‘Long Short-Term Memory’. *Neural Computation* 9, no. 8 (1997): 1735–1780.

7. Kowsari, Kamran, Kiana Jafari Meimandi, Mojtaba Heidarysafa, Sanjana Mendu, Laura Barnes, and Donald Brown. 'Text Classification Algorithms: A Survey'. *Information* 10, no. 4 (2019): 150.
8. Lipton, Zachary C., John Berkowitz, and Charles Elkan. 'A Critical Review of Recurrent Neural Networks for Sequence Learning'. *ArXiv:1506.00019 [Cs]*, 17 October 2015. <http://arxiv.org/abs/1506.00019>.
9. Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. 'Efficient Estimation of Word Representations in Vector Space'. *ArXiv:1301.3781 [Cs]*, 6 September 2013. <http://arxiv.org/abs/1301.3781>.
10. Nithyanand, Rishab, Brian Schaffner, and Phillipa Gill. 'Online Political Discourse in the Trump Era'. *ArXiv:1711.05303 [Cs]*, 14 November 2017. <http://arxiv.org/abs/1711.05303>.
11. Pennington, Jeffrey, Richard Socher, and Christopher D. Manning. 'Glove: Global Vectors for Word Representation'. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543, 2014.
12. Pratama, Bayu Yudha, and Riyanarto Sarno. 'Personality Classification Based on Twitter Text Using Naive Bayes, KNN and SVM'. In *2015 International Conference on Data and Software Engineering (ICoDSE)*, 170–174. IEEE, 2015.
13. Rao, Adithya, and Nemanja Spasojevic. 'Actionable and Political Text Classification Using Word Embeddings and LSTM'. *ArXiv:1607.02501 [Cs]*, 13 July 2016. <http://arxiv.org/abs/1607.02501>.
14. Sundermeyer, Martin, Ralf Schlüter, and Hermann Ney. 'LSTM Neural Networks for Language Modeling'. In *Thirteenth Annual Conference of the International Speech Communication Association*, 2012.
15. Uysal, Alper Kursat, and Serkan Gunal. 'The Impact of Preprocessing on Text Classification'. *Information Processing & Management* 50, no. 1 (2014): 104–112.

# CRMtex. An ontological model for ancient textual entities

Francesca Murano<sup>1</sup>, Achille Felicetti<sup>2</sup>

<sup>1</sup> Università di Firenze – francesca.murano [at] unifi.it

<sup>2</sup> Università di Firenze – achille.felicetti [at] pin.unifi.it

## ABSTRACT

This poster presents the CRMtex, an ontological model based on CIDOC CRM developed since 2015 to support the study of ancient documents by identifying relevant textual entities and by modelling the scientific process related to the investigation of ancient texts and their features in order to foster integration with other cultural heritage research fields.

CRMtex is intended to identify and define in a clear and unambiguous way the main entities involved in the study and edition of ancient handwritten texts and to describe them by means of appropriate ontological instruments in a multidisciplinary perspective.

The CRMtex extension provides tools for managing this kind of complexity by defining classes and properties for describing a handwritten text in all its aspects, from its creation (and/or destruction) in the past, down to its present conservation, investigation and study by scholars, including its transcription, translation, interpretation and publication.

The full compatibility of CRMtex with the CIDOC CRM ontology and its extensions ensures persistent interoperability of data encoded by means of its entities with other semantic information produced in cultural heritage and digital humanities.

## KEYWORDS

ancient texts, ontologies, linguistics, CIDOC CRM, epigraphy

## POSTER

### 1 INTRODUCTION AND BACKGROUND

Although from the semiotic point of view there is a single mechanism of production of written texts (i.e. supports, techniques, etc. are not relevant), traditionally, their study falls within different disciplines, generally grown around the specific characteristics of each class of documents (e.g., papyrology for the study of papyri and epigraphy for inscriptions). Nevertheless, an interdisciplinary approach is essential, and the identification of common elements is paramount in order to confer uniformity and interoperability to all these disciplines, as well as to exploit complementary skills from different approaches.

In comparison to modern digital or printed texts, ancient ones are characterised by their uniqueness, being they the result of manual work rather than a mechanised process, as occurs with modern printing. Therefore, a special relationship exists between the texts and its supports. This and other characteristics make the study and digitisation of this type of documentation particularly arduous: the close relationship between the text and its support requires careful analysis since they are inextricably linked to form a unique object of study.

The gigantic integration effort established by Papyri.info<sup>1</sup> and Trismegistos<sup>2</sup> and the various attempts made by projects such as EAGLE<sup>3</sup> to develop a semantic model in the field of epigraphy, testify to a constant and growing interest in the use of advanced and efficient conceptual tools for the generation of standardised, integrated and interoperable information in these disciplines. In the same perspective, many European and international initiatives are also focussing their attention on information concerning ancient texts and on the interoperability challenges in which they are involved<sup>4</sup>.

## 2 OBJECTIVES

The first aim of CRMtex is to identify and define in a clear and unambiguous way the main entities involved in the study and edition of ancient handwritten texts and then to describe them by means of appropriate ontological instruments in a multidisciplinary perspective [1][2].

The concept of “written text”, provided by CRMtex, is intended to identify a common entity consisting in a particular feature (i.e., set of glyphs) created (i.e., written) on various kinds of support, having semiotic significance and the declared purpose of conveying a specific message towards a given recipient or group of recipients. The modelling of the scientific autoptic examination of the document, consisting of an accurate analysis of the surface and the signs and prescribing the use of specific tools and procedures, is another key aspect taken into account by the CRMtex extension.

In addition to dealing with text as an object, our model also focuses on the aspects of the research and provides classes and relationships to describe the typical operations that scholars from different disciplines put in place to gain knowledge about texts. It is evident, in this perspective, that the study of ancient texts typically starts from the analysis of the physical characteristics of the text itself before moving to the investigation of their archaeological, palaeographic, linguistic and historical features.

The model we propose allows to describe the textual entities in their entirety, capturing the differences between the various writing systems. We intend to enrich this model through a more detailed semantic coding of the relationship between grapheme and codified linguistic unit, taking into account not only the phonographic scripts, but the glottographic scripts as a whole. A future development of the model will allow to extend the description also to writing systems based on logographic and ideographic models [5].

---

1 <http://papyri.info>

2 <http://www.trismegistos.org>

3 <https://www.eagle-network.eu>

4 <http://www.parthenos-project.eu>

### 3 METHODOLOGIES AND EXPECTED OUTCOMES

The need to create a new ontology for ancient texts started from the assumption that, unlike printed texts, non-mechanised written texts (including inscriptions, papyri and manuscripts) have specific features that must be taken into account for their study.

Since writing is an intellectual process aimed at the encoding of a language, CRMtex has its foundation in the semiotic aspects of language and text. The core concept of our model is therefore the notion of “text” as the product of a semiotic process involving an encoding (“writing”) and a decoding (“reading”) process. Writing is in turn a particularly sophisticated human technology allowing the encoding of a linguistic message through a series of signs specifically selected for this purpose [3][4].

In writing, as in any semiotic system, every component (sign) possesses a dual nature, one physical and another conceptual. Writing, therefore, appears as a code requiring an encoding process by the creator or writer and a decoding one by the receiver or reader to be properly understood. It is necessary to distinguish between the physical manifestation of the text, understood as a set of physical features shown on a given support through the use of a specific technique (e.g. scribbled with ink, painted, engraved, etc.), from its abstract dimension, i.e. from the set of mental images represented by these same physical features.

We have based our model on the solid foundations of CIDOC CRM<sup>5</sup> because it constitutes one of the most widely used ontologies in the field of Cultural Heritage. In its core version, CIDOC CRM already provides most of the entities necessary to model common elements such as actors, objects, places, events and their mutual interrelations on a chronological basis.

Because of the native ability of CRMtex to describe relationships between text and artefacts by efficiently placing the text in the context of the life and history of ancient objects, our model has been chosen as a candidate for the encoding and integration of inscription and graffiti data in the semantic infrastructure that the ARIADNEplus project<sup>6</sup> is building. CRMtex has also been selected among the basic models for the ontology in the process of definition by the community of epigraphists in the framework of the epigraphy.info initiative<sup>7</sup> for the interoperability of epigraphic data.

---

5 <http://www.cidoc-crm.org/Version/version-6.2.3>

6 <https://ariadne-infrastructure.eu>

7 <http://epigraphy.info>

## REFERENCES

1. Felicetti, Achille, and Francesca Murano. 'Scripta Manent: A CIDOC CRM Semiotic Reading of Ancient Texts'. *International Journal on Digital Libraries* 18, no. 4 (2017): 263–270.
2. Felicetti, Achille, Francesca Murano, Paola Ronzino, and Franco Niccolucci. 'CIDOC CRM and Epigraphy: A Hermeneutic Challenge.' In *EMF-CRM@ TPD*, 55–68, 2015. <http://ceur-ws.org/Vol-1656/paper5.pdf>.
3. Haas, Gisèle. 'Sémiologie de l'écriture'. *Linx* 31, no. 2 (1994): 173–178.
4. Harris, Roy. *Signs, Language, and Communication: Integrational and Segregational Approaches*. Psychology Press, 1996.
5. Pulgram, Ernst. *Writing Without Letters*. Vol. 4. Manchester University Press, 1976.



# MIMA: a data model to represent multi-disciplinary analysis on manuscripts. Use case on Pellegrino Prisciani's *Historiae Ferrariae*

Valentina Pasqual<sup>1</sup>, Marilena Daquino<sup>2</sup>, Francesca Tomasi<sup>3</sup>

<sup>1</sup> Università di Bologna – valentina.pasqual2 [at] unibo.it

<sup>2</sup> Università di Bologna – marilena.daquino2 [at] unibo.it

<sup>3</sup> Università di Bologna – francesca.tomasi [at] unibo.it

## ABSTRACT

This research aims to explore and advance scholars' understanding of the coexistence of multiple interpretations in a formal framework, by providing a practical solution to represent them. The state of the art, the adopted methodology and approach, and the results applied on a significant case study are shown. The research has been narrowed by working on a use case – the collection of lectures called *Scrivere, rappresentare, conoscere nel rinascimento. Pellegrino Prisciani, un intellettuale eclettico tra la corte e il mondo*. As a result, MIMA (Multi-disciplinary Interpretations model on Manuscript Apparatus) aims to formally represent these aspects by leveraging Semantic Web technologies and the systematic reuse of already existing ontologies.

## KEYWORDS

semantic web, cultural heritage, manuscript, data modelling, interpretation, contestability

## POSTER

### 1 INTRODUCTION

Despite Digital Humanities research widely addresses aspects related to knowledge representation of cultural heritage, there is a lack of formal representations addressing multi-disciplinary interpretations over artefacts along with their contestability. Considering this gap in literature, this research wants to investigate which are the underlying methods of experts' interpretation – extracted by the use case which presents philological, iconographical and palaeographic analyses over Pellegrino Prisciani's *Historiae Ferrariae* illuminated manuscript - and which is the best solution to formalise such information in an expressive and efficient way. It focuses on experts' comments peculiarities and shared patterns -, and the interpretative act underlying each of them, considering their coexistence into a single environment while preserving their contestability (Daquino, Pasqual and Tomasi 2020).

## 2 STATE OF THE ART

Dealing with cultural heritage and semantic technologies means that knowledge acquisition with domain experts is required (i.e. metadata definition, content interpretation). *What* an expert state about a cultural object has been broadly investigated by CH domain ontologies such as FRBR (IFLA Study Group 2008) – for bibliographic records, along with FRBR-aligned ontologies (i.e. FaBio<sup>1</sup>); CIDOC CRM – designed as golden standard for representing cultural objects life cycle in museum domain (Doerr 2009) along with its extensions (e.g. CRMtex<sup>2</sup>); EDM – base ontology to represent descriptions provided by Europeana content providers (Charles and Isaac 2015); also handling with manuscript representation (e.g. STITCH<sup>3</sup>, DM2E model<sup>4</sup>, Sharing Ancient Wisdoms (SAWS<sup>5</sup>), Henry the III Fine Rolls (FRH3<sup>6</sup>)).

For what concerns *how* experts state their interpretations, Van den Akker (2011) propose a model to express hermeneutic analysis on artworks about historical content. VIR<sup>7</sup> is a CIDOC extension model to represent visual recognition on iconographical objects. HiCo<sup>8</sup> expresses specific contextual information about an interpretative act on artefacts. *Bufalini Notebook*<sup>9</sup> reuses HiCo and nanopublication<sup>10</sup> (Groth et al. 2010) to represent *Quaderno di appunti* dense annotation network together with its contextual information (Daquino, Giovannetti and Tomasi 2019).

## 3 TITLE

This research is based on the design-science method proposed by (Hevner et al. 2004, 75–105), which seeks to extend the boundaries of human and organizational capabilities by creating new and innovative artefacts.

Considering the qualitative nature of this research, it has been approached firstly with document analysis (Bowen 2009) on source material, involving skimming (superficial examination to choose excerpts which clearly express disciplines points of view on the same manuscript element), reading (thematic analysis and pattern recognition within the data, seeking for categories for analysis), and interpretation (clarification of incomplete or implicit knowledge in the excerpts). Then, mind maps for each case study have been created to set all the pieces of content in a triple-fashion structure that will be refactored in ontological terms.

---

1 <https://sparontologies.github.io/fabio/current/fabio.html>

2 <http://www.cidoc-crm.org/crmtext/>

3 <https://www.cs.vu.nl/STITCH/>

4 <https://github.com/DM2E/dm2e-ontologies>

5 <http://purl.org/saws/ontology>

6 <https://data.kdl.kcl.ac.uk/dataset/frh3>

7 <https://ncarboni.github.io/vir/>

8 <http://hico.sourceforge.net/>

9 <http://projects.dharc.unibo.it/bufalini-notebook/introduction>

10 <http://www.nanopub.org/nschema>

Data modelling activity consisted in translating mind maps in competency questions addressing aspects emerged from case studies (i.e. motivating scenarios). Then, a model that does not include any term belonging to any other existing ontology has been sketched, so as to represent pragmatically the scenario originally expressed in natural language. Finally existing ontologies literature has been reviewed and terms have been selected from existing ontologies so as to refactor the model.

MIMA<sup>11</sup> has been reapplied to case studies, first to test each discipline representativeness and then to test points of interaction between information. The model consistency has been tested through the use of a reasoner. Then a toy RDF dataset has been produced and queried (translating competency questions in SPARQL<sup>12</sup>) to test its expressivity.

## 4 RESULTS: MODEL AND CASE STUDY

The result is a four layered model organised in nanopublication structure, as shown in figure 1.

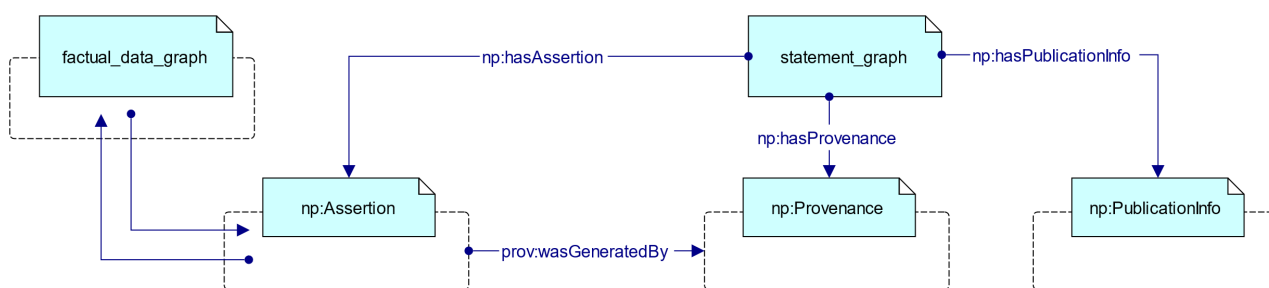


Figure 1: MIMA macrostructure showing how RDF triples are stored and connected using named graphs and nanopublications

Figure 1 shows how the statement graph is connected with the four layers, expressed in MIMA as named graphs<sup>13</sup>. For each layer several ontologies have been reused, integrated, and extended if needed.

Layer 0: Factual data graph includes the information deemed not to be questionable. It includes bibliographic metadata (e.g. the edition of a work), the physical and logical description of artefacts (e.g. the folios). Artefacts are considered as an *unicum* (i.e. the manuscript physical structure, cited artworks) and have been modelled using CIDOC, VIR and CRMtex. Serial artefacts (i.e. books editions) have been modelled using FaBiO;

11 <https://mima-data-model.github.io/mima-documentation/>

12 <https://www.w3.org/TR/rdf-sparql-query/>

13 <https://www.w3.org/2009/07/NamedGraph.html>

Layer 1: Assertion graph includes scholars' questionable and/or competing statements formulated over the artefact. Assertions on the iconographical apparatus have been modelled using VIR ontology; events, places, and periods involved in the manuscript history or described in the manuscript have been modelled with CIDOC; relations between artefacts (e.g. manuscript sources and influences), have been described with CIDOC and OA Entry ontology<sup>14</sup>; handwriting analysis has been refactored with CRM-tex and CIDOC.

Layer 2: Provenance graph includes contextual information on a statement (np:Assertion prov:wasGeneratedBy np:Provenance). Sources, motivations, types and performers of interpretative acts are represented by using terms of HiCo and PROV; the interpretation certainty degree is expressed reusing CWRC ontology<sup>15</sup> terms.

Layer 3: Publication information graph represents the meta-context of a statement that has been automatically or semi-automatically generated. Publication information has been addressed with terms from PROV ontology.



Figure 2: Historiae Ferrariae, Vol. I, ASMo, ms. 129, proem

14 <http://oentry-ontology.sourceforge.net/>

15 <http://sparql.cwrc.ca/ontologies/cwrc#>

Consider the following scenarios in natural language (MS) referred to *Historiae*, Vol. I, proem - shown in figure 2.

- **Philologic statement (MS1):** The illustration represents Prisciani giving *Historiae* manuscript to Ercole I d’Este;
- **Palaeographic statement (MS2):** The monumentality of *Historiae* is conveyed by both ‘capitale epigrafica’ in the title and the illustration;
- **Iconographic statement (MS3):** The proem illustration depicts the consignment of *Historiae* to the commissioner by the author. The motif can be found also in *Orthopasca*.

Figure 3 summarises how the MS can be represented with MIMA and how MS1, MS2 and MS3 entities are shared among the graphs. Each MS (e.g. philologic\_statement individual) is constituted by the respective assertion graph linked to its provenance information (respective provenance graph) and the manuscripts description (included in factual data graph). The pivotal element of the interaction between the assertions is the individual `hf_I_p1r_ill1` (instance of `vir:IC1_Iconographic_Atom`, representing the illustration in *Historiae* I, proem). Considering the disciplines different interests, each assertion is then specialised in its specific domain. For instance, MS3 states the illustration symbolic meaning (i.e. monumentality) to create a bridge between *Historiae* and *Orthopasca*; MS2 defines the illustration conceptual meaning to highlight the monumentality of a textual feature. The interpretation criterion (respectively `iconographicApproach` and `palaeographicApproach` individuals) and interpretation type (respectively `similarityDetectionBetweenArtworks` and `handwritingAnalysis` individuals) are stored in the respective provenance graphs to preserve this kind of peculiarities.

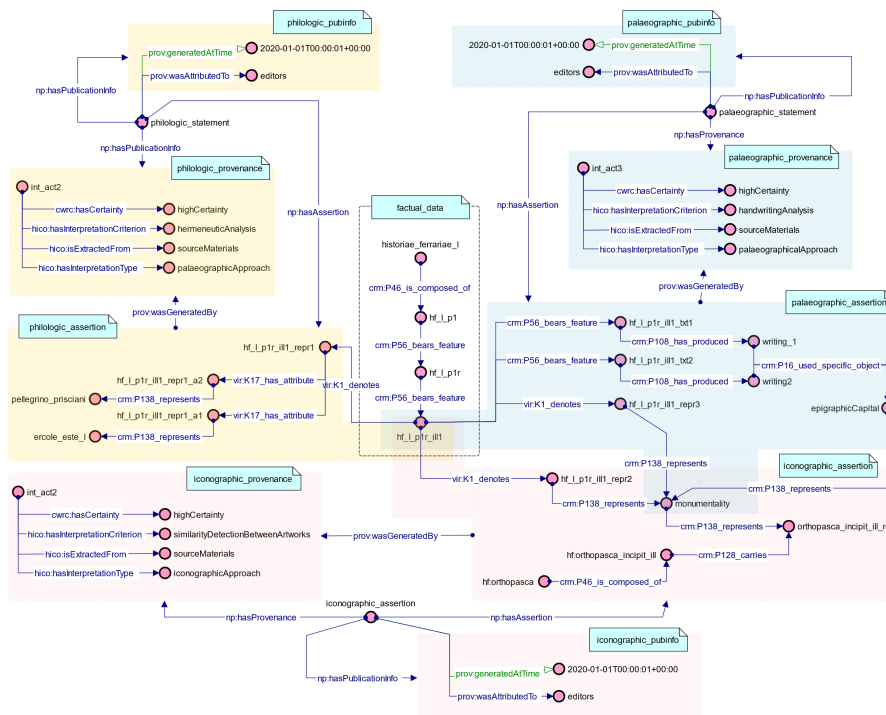


Figure 3: illustrates MIMA applied on MS1, MS2 and MS3.

## 5 CONCLUSIONS

The case study wants to show how MIMA leverages potentialities of the Semantic Web technologies when re-organizing, disassembling and reassembling sources of information and connecting bits into a big network of entities with external and internal linking. MIMA allows different disciplines analysis to coexist in a single environment and to easily compare them over their four layers (e.g. what and how an expert states an interpretation). MIMA opens to layered data analysis (e.g. how disciplines interact between them? Are there shared patterns among them?) and visualisations (e.g. filtering cultural objects over layers). Even considering its verbosity (inherited both from CIDOC and nanopublications), MIMA offers an effective and reusable way to represent such a complex and multifaceted domain.

## REFERENCES

1. Bowen, Glenn A. 'Document Analysis as a Qualitative Research Method'. *Qualitative Research Journal* 9, no. 2 (3 August 2009): 27–40. <https://doi.org/10.3316/QRJ0902027>.
2. Charles, Valentine, and Antoine Isaac. 'Enhancing the Europeana Data Model (EDM)'. *EDM WHITE PAPER*, 2015.
3. Daquino, Marilena, Francesca Giovannetti, and Francesca Tomasi. 'Linked Data per le edizioni scientifiche digitali. Il workflow di pubblicazione dell'edizione semantica del quaderno di appunti di Paolo Bufalini'. *Umanistica Digitale*, 18 December 2019, No 7 (2019). <https://doi.org/10.6092/ISSN.2532-8816/9091>.
4. Daquino, Marilena, Valentina Pasqual, and Francesca Tomasi. 'Knowledge Representation of Digital Hermeneutics of Archival and Literary Sources'. *JLIS: Italian Journal of Library, Archives and Information Science = Rivista Italiana Di Biblioteconomia, Archivistica e Scienza Dell'informazione* : 11, 3, 2020, no. 3 (2020). <https://doi.org/10.4403/jlis.it-12642>.
5. Doerr, Martin. 'Ontologies for Cultural Heritage'. In *Handbook on Ontologies*, edited by Steffen Staab and Rudi Studer, 463–86. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009. [https://doi.org/10.1007/978-3-540-92673-3\\_21](https://doi.org/10.1007/978-3-540-92673-3_21).
6. Groth, Paul, Andrew Gibson, and Jan Velterop. 'The Anatomy of a Nanopublication'. *Information Services & Use* 30, no. 1–2 (21 September 2010): 51–56. <https://doi.org/10.3233/ISU-2010-0613>.
7. Hevner, March, Park, and Ram. 'Design Science in Information Systems Research'. *MIS Quarterly* 28, no. 1 (2004): 75. <https://doi.org/10.2307/25148625>.

8. IFLA Study Group. *Functional Requirements for Bibliographic Records*. Munich: International Federation of Library Associations and Institutions, 2008.
9. Van Den Akker, Chiel, Susan Legêne, Marieke Van Erp, Lora Aroyo, Roxane Segers, Lourens van Der Meij, Jacco Van Ossenbruggen, Guus Schreiber, Bob Wielinga, and Johan Oomen. 'Digital Hermeneutics: Agora and the Online Understanding of Cultural Heritage'. In *Proceedings of the 3rd International Web Science Conference*, 1–7, 2011.

# Testimoniare il Lager: l'informatica al servizio della memoria

Marina Riccucci<sup>1</sup>, Angelo Mario Del Grosso<sup>2</sup>, Frida Valecchi<sup>3</sup>, Giulia Causarano<sup>4</sup>  
Università di Pisa – marina.riccucci [at] unipi.it  
ILC-CNR – angelo.delgrosso [at] ilc.cnr.it  
MIUR – fridavalecchi [at] gmail.com  
Università di Pisa – g.causarano [at] studenti.unipi.it

## SINTESI

Il contributo illustra il lavoro condotto in seno al progetto di ricerca *Voci dall'inferno*: le parole per dirlo, condotto e coordinato da Marina Riccucci (Università di Pisa) con il supporto del Dott. Angelo Mario Del Grosso (Università di Pisa) e della Prof.ssa Frida Valecchi. Nel caso specifico sarà dedicata particolare attenzione alla rappresentazione digitale XML-TEI di una testimonianza manoscritta inedita di una donna sopravvissuta alla deportazione in Lager, nonché allo sviluppo di un componente innovativo per l'elaborazione dei documenti digitali, integrato nell'applicazione web *Memoriarchivio* - software creato specificamente per *Voci dall'inferno* e che rappresenta un fondamentale strumento per archiviare e analizzare le risorse testuali oggetto d'analisi.

## PAROLE CHIAVE

italianistica digitale, codifica del testo, testimonianze, lager, lessico dantesco, XML-TEI, Saxon-js.

## INTERVENTO

### 1 INTRODUZIONE

Le memorie dei sopravvissuti all'Olocausto ci arrivano vive e nitide in due forme o tipologie: sono, rispettivamente, le testimonianze letterarie e le testimonianze non letterarie [5]. Le prime sono opere composte secondo i canoni della narrativa, della letteratura; i testimoni scelgono di riferire l'esperienza vissuta sotto forma di racconto organizzato, tematicamente e stilisticamente strutturato. Le seconde sono resoconti orali, in forma di intervista, e resoconti scritti, in forma di diario, memoriale, epistolario, solo di rado in tangenza con la letterarietà.

La ricerca ruota intorno al progetto *Voci dall'inferno*, che è un progetto ampio e per il quale il supporto dell'informatica è fondamentale. *Voci dall'inferno* è finalizzato al raggiungimento di almeno due obiettivi primari: 1) la costituzione di un corpus digitale delle testimonianze non letterarie dei sopravvissuti ai Lager (diari, lettere, memoriali); 2) la verifica della presenza, all'interno di quelle testimonianze, di lessico dantesco [4].



Il lavoro si svolge dunque su più fronti: quello della creazione di un database che raccolga le testimonianze italiane non letterarie dei sopravvissuti ai Lager nazisti e che le renda interrogabili in merito alla presenza di lessico dantesco (a questo scopo è stato creato il software Memoriarchivio<sup>16</sup>); quello specificamente critico, che fa dei singoli testi oggetti di analisi linguistico-filologica, e quello più marcatamente tecnico-informatico che mira alla rappresentazione digitale delle testimonianze condotta tramite l'uso di linguaggi di marcatura dichiarativi. Nello specifico, oggetto di studio e di indagine è la testimonianza manoscritta inedita di una donna di origini polacche, Romana Feld: le pagine autografe che compongono il testo sono state codificate seguendo le linee guida della TEI [1], quindi incluse all'interno del database Memoriarchivio. Grazie ad una apposita sezione della banca dati, dove sono presenti le anagrafiche dei testimoni e dei curatori, nonché le fonti da cui sono state tratte le testimonianze, è possibile mantenere le tracce sui contesti di provenienza e sulla storia della loro conservazione. Infine il contenuto del documento è stato analizzato sotto un profilo strettamente linguistico.

## 2 METODI E RISULTATI

La prima fase del lavoro ha portato alla selezione di una testimonianza dall'archivio della Fondazione CDEC, ente che ricopre un ruolo fondamentale nel panorama storico-culturale ebraico nazionale e internazionale. Tra le varie testimonianze, la scelta è ricaduta sulla testimonianza di Romana Feld, donna polacca deportata nel campo di concentramento e sterminio di Birkenau nel 1944. Nella seconda fase ci si è concentrati sulla codifica del testo [2] e, in particolare, è stato definito il modello di rappresentazione digitale XML-TEI e la resa grafica del documento stesso (edizione diplomatica). Una volta terminato il processo di codifica della testimonianza si è fatto uso del software open-source EVT [6], necessario a visualizzare l'output della codifica in ambiente web (Fig. 1). In particolare, è stato inserito il documento XML contenente la codifica della testimonianza (e le relative immagini) fra le risorse dell'applicazione. Tale strumento software consente la visualizzazione distinta dei due livelli di edizione (diplomatico e interpretativo) mediante l'uso di tecnologie web e di fogli di stile presenti di default in EVT. Il software quindi ha lo scopo di rendere graficamente all'interno di un browser i fenomeni testuali precedentemente esplicitati attraverso i marcatori XML/TEI nel documento associato. Il collegamento "testo-immagine", che visualizza la corrispondenza tra riga di testo e area dell'immagine, è implementato adottando il metodo di codifica noto con il nome *parallel transcription*.

Le funzioni originarie dell'applicativo Memoriarchivio sono state estese grazie allo sviluppo di un componente client-side basato sull'innovativa libreria JavaScript Saxon-JS 2.0 [2] (Fig. 2, Fig. 3, Fig. 4). Nella terza fase è stata condotta un'analisi critico-interpretativa della testimonianza, volta a determinare e a quantificare l'apporto del lessico dantesco (Fig.5). Il metodo qui introdotto consente di adottare soluzioni di codifica standard e di sviluppare strumenti software aperti che consentano allo studioso di fruire facilmente delle diverse analisi del testo e di poterne estrarre liberamente i contenuti.

---

16 L'applicazione è consultabile all'indirizzo web  
<https://memoriarchivio.org/>

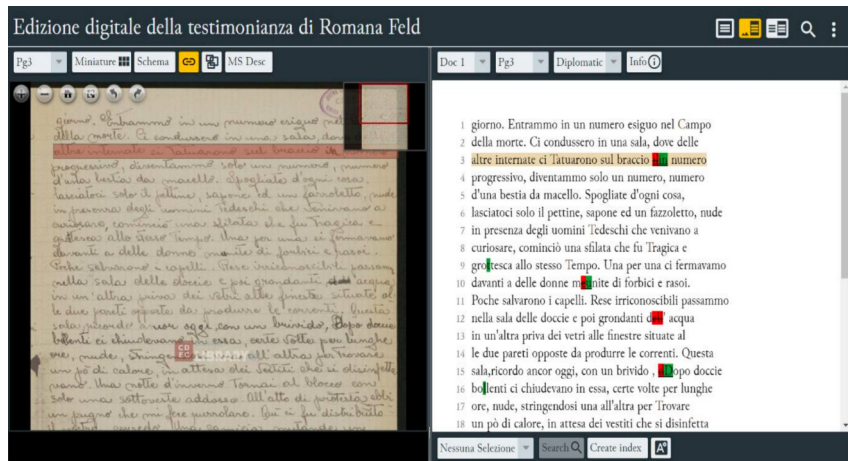


Figura 1 - Visualizzazione dell'edizione diplomatica della testimonianza resa a video grazie al software EVT

```

<!DOCTYPE html>
<html>
<head>
<meta charset="UTF-8" />
<title>SAXON-JS</title>
<script type="text/javascript"
src="Saxon-JS-2.0/SaxonJS2.js"></script>
<script type="text/javascript"
src="js2.js"></script>
<link rel="stylesheet" href="style.css">
</head>
<body>
  <div id="Info"></div>
  <div id="Testo"></div>
</body>
</html>
<xsl:template name="main" match="/">
<html>
<head></head>
<body>
  <xsl:result-document href="#Info">
  <xsl:apply-templates select=
    "//tei:fileDesc"/>
  <xsl:apply-templates select=
    "//tei:profileDesc" />
  </xsl:result-document>
  <xsl:result-document href="#Testo">
  <xsl:apply-templates select=
    "//tei:text"/>
  </xsl:result-document>
</body>
</html> </xsl:template>

```

Figura 2 - Esempio di utilizzo della libreria Saxon-js. Sulla sinistra il documento HTML e sulla destra il documento XSLT

|    |                                |         |         |   |           |  |  |  |
|----|--------------------------------|---------|---------|---|-----------|--|--|--|
| 10 | Auschwitz, un mondo rovesciato | Marta   | Ascoli  | sopravvissuti ai campi di concentramento tedeschi |           |  |  |  |
| 11 | La mia vita ad Auschwitz       | Samuele | Modiano | Sami Modiano: la mia vita ad Auschwitz            | 27-1-2015 |  |  |  |

| File .xml inseriti       |                              |           |  |                              |  |
|--------------------------|------------------------------|-----------|--|------------------------------|--|
| ID del file sorgente XML | Testimonianza relativa       | File name | Corpo  | ID del file destinazione TXT |  |
| 20                       | Testimonianza di Romana Feld | main.xml  | <?xml version="1.0" encoding="UTF-8"?> <!DOCTYPE TEI SYSTEM "myTEI.dtd"> <TEI xmlns="http://www.tei-c.org/ns/1.0"> <teiHeader> | 48                           |  |

Figura 3 - Schermata del software memoriarchivio relativa alla gestione dei file TEI

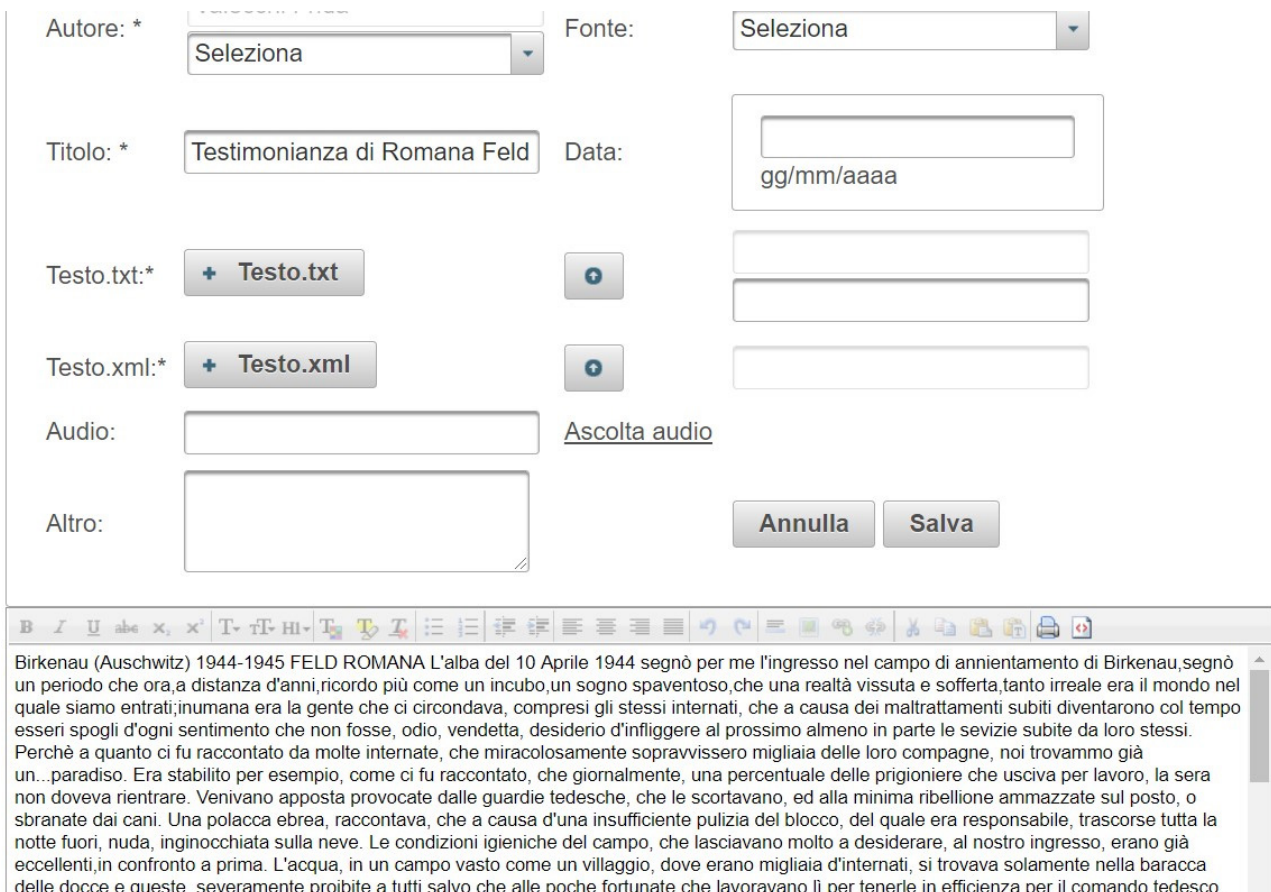


Figura 4 - Schermata del software memoriarchivio relativa all'inserimento di una testimonianza con relativo file TEI xml, se presente

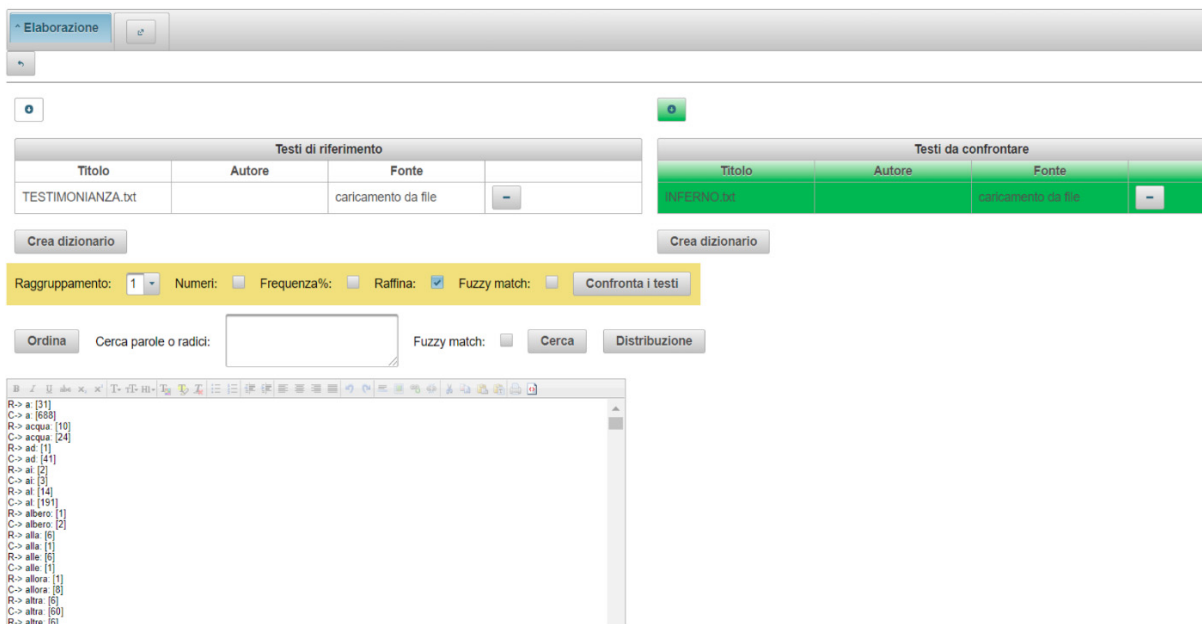


Figura 5 - Schermata del software memoriarchivio relativa al confronto tra il testo dantesco e il vocabolario della testimonianza

### 3 CORPUS DIGITALE

L'applicazione web Memoriarchivio è strutturata in modo da consentire all'utente l'agevole consultazione delle testimonianze, delle fonti da cui sono state tratte, delle anagrafiche dei testimoni e dei curatori delle fonti (Fig. 6, Fig.7). L'accesso con le credenziali di amministratore permette altresì l'inserimento, la modifica e la cancellazione dei dati, compresa la gestione della rappresentazione digitale XML-TEI. In entrambe le modalità di accesso al sito, è comunque garantita la funzionalità di conteggio e ricerca, anche in modalità fuzzy, delle parole nei testi o gruppi di testi in plain text presenti già nel sito o caricati direttamente dall'utente, nonché la funzionalità di confronto fra testi, contando e visualizzando quali parole presenti nel testo di riferimento possano essere ritrovate anche nei testi da confrontare (Fig. 8).

ARCHIVIO DELLA MEMORIA: Testimonianze

| Testimonianze |   |         |                  |  |            |   |       |       |  |
|---------------|---|---------|------------------|--|------------|---|-------|-------|--|
| id            | Titolo  | Nome    | Cognome          | Fonte  | Data       | Altro   | Testo | Audio |  |
| 3             | Il bel viaggio  | Antonio | Falanga          | L'ultimo treno. Racconti del viaggio verso il lager, Roma, Donzelli Editore                      |            | Poesia in dialetto milanese                         |       |       |  |
| 4             | Ricordo di Bruno Piazza                               | Bruno   | Piazza           | Perché gli altri dimenticano. Un italiano ad Auschwitz, Milano (Feltrinelli)                     |            | Ricordo di Bruno Piazza del momento precedente alla |       |       |  |
| 2             | Relazione di una reduce da un campo di concentramento | Lina    | Verona Valabrega | Archivio Vicissitudini dei singoli   | 30-11-1945 |   |       |       |  |
| 5             | Un'infanzia perduta                                   | Liliana | Segre            | Voci dalla Shoa. Testimonianze per non dimenticare, Scandicci(Firenze), La Nuova Italia Editrice |            |   |       |       |  |
| 6             | Oggi vi racconterò l'inferno                          | Nedo    | Fiano            | Voci dalla Shoa. Testimonianze per non dimenticare, Scandicci(Firenze), La Nuova Italia Editrice |            |   |       |       |  |

Figura 6 - Schermata principale del software memoriarchivio dalla cui barra del menu si evincono le principali funzioni

ARCHIVIO DELLA MEMORIA: Fonti

i campi con l'asterisco sono obbligatori

Nome: \*  Anno:

Tipo:  Curatori:

Altro:

| Fonti inserite   |      |                           |   |  |
|--|------|---------------------------|---|--|
| Nome   | Anno | Tipo                      | Altro   |  |
| L'ultimo treno. Racconti del viaggio verso il lager, Roma, Donzelli Editore  | 2012 | libro                     |   |  |
| Perché gli altri dimenticano. Un italiano ad Auschwitz, Milano (Feltrinelli) | 1995 | libro                     |   |  |
| Intervista a Vittorio Cremisi  | 1982 | audio intervista sito-web | <a href="http://www.70resistenza.it/doc.php?id=295">http://www.70resistenza.it/doc.php?id=295</a> |  |

Figura 7 - Schermata del software memoriarchivio per la visualizzazione e gestione delle fonti

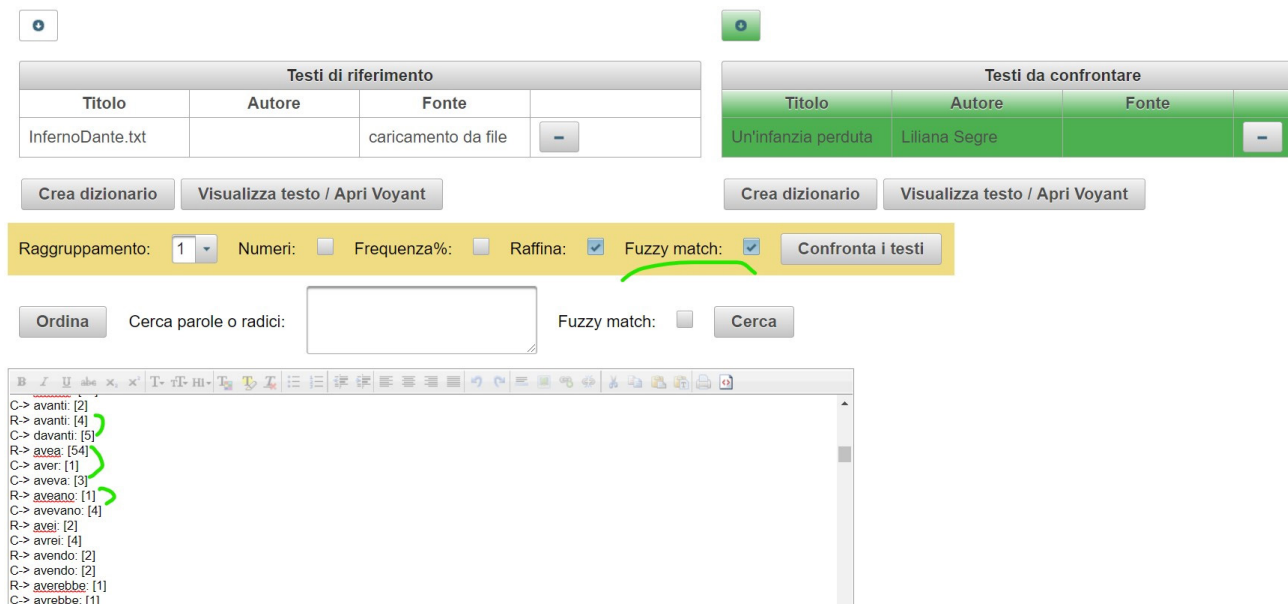


Figura 8 - Schermata del software memoriarchivio per la visualizzazione e gestione delle fonti

## BIBLIOGRAFIA

1. Burnard, Lou, e C. Michael Sperberg-McQueen. «TEI Lite: introduzione alla codifica dei testi». *TEI Lite: introduzione alla codifica dei testi*, 2004. [https://tei-c.org/Vault/P4/Lite/teiu5\\_it.html](https://tei-c.org/Vault/P4/Lite/teiu5_it.html).
2. Ciotti, Fabio. «Cosa è la codifica informatica dei testi». In *Atti del Convegno Umanesimo & Informatica (Trento 1996)*. Trento: Università degli Studi di Trento, 1996.
3. Lockett, Debbie, e Kay Michael. «Saxon-JS: XSLT 3.0 nel browser». *Atti di Balisage: The Markup Conference 2016 17* (2016).
4. Peron, Sabrina. «Dante ad Auschwitz. La poetica di Dante nell'opera di Primo Levi.» *Itinera* 3 (2012).
5. Riccucci, Marina, e Sara Calderini. «L'ineffabilità della nefandezza: Dante 'per dire' il Lager. Un sondaggio preliminare nelle testimonianze non letterarie». *Italianistica* XLIX, n. 1 (2020).
6. Rosselli Del Turco, Roberto, Chiara Di Pietro, e Chiara Martignano. «Progettazione e implementazione di nuove funzionalità per EVT 2: lo stato attuale dello sviluppo». *Umanistica Digitale*, 2019.

# Mu.Vi.A. – Museo Virtuale degli Acquaviva

Carmine Christian Ruocco  
Università di Teramo – cruocco [at] unite.it

## PAROLE CHIAVE

musei virtuali – patrimonio – realtà virtuale

## POSTER

Negli ultimi anni la virtual reality si è configurata come una delle modalità più idonee a dar vita a nuove narrazioni del patrimonio museale tali da permettere un rapporto più intenso tra i visitatori e i musei stessi.

L’immersione totale del VR permette infatti di rivivere non solo il bene museale in sé ma anche il contesto storico in cui esso era originariamente inserito. Fin qui però le esperienze immersive si limitano a riproduzioni digitali di alcuni singoli beni museali lasciando la contestualizzazione del bene stesso all’esterno dell’esperienza VR, mentre, meno frequentemente si presenta una contestualizzazione interna tramite storytelling e, ancora più rare sono le dirette interazioni degli utenti. Nella maggior parte dei casi, infatti, gli utenti non possono muoversi liberamente né le loro interazioni modificano in qualche modo il mondo in cui sono immersi. Al fine di sfruttare appieno le potenzialità dello strumento e renderlo qualcosa in più di una mera curiosità si potrebbero innestare nell’esperienza VR pratiche comuni al mondo dell’entertainment, come la meccanica degli achievement, capaci di catturare maggiormente l’attenzione degli utenti (come ad esempio provato con Toti Submarine VR Experience del Museo Nazionale della Scienza e Tecnologia Leonardo da Vinci).

Andrebbe pertanto tenuto in conto il nucleo realmente innovativo delle nuove tecnologie, evitando di limitarsi a chiedere a società esterne un qualche applicativo tecnologico, con il rischio di delegare anche il processo decisionale sulla “forma”.

In tale filone si colloca il progetto PON Museo Virtuale degli Acquaviva: percorsi turistico-culturali tra Abruzzo, Campania e Puglia, dell’Università degli Studi di Teramo. Tale progetto persegue il tentativo di creare non solo un museo virtuale che offra diversi gradi di interazione – uno di semplice virtual tour e un altro che preveda vere e proprie missioni interattive – ma anche un museo che possa essere interamente progettato e realizzato da chi non è esperto d’informatica, con la sola supervisione tecnica di un’azienda del settore.

Al giorno d’oggi, chiunque può creare e animare i più svariati oggetti e avatar inserendoli in vere e proprie scene interattive, utilizzando software quali Autocad e Unity che, pur presentando una notevole complessità, possono essere facilmente appresi tramite tools messi a disposizione dalle aziende produttrici e reperibili in rete: ad esempio, Unity Learn, diventato gratuito a seguito dell’emergenza pandemica, ha reso alla portata di tutti aspetti fin qui limitati agli specialisti.

In particolare, il Mu.Vi.A. punta ad una ricostruzione interattiva della vita aristocratica nel Regno di Napoli in età moderna, partendo da alcune residenze nobiliari di una delle maggiori casate aristocratiche del tempo, quella degli Acquaviva.

Tale obiettivo è perseguito sia ricostruendo le residenze secondo lo stato del tempo, sia tramite uno storytelling che ha selezionato alcuni eventi occorsi ad esponenti della famiglia permettendo all'utente di riviverli in prima persona. In tal modo sarà, ad esempio, possibile partecipare prendere parte alla riconquista di Atri nel 1508 impersonando lo stesso Andrea Matteo Acquaviva III o incontrare Jacopo Sannazaro divenendone mecenate, o, ancora, giocare una partita a carte esattamente come era solita fare la famiglia nelle giornate invernali a Giulianova. Non si vuole creare una sorta di videogioco sugli Acquaviva, ma si ritiene che un tale approccio possa permettere una conoscenza diretta degli aspetti intangibili del patrimonio umano, quali attività di vita quotidiana, costumi e rituali, specificatamente riferiti al contesto aristocratico del Regno di Napoli. Il progetto, attualmente all'esatta metà della sua durata complessiva di 3 anni, ha attualmente individuato come core quattro residenze principali di cui 2 in Abruzzo (Atri e Giulianova), 1 in Puglia (Marchionne) e 1 in Campania (Napoli).



Figura 1: il castello di Marchionne ricostruito in Unity



Figura 2: la ricostruzione virtuale della biblioteca Acquaviva nella residenza di Giulianova

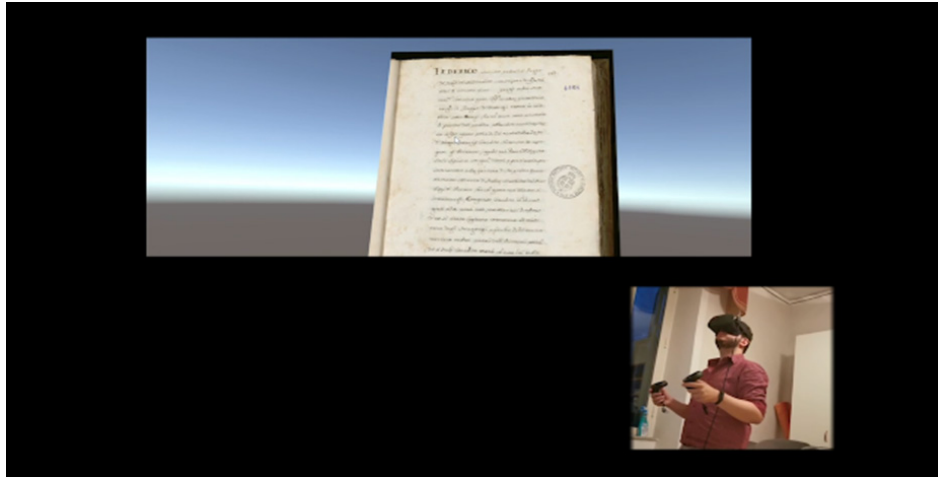


Figura 3: la visualizzazione a 360° di un documento digitalizzato

## BIBLIOGRAFIA

1. Bast, Gerald, Elias G. Carayannis, e David FJ Campbell. *The future of museums*. Springer, 2018.
2. Bonacini, Elisa. «Il museo partecipativo sul web: forme di partecipazione dell'utente alla produzione culturale e alla creazione di valore culturale/The participatory museum on the Web: forms of user participation in cultural production and the creation of cultural value». *Il capitale culturale. Studies on the Value of Cultural Heritage*, n. 5 (2012): 93–125.
3. Izzo, Filomena. *Musei e tecnologie: Valorizzare il passato per costruire il futuro*. CEDAM, 2017.
4. Luigini, Alessandro, e Chiara Panciroli. *Ambienti digitali per l'educazione all'arte e al patrimonio*. FrancoAngeli, 2018.
5. Museo Nazionale Etrusco di Villa Giulia. «Un metodo integrato per valutare i Musei Virtuali e l'esperienza dei visitatori. Il caso del Museo Virtuale della Valle del Tevere». In *Atti del Workshop*, 1–69. Roma: Museo Nazionale Etrusco di Villa Giulia, 2016.
6. Palombini, Augusto. «Musei Virtuali e nuove tecnologie per i musei: esperienze e prospettive europee». In *Il Museo Sensibile*, a cura di Martini Di Berta, 139–53. Milano: Franco Angeli, 2016.
7. ———. «Narrazione e virtualità: possibili prospettive per la comunicazione museale». *Digitalia, Rivista del digitale nei beni culturali* 1 (2012): 9–22.



# Towards the unchaining of symbolism from knowledge graphs: how symbolic relationships can link cultures.

Bruno Sartini<sup>1</sup>, Aldo Gangemi<sup>2</sup>

<sup>1</sup> Università di Bologna – bruno.sartini3 [at] unibo.it

<sup>2</sup> Università di Bologna – aldo.gangemi [at] unibo.it

## ABSTRACT

The aim of the work here proposed is to address the lack of information concerning symbolic meaning in linked open data of the Cultural Heritage domain. Moreover, it is emphasized how this issue limits the interconnections between cultures and cultural heritage items. A review of the current semantic databases and their methods to encode symbolism is presented. Then, an empirical experiment is conducted by describing the symbolism of 15 elements depicted in a CH item using a prototype ontology. The symbolism of those elements has been expanded by including information from a renowned source, and a knowledge base has been created which includes their potential symbolic meaning according to different cultures. This KB has later been matched to a dataset of 3197 paintings, belonging to various genres, extracted from Wikidata. An initial quantitative and qualitative analysis of the results of the matching is presented to demonstrate the potential of a linked data-based semantic representation of symbolism.

## KEYWORDS

symbolism, knowledge graphs, cultural heritage, semantic web

## POSTER

### 1 INTRODUCTION

The definition of a symbol is a highly debated topic in many contexts such as Philosophy, Linguistics, Anthropology as it is emphasized by Eco [3]. In this work, symbols are considered as elements that convey figurative or implicit meanings. These meanings can change depending on the context. The anchor is a “symbol of security and safety” in the Japanese culture [8]. On the other hand, it is used by the German painter Caspar David Friedrich in his painting *The Cross at Rügen* to express “hope of resurrection” [9].

Semantic web-based knowledge graphs have been applied to cultural heritage due to their dynamic nature able to encode the heterogeneous links that exist between CH objects [6]. But, because of the lack of structured symbolic information in these KGs, few connections can be highlighted on a cultural level, therefore limiting the potential offered by linked data with links based only on standard metadata. Additionally, most of the symbolic information of cultural objects are “chained” to unstructured descriptions that cannot exploit the full potentiality of Semantic Web technologies.

## 2 STATE OF THE ART

In this section, the following Semantic databases i) ArCo ii) Europeana iii) Wikidata will be analyzed and compared on their extractable information, especially in the context of symbolism and cultural interconnections.

ArCo<sup>1</sup> is the knowledge graph of Italian Cultural Heritage. Its data is encoded with its own data model which introduces new classes and properties and imports external ontologies [1]. Every CH object is described according to cataloguing standards converted to RDF. In ArCo the property “subject” along with Dublin Core (i.e. dc:subject)<sup>2</sup> is used to link the CH objects with their subjects which are not furtherly linked to any symbolic meaning and neither is the CH object. Nevertheless, in some strings encoded using the Dublin Core property “description”, there are some references to iconographic analyses.<sup>3</sup>

Europeana<sup>4</sup> is a semantic database developed for the diffusion, digital transformation, and preservation of cultural heritage. It uses the Europeana Data Model to encode the data<sup>5</sup>. It shares with ArCo the use of dc:subject and dc:description.

Wikidata<sup>6</sup> is a knowledge base that gathers data from the other branches of Wiki (such as Wikipedia, Wikimedia) converting it to a structured linked data format following its own data model. Differently from ArCo and Europeana, Wikidata has a property named P180 or “depicts” that links a painting to the elements that appear inside it.

The unstructured descriptions provided by ArCo and Europeana limit both the cultural interlinking through symbolism and the information extraction potential. SPARQL queries on those only work by including vague regular expressions (such as “symbol of”) which do not guarantee to retrieve all the relevant data (recall). Moreover, their unstructured form makes them incompatible with logical inferences. The combination of these issues tends to obfuscate (or “chain”) the potential of knowledge graphs to express symbolic content.

---

1 <http://wit.istc.cnr.it/arco>

2 <https://dublincore.org/specifications/dublin-core/dcmi-terms/>

3 An example of this description can be found in

<https://dati.beniculturali.it/lodview-arco/resource/HistoricOrArtisticProperty/0800084295.html>

4 <https://www.europeana.eu/>

5 <https://pro.europeana.eu/page/edm-documentation>

6 <https://www.wikidata.org/>

### 3 EXPERIMENT METHODOLOGY<sup>7</sup>

The painting *Attributes of rDo-rje Kon-btsun De-mo* (fig.1) was chosen as the starting point for the experiment. Its natural language description<sup>8</sup> was manually de-structured (or “unchained”) to get the elements depicted in it (crown, gold, mule, white horse, golden earring, peony, mirror, thunderbolt, arrow, flaming sword, lotus, white robe, necklace, veil, phoenix) and their symbolic meanings. Their symbolism was then enriched using knowledge extracted from Olderr’s Dictionary of Symbols [7], which added potential symbolic meanings of those elements according to *Chinese, Buddhist, Tibetan* contexts.

The data was encoded using the turtle serialization of RDF and following a rapidly prototyped ontology, created for the sake of the experiment. In this model, the symbolic relation holding between a symbol and its meaning is called *Simulation* and is “reified” so that we can talk about it. Simulations are associated with a *Cultural Heritage Item, a Cultural Context*. Appropriate relations (*personification of, emblem of*) are used to formally represent simulations so that they fit the expressive power of RDF and OWL<sup>9</sup>. A symbol (called *Simulacrum*) and its meaning (called *Reality Counterpart*) are linked to their Simulation relation.

The code representing the symbolic knowledge of the painting contains 745 triples. A summary of the process that “unchained” the symbolic meaning from the natural language description can be found in the table 1.

| Part of Painting Description from Dorje (2003)   | Element chosen                | Symbolic meaning (from description)        | Example of Symbolic meaning extracted (from Olderr’s Dictionary) | Turtle encoding of <b>peony-long life symbolic relationship</b> using the ontology   |
|--|-------------------------------|--|--|--|
| The painting shows the attributes of Dorje Kongsun Demo [...] Below the lotus cushion a row of offerings is arrayed:[...], a vase filled with <b>peonies symbolizing long life</b> | Peonies (lemmatized as peony) | long life; Attribute of Dorje Kongsun Demo | CHINESE CONTEXT: love; affection; wealth; honor.                 | <code>symb:peonylonglife a symb:Simulation ; symb:hasContext symb:Tibetan, symb:Chinese, symb:Buddhist ; symb:hasRealityCounterpart symb:longlife ; symb:hasSimulacrum symb:peony .</code> |

Table 1: process of the unchaining of “peony” symbolism from natural language to turtle structured description

7 All the scripts and results of the experiment are available here:

<https://github.com/br0ast/symbolisminKGexperiment>

8 <https://wellcomecollection.org/works/dhewzen3>

9 <https://www.w3.org/OWL/>



Figure 1: detail of Attributes of rDo-rje Kon-btsun De-mo provided by Wellcome Collection

Then, a knowledge base was created by extracting from the dictionary the symbolic meanings associated with these 15 elements belonging to *general*, *Greco-Roman*, *flower language* and *Christian* contexts.

Once the knowledge base was set, 3197 paintings were extracted from Wikidata through its SPARQL portal along with the entities depicted in them (property P180). Resources were chosen from 5 different genres (property P136): Mythological paintings (1000 paintings), Allegory (473), Animal Art (378), Floral Painting (376) and Religious art (1000).

The mapping methodology consisted in string matching between the 15 elements of the knowledge base and the depicted elements of the paintings. For every match, symbolical information was automatically encoded for the painting according to its genre. Table 2 shows how the dictionary contexts were associated to the genres.

|              | Myth. Paintings | Allegory | Animal Art | Religious Art | Floral Painting |
|--------------|-----------------|----------|------------|---------------|-----------------|
| General      | X               | X        | X          | X             | X               |
| Christian    |                 |          |            | X             |                 |
| Greco-Roman  | X               |          |            |               |                 |
| Flower Lang. |                 |          |            |               | X               |

Table 2: mapping between painting genres and the dictionary contexts

## 4 CONCLUSION

In this section, a quantitative analysis on the results of the matching is accompanied by an estimation of possible information that could have been added with a more robust knowledge base. Then, two findings about intercultural linking are presented.

Out of all the 3197 paintings extracted, 288 contain at least one of the 15 elements present in the knowledge base: 72 belong to the Allegory Genre, 53 to Animal Art, 18 to Floral painting, 99 to Mythological Paintings and 50 to Religion. The symbolic relationships matched are 393, described by 7587 triples. By considering the symbolism of all the 2564 unique subjects depicted in the paintings, more than one million triples could be generated.

By comparing the Tibetan painting to the knowledge base, it was found that the Buddhist Goddess Dorje Kongsun Demo shares the attributes of phoenixes, veil and white robe with the Christian personification of Chastity.

Another discovery regards two paintings: *The Magdalen Reading* by Rogier van der Weyden and *Saint George Slaying the Dragon*<sup>10</sup> by Édouard Debat-Ponsan. Looking at their standard Metadata, the only thing that they share is the Wikidata genre of Religious Art. On the symbolic level, after the matching, both express the concept of lust. The literature on these paintings or the characters depicted inside them [4; 5] confirmed their potential symbolic meaning and corresponding connection.

This work presented an experiment on the inclusion of structured symbolic information into linked data of Cultural Heritage. The initial results of the enrichment of a knowledge database on both the quantitative and qualitative sides are promising.

## REFERENCES

1. Carriero, Valentina Anita, Aldo Gangemi, Maria Letizia Mancinelli, Ludovica Marinucci, Andrea Giovanni Nuzzolese, Valentina Presutti, e Chiara Veninata. «ArCo: The Italian cultural heritage knowledge graph». In *International Semantic Web Conference*, 36–52. Springer, 2019.
2. Dorje, G. «A rare series of Tibetan banner». In *Pearls of the Orient: Asian Treasures from the Wellcome Library*. London, 2003.
3. Eco, Umberto. *Semiotica e filosofia del linguaggio*. Einaudi, 1997.
4. Erhardt, Michelle, e Amy Morris. *Mary Magdalene, iconographic studies from the Middle Ages to the Baroque*. Brill, 2012.
5. Fraser, Lorinda. «Saint George and the Dragon: Sainly and Othered Bodies». *The Corvette* 4, n. 1 (2017): 85–97.
6. Lodi, Giorgia, Luigi Asprino, Andrea Giovanni Nuzzolese, Valentina Presutti, Aldo Gangemi, Diego Reforgiato Recupero, Chiara Veninata, e Annarita Orsini. «Semantic web for cultural heritage valorisation». In *Data Analytics in Digital Humanities*, 3–37. Springer, 2017.
7. Olderr, Steven. *Symbolism: a comprehensive dictionary*. McFarland, 2012.
8. Otto, Alexander Francis, e Theodore S Holbrook. *Mythological Japan, or The Symbolisms of Mythology in Relation to Japanese Art*. Philadelphia: Drexel Biddle, 1902.
9. Roberts, Helene E. *Encyclopedia of comparative iconography: themes depicted in works of art*. Routledge, 2013.

---

10 <https://www.wikidata.org/wiki/Q3039856>  
<https://www.wikidata.org/wiki/Q3947220>

# Verso la definizione di un modello di codifica per l'edizione digitale delle postille di Giorgio Bassani

Angela Siciliano<sup>1</sup>, Giovanni Alberto Amato<sup>2</sup>

<sup>1</sup>Università di Pisa, Université Grenoble Alpes – [angela.siciliano \[at\] phd.unipi.it](mailto:angela.siciliano[at]phd.unipi.it)

<sup>2</sup>Università di Pisa – [g.amato10 \[at\] studenti.unipi.it](mailto:g.amato10[at]studenti.unipi.it)

## SINTESI

Il contributo illustra il lavoro di definizione di un metodo per la codifica e lo studio delle postille di Giorgio Bassani.

## PAROLE CHIAVE

filologia digitale, filologia d'autore, TEI-XML, Giorgio Bassani, postille

## POSTER

### 1 INTRODUZIONE

In filologia la scelta dei criteri di edizione di un testo, che consentono di restituirne le caratteristiche, la genesi, la storia dei movimenti interni, non può prescindere da una riflessione di ordine metodologico. Occorre riflettere, ancor prima che sull'oggetto di studio, sulla tipologia testuale che esso rappresenta. Ogni tipologia testuale presenta delle peculiarità che richiedono l'adozione di specifici criteri, ritagliati sul caso, funzionali ad una rappresentazione che intende essere chiara, sistemica ed esaustiva. Questo momento di riflessione è tanto più necessario per una forma di testualità come la postilla: un testo che dipende da un altro testo, che nasce da un altro testo per spiegarlo, illustrarlo, commentarlo in vario modo. La postilla implica quindi – a parte casi eccezionali (note slegate dalla lettura di un libro, ad es. versi o riflessioni appuntate sul colophon o semplici promemoria) – un altro testo: di questo dialogo il filologo dovrà rendere conto nel modo più opportuno. I due testi non hanno infatti lo stesso peso specifico: esiste tra di essi una gerarchia che sarà necessario definire, e che può variare a seconda del caso in esame, e in particolare del tipo di testo con cui la postilla entra in rapporto.

La postilla vergata da un autore sul suo manoscritto/dattiloscritto (Fig. 1) e la nota che lo stesso appunta su un libro a stampa, per esempio, non hanno esattamente lo stesso “valore”.

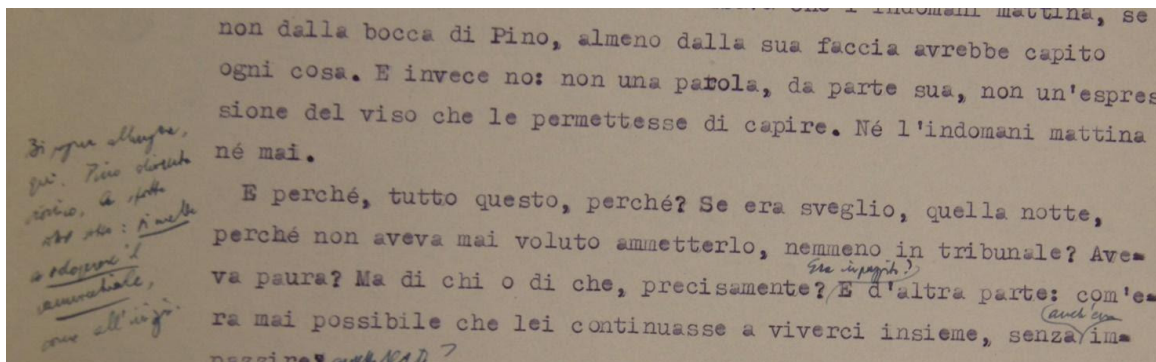


Figura 1: esempio di nota d'autore su dattiloscritto (Giorgio Bassani, Una notte del '43)

## 2 METODOLOGIA E CODIFICA

La postilla in Fig. 1<sup>1</sup> è una nota compositiva con cui Bassani fornisce istruzioni su come procedere nella rielaborazione del finale del racconto *Una notte del '43*. È in posizione di servizio rispetto al testo e nell'edizione critica del racconto sarà ad esso subordinata da un punto di vista grafico e tipografico: collocata nella sezione delle postille, alla fine dell'edizione, e richiamata dal simbolo “<” (segno di parentesi angolare aperta) accanto al testo restituito [9-10].

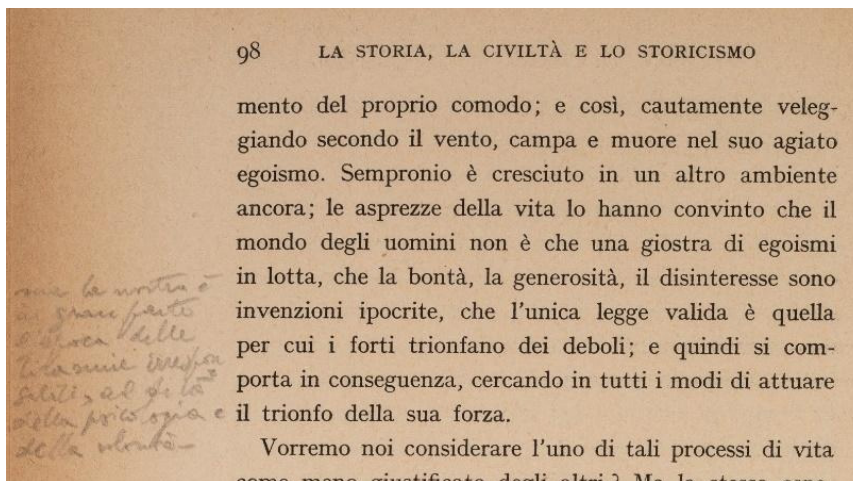


Figura 2: Esempio di nota d'autore su volume a stampa (postilla di Bassani a La scuola dell'uomo di Guido Calogero)

Se, invece, la postilla è una nota d'autore sul margine di un testo a stampa (Fig. 2) il rapporto si rovescia. Dal punto di vista del filologo e dell'editore è la postilla, il testo d'autore, ad essere gerarchicamente superiore al testo a stampa, sebbene nasca in relazione ad esso.

1 Si riporta la trascrizione della postilla: “ Bisogna allungare, | qui, Pino diventa | ironico: la sotte | sotto sotto: si mette | a adoperare il | cannocchiale, | come all'inizio”.

Nell'edizione cartacea questa differenza gerarchica sarà espressa attraverso l'uso di un diverso corpo tipografico, come mostrato in Fig. 3.

41

Sempronio è cresciuto in un altro ambiente ancora; e asprezze della vita lo hanno convinto che il mondo degli uomini non è che una giostra di egoismi in lotta, che la bontà, la generosità, il disinteresse sono invenzioni ipocrite, ♦ che l'unica legge valida è quella per cui i forti trionfano dei deboli; e quindi si comporta in conseguenza, cercando in tutti i modi di attuare il trionfo della sua forza. [p. 98, rr. 3-10] (L)

ma la nostra è | in gran parte | l'epoca delle | tirannie irrespon| sabili, al di là | della psicologia e | della volontà.

### Figura 3: esempio di edizione critica delle postille di Giorgio Bassani

Nell'edizione digitale questa differenza gerarchica riveste un'importanza maggiore, rappresentandone il principio strutturante sul piano della codifica dei dati e della successiva visualizzazione. La codifica delle postille di Bassani è stata eseguita seguendo le linee guida della *Text Encoding Initiative*<sup>2</sup> [2] definendo uno schema ritagliato sul caso specifico. Non esiste infatti un modello, un sistema univoco per codificare e descrivere le postille. Nella sua elasticità, la TEI prevede la possibilità di rappresentare le postille/ i *marginalia*: sul piano dei metadati nel `<msDesc>`, in particolare in `<PhysDesc>` con il tag `<additions>`, e poi nella rappresentazione effettiva del testo in `<text>` con l'elemento `<note>` oppure adottando l'elemento `<add>` [7]<sup>3</sup>, sempre come aggiunta, appendice a un testo principale. Anche se nelle linee guida delle TEI ci si riferisce generalmente a note apposte su manoscritti, questo sistema è stato adottato in casi simili al nostro, per esempio nell'edizione dei marginalia di Whitman (Fig. 4)<sup>4</sup>.

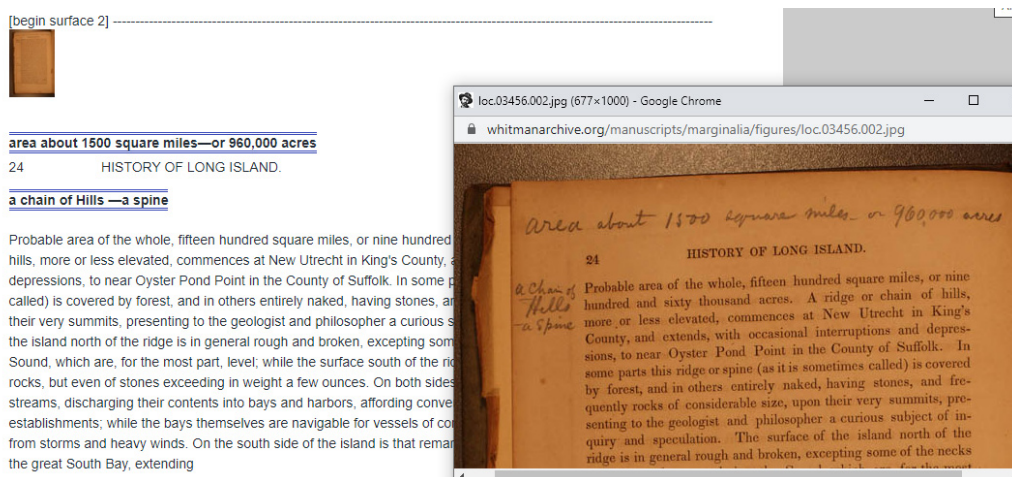


Figura 4: edizione digitale dei marginalia di Whitman

Questa edizione dà maggiore visibilità al testo a stampa, trascritto e riprodotto: come si vede anche dalla collocazione delle postille, fornisce una fotografia, una riproduzione prettamente diplomatica.

2 <https://tei-c.org/guidelines/>

3 Non è l'unica soluzione sperimentata: si ricorre anche a `<label>`, `<milestone>` e altri tag [5].

4 <https://whitmanarchive.org/manuscripts/notebooks/index.html>

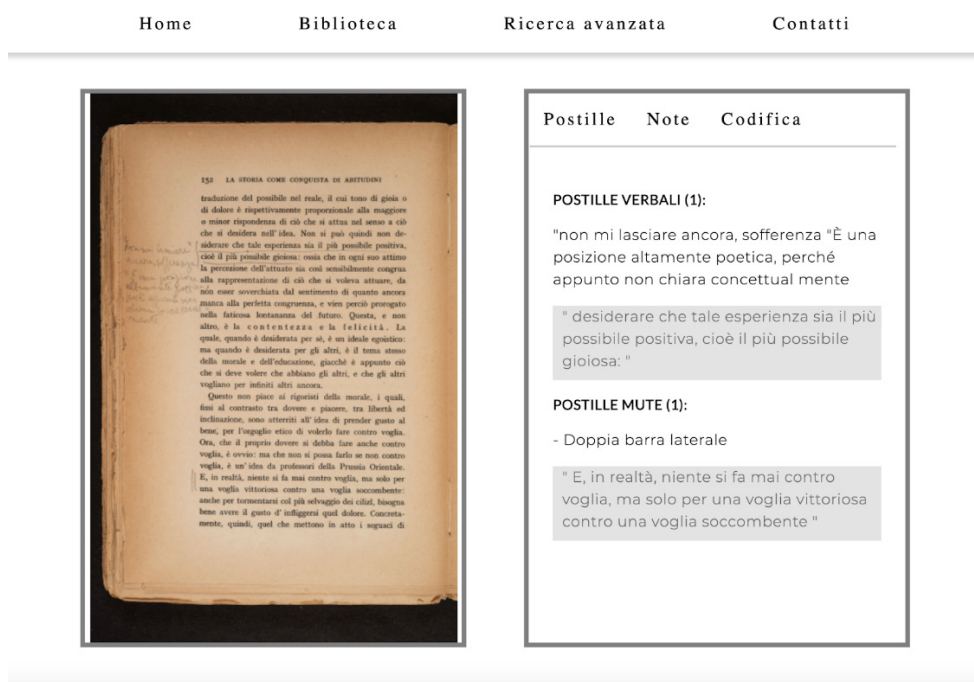


È disponibile anche la riproduzione facsimilare, ma dev'essere aperta in una finestra separata, rendendo difficile il confronto tra l'originale e la trascrizione della postilla. È un sistema poco funzionale.

L'edizione delle postille di Bassani procede in una direzione opposta:

1) dal punto di vista della rappresentazione digitale del testo [3]: annidare la postilla adottando l'elemento <note> oppure l'elemento <add> intercalandolo alla trascrizione del testo a stampa significa riconoscerle una posizione subordinata mentre, nella nostra prospettiva, è la postilla ad essere il testo principale.

2) dal punto di vista della presentazione: l'edizione affianca al facsimile della pagina il testo della postilla, seguito dalla pericope di testo a stampa, distinta ed evidenziata con un fondino grigio (Fig. 5).



3 Figura 5: Prototipo di visualizzazione per l'edizione digitale delle postille di Bassani

In assenza di un modello di riferimento, e sulla base del confronto con altre edizioni<sup>5</sup>, è stata così definita una rappresentazione digitale del testo rispondente alle esigenze editoriali e al caso di studio bassaniano, che potrà diventare un modello di riferimento ed essere adattato anche ad altri autori con esigenze simili.

Una prospettiva di codifica che si basa, come detto in precedenza, sulla priorità gerarchica della postilla sul testo a stampa. Ne consegue una diversa funzione della postilla e quindi un migliore trattamento digitale, sia nella sezione descrittiva dei metadati (<TeiHeader>), che nella sezione dedicata alla trascrizione del testo principale (<text>).

5 Lo studio delle annotazioni che illustri autori (scrittori, filosofi, scienziati) hanno lasciato sulle pagine dei volumi della loro biblioteca suscita un interesse sempre maggiore in progetti di informatica umanistica [1, 6, 7].

A livello dei metadati ciascuna postilla è stata considerata e descritta impiegando il descrittore `<msItem>` del tagset `<msDesc>`, e non con come `<additions>`, cioè aggiunta (Fig. 6):

```
<msItem n="155" xml:id="p155" class="#citazione #barra_laterale_ondulata_e_sottolineatura_interlineare" facs="#pag152.p1">
  <msItem n="155.1" class="#citazione">
    <locus target="#t152.1" facs="#pag152.p1det"> p. 152 </locus>
    <textLang> italiano </textLang>
    <note resp="#AS"> postilla verbale a lapis, sul margine sinistro </note>
  </msItem>
  <msItem n="155.2" class="#barra_laterale_ondulata">
    <locus facs="#pag152.p1det1m1"> p. 152 </locus>
    <note resp="#AS"> postilla muta, costituita da barra laterale ondulata, sul margine sinistro; a lapis </note>
  </msItem>
  <msItem n="155.3" class="#sottolineatura_interlineare">
    <locus facs="#pag152.p1det1m2"> p. 152 </locus>
    <note resp="#AS"> postilla muta, costituita da sottolineatura interlineare; a lapis </note>
  </msItem>
</msItem>
<msItem n="156" xml:id="p156" class="#doppia_barra_laterale">
  <locus target="#t152.2" facs="#pag152.p2det"> p. 152 </locus>
  <note resp="#AS"> postilla muta, costituita da doppia barra laterale, sul margine sinistro; a lapis </note>
</msItem>
```

Figura 6: frammento di codice XML-TEI relativo alla sezione `<msItem>`.

Si evidenziano l'attributo `@target` e l'elemento `<locus>` per registrare le connessioni tra le postille, il facsimile e la loro trascrizione.

È stata poi trascritta nella sezione `<text>`, costituita da più elementi `<div>`, ciascuna corrispondente ad una postilla edita dal curatore. Possiamo infatti pensare la postilla come frammento (quindi parte) di quel complesso e articolato dialogo, o confronto critico, che l'autore istituisce con il libro attraverso l'atto della lettura: una sorta di testo ricostruibile dall'insieme dei frammenti, spesso in relazione tra di loro.

Segue poi un'altra sezione `<text>`, che accoglie il testo a stampa ciascuna postilla è riferita, che si è deciso di trascrivere, in una posizione subordinata – in coda alla codifica –, per utilizzarlo in un secondo tempo per indagini di tipo lessicale e interpretativo. Confrontandolo col testo delle postille sarà possibile ricavare informazioni sul metodo di lettura e postillatura di Bassani: verificare se, per esempio, riprende le parole del testo a stampa, le rielabora, le rovescia semanticamente, le sintetizza, ecc.

Inoltre, la postilla è stata descritta con due prospettive complementari: nella sezione `<sourceDoc>` e in quella `<text>` (Fig. 7):

```
<surface xml:id="pag152">
  <graphic url="pag152.jpg" width="1668px" height="2500px"/>
  <zone xml:id="pag152.p1" points = ' 155,641 161,1055 499,1043 503,781 961,785 967,733 1443,729 1445,655 ' >
    <zone xml:id="pag152.p1det" type="authornote" points = ' ' hand="#m1" corresp="#pag152.p1t">--
    </zone>
    <zone xml:id="pag152.p1det1" points = ' 473,659 481,789 961,783 965,729 1439,725 1435,653 ' >
    <zone xml:id="pag152.p1det1m1" ulx="478" uly="670" lrx="501" lry="793" hand="#m1" corresp="#pag152.p1t">
      <metamark function="highlighting" rend="wavy" place="vertical_line" target="#pag152.p1det1 #pag152.p1det"/>
    </zone>
    <zone xml:id="pag152.p1det1m2" points = ' 499,672 500,778 965,781 966,730 1440,728 1437,665 ' hand="#m1" corresp="#pag152.p1t">
      <metamark function="highlighting" rend="underlining" place="inline" target="#pag152.p1det1m2 #pag152.p1det"/>
    </zone>
    </zone>
    <zone xml:id="pag152.p1t" ulx="496" uly="614" lrx="1443" lry="773" type="printedtext" start="idtext176"/>
  </zone>
  <zone xml:id="pag152.p2" ulx="452" uly="1707" lrx="1442" lry="1830">
    <zone xml:id="pag152.p2det" ulx="453" uly="1701" lrx="493" lry="1805" hand="#m1" corresp="#pag152.p2t">
      <metamark function="highlighting" rend="double_straight" place="vertical_line" target="#pag152.p2"/>
    </zone>
    <zone xml:id="pag152.p2t" ulx="492" uly="1709" lrx="1435" lry="1832" type="printedtext" start="idtext177"/>
  </zone>
</surface>
```

Figura 7: frammento di codifica XML-TEI per la rappresentazione diplomatica delle postille con metodo embedded-transcription

Non si tratta di una ridondanza: la novità della proposta consiste nell’approccio di codifica multidimensionale, che consente di esprimere in modo razionale, chiaro e strutturato il corredo di informazioni relative alla postilla.

In <text> si dà il testo critico della postilla: il testo “pulito”, finale, quale risulta degli interventi di Bassani e dell’editore (limitati a pochi casi sulla base di un approccio di tipo conservativo seguito anche nell’edizione cartacea), che non sono documentati in questa sezione, ma in quella relativa al <sourceDoc>.

Nel <sourceDoc>, in cui con il metodo dell’embedded-transcription si associa la riproduzione facsimilare alla trascrizione della postilla, si fornisce invece un testo diplomatico. È in questo modulo, infatti, che si riproduce e dettaglia la dimensione documentale. Si esplicitano così – in una forma interrogabile – tutti gli elementi che pertengono al farsi del testo e alle sue caratteristiche materiali (ad. es. lo strumento scrittoio, sottolineature, interventi dell’autore). In più, le caratteristiche materiali/documentali si innestano in una struttura reticolare connessa sia ai segni postillatori sia al testo postillato.

## BIBLIOGRAFIA

1. Bleeker, Elli. «Melville’s Marginalia Online – RIDE». *RIDE A Review Journal for Digital Editions and Resources* (blog), 2015. <https://ride.i-d-e.de/issues/issue-3/melville/>.
2. Burnard, Lou, e C. Michael Sperberg-McQueen. «TEI Lite: introduzione alla codifica dei testi». *TEI Lite: introduzione alla codifica dei testi*, 2004. [https://tei-c.org/Vault/P4/Lite/teiu5\\_it.html](https://tei-c.org/Vault/P4/Lite/teiu5_it.html).
3. Ciotti, Fabio. «Cosa è la codifica informatica dei testi». In *Atti del Convegno Umanesimo & Informatica (Trento 1996)*. Trento: Università degli Studi di Trento, 1996.
4. Cohen, Matt, Ed Folsom, e Kenneth M. Price. «Notebooks - The Walt Whitman Archive». *The Walt Whitman Archive*. Consultato 13 gennaio 2021. <https://whitmanarchive.org/manuscripts/notebooks/index.html>.
5. Estill, Laura. «Encoding the Edge: Manuscript Marginalia and the TEI». *Digital Literary Studies* 1, n. 1 (2016).
6. Martinelli, Donatella. «L’edizione digitale delle postille manzoniane a Plauto: problemi ecdotici». *Ecdotica* 14 (2017).
7. Nachwuchsgruppe Berliner Intellektuelle 1800 – 1830 Digital Edition group. *Edition-specific TEI encoding guidelines*, 2016. <https://www.berliner-intellektuelle.eu/encoding-guidelines.pdf>.
8. Pionke, Albert D. «Handwritten Marginalia and Digital Search: The Development and Early Research Results of Mill Marginalia Online». *ILCEA. Revue de l’Institut Des Langues et Cultures d’Europe, Amérique, Afrique, Asie et Australie*, n. 39 (3 marzo 2020). <https://doi.org/10.4000/ilcea.8582>.
9. Siciliano, Angela. «Una notte del ’43” di Giorgio Bassani: edizione e studio critico della versione “originale” - parte I». *Studi di Filologia Italiana* LXXVI (2018).
10. ———. «Una notte del ’43” di Giorgio Bassani: edizione e studio critico della versione “originale” - parte I». *Studi di Filologia Italiana* LXXVII (2019).
11. TEI Consortium. «Guidelines – TEI: Text Encoding Initiative». *TEI - Text Encoding Initiative*, 2020. <https://tei-c.org/guidelines/>.

# Discovering Stories using Visual GISting

Irene Sucameli

Università di Pisa – irene.sucameli [at] phd.unipi.it

## ABSTRACT

Nowadays, journals and newspapers describe events happened all over the world, taking into consideration from the smallest village to the biggest and crowded city of a Country. From these articles it is possible to identify two different narrative levels: the first one is related to the content of the journalistic story and it is the more immediate and accessible level; the second one is more complex and it allows to extract more transversal information on the characteristics of the place itself, going beyond the simple reading of the text to analyse the spatial and cultural factors that led to the event described. To access to this second level, in this paper is proposed an approach based on the Visual GISting method. The aim is to demonstrate how the geo-visualization is able to provide important and substantial contributions to the Humanities, extracting data which are not directly deductible from the literary text.

## KEYWORDS

GIS, journalistic geography, digital public humanities

## POSTER

### 1 INTRODUCTION

Space and narrative have always been intrinsically linked, and they continuously receive reciprocal influences. In fact, the space provides the context for the stories and, at the same time, it is possible to understand its core by knowing the stories which are embodied in that place [2]. Nowadays, the reading of these hidden contents, which are not easily accessible, is simplified by modern geographic information systems (GIS), which help to rediscovery the physical space through a transversal reading which involves the humanities [1; 3; 4]. Indeed, Visual GISting provides an essential contribution in the field of Digital Humanities, and it allows to place a literary phenomenon within a specific space. In the project here described, Visual GIS is applied to Digital Humanities to perform a more in-depth reading of newspaper articles, rediscovering the stories and habits of a Country.

## 2 USING VISUAL GIS

The main question that this paper wants to investigate is whether there is a connection between places mentioned in newspapers' articles and the topics they are nominated for. Furthermore, this research aims to find out if the size of a city is the only element which influences the frequency of appearance of a place or if there are even other components to consider. To answer to these research questions, I selected ten articles per six different topics (*politics, current news, economy, sport, culture, and health & nature*) from the English newspaper "The Guardian", for a total of 60 articles. From these articles, the names of each geographic place together with their frequency of appearance were extracted; then, they were geo-localised within a map of the United Kingdom using QGIS<sup>1</sup>.

Thanks to the visualization of the coordinates of the places mentioned in the articles of The Guardian, it was possible to move from a corpus of journalistic texts to the creation of a spatial database. Using QGIS, it was assigned a different coloured indicator for each topic (**green** for current news, **red** for politics, **orange** for business, **blue** for health & nature, **purple** for sport and **indigo** for culture) and it was visualized on the map the position of those places cited according to the frequency with which they appear in the text; the higher the frequency, the bigger the size of the indicator was (Fig. 1). From this analysis emerged a strong connection between the frequency of appearance of a city in articles and the city's population density. Moreover, as can be easily understood, wider spatial territories, like England, resulted to be more cited than smaller areas, like the Northern Ireland.

Nevertheless, the analysis performed showed that it was necessary to take other elements into consideration since the presence of sports infrastructures, political and cultural centres, natural parks had influence on how frequently the place was mentioned in the newspaper. For example, for the *business* topic, most of the indicators located in England, Scotland and Wales coincided with the presence of shipyards and industrial centres (such as Port Talbot, Redcar, Stockport). Similarly, for the *health* topic a great number of indicators were in the south-east area of England, where there are counties (Surrey, Sussex, Kent and so on) with a great number of coastal localities, wildlife centres and natural parks.

In addition to this first geo-visualization analysis and in order to discover which was the topic most widely spread for each Country, there have been made some spatial queries whose results were graphically represented with convex hulls.

Thanks to this second investigation emerged, for example, that while in England the most spread topic was the *health & nature's* one, in Wales it was the *current news's* one (hull in green in Figure 2).

---

1 <https://www.qgis.org/it/site/>

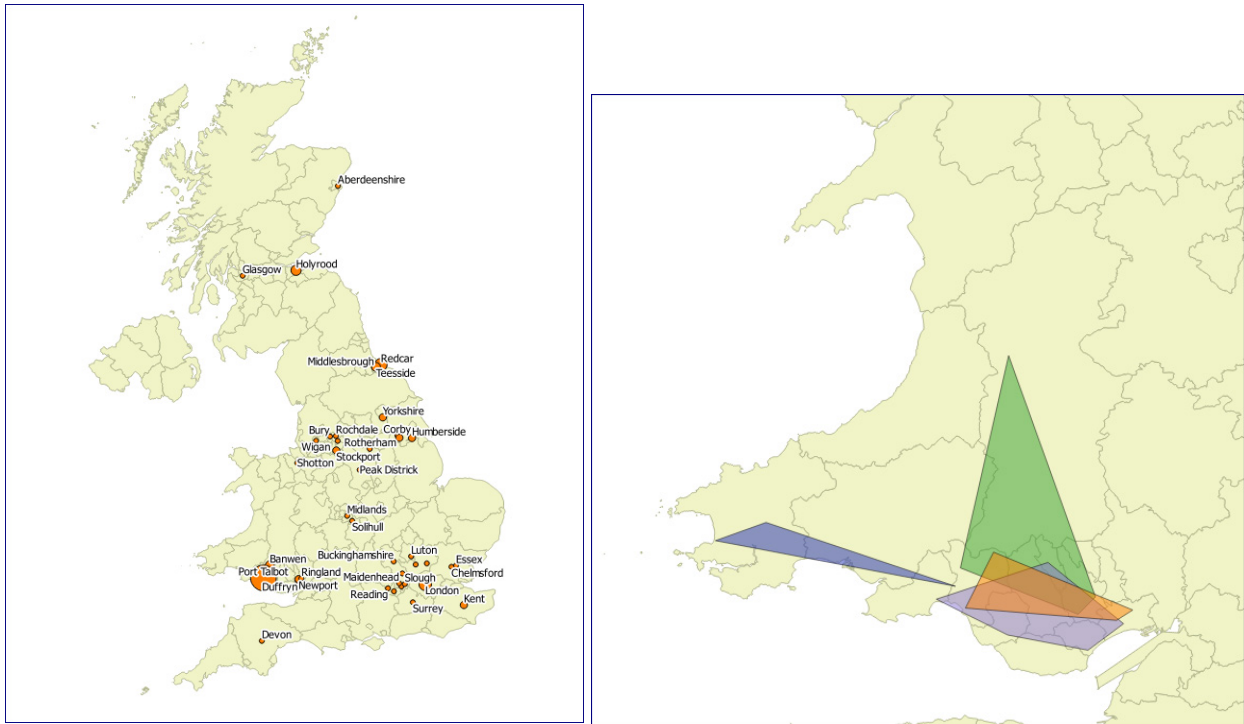


Figure 1: example of geo-visualization of the business articles based on the frequency of appearance of UK cities.

Figure 2: example of geo-visualization of the articles related to Wales, realised using convex hulls

### 3 CONCLUSION

With the present research project, that could be surely implemented and improved in the future, it was demonstrated that there is a strong connection between traditional disciplines and Digital Humanities. Indeed, the adoption of GIS's tools was proved to be extremely useful for digital journalism. The geo-visualization of the articles from The Guardian allowed to shift from a simple question such as “what places is the text discussing” to more complicated questions such as “what is the text saying about these places” [5]. In fact, from the data processed using QGIS emerged pieces of information that could not be deduced from a simple reading of the newspaper's articles.

In conclusion, it was possible to answer the initial questions thanks to a combination of both quantitative and qualitative data and, doing so, it was also demonstrated the enormous potential of geo-visualization systems and how these can be effectively applied to the Humanities.

## REFERENCES

1. Cooper, David, e Ian N. Gregory. «Mapping the English lake district: a literary GIS». *Transactions of the Institute of British Geographers* 36, n. 1 (2011): 89–108.
2. Cresswell, Tim. *Place: an introduction*. John Wiley & Sons, 2014.
3. Earley-Spadoni, Tiffany. «Spatial History, deep mapping and digital storytelling: archaeology's future imagined through an engagement with the Digital Humanities». *Journal of Archaeological Science* 84 (2017): 95–102.
4. Gregory, Ian, Christopher Donaldson, Patricia Murrieta-Flores, e Paul Rayson. «Geoparsing, GIS, and textual analysis: current developments in spatial humanities research». *International Journal of Humanities and Arts Computing* 9, n. 1 (2015): 1–14.
5. Gregory, Ian N., e Andrew Hardie. «Visual GISting: bringing together corpus linguistics and Geographical Information Systems». *Literary and linguistic computing* 26, n. 3 (2011): 297–314.

# Alice's Adventures in Digital Humanities

Mario Valori  
valori [at] tutanota.com

## SINTESI

‘Digital humanities’ è un concetto sfumato e, per certi versi, chimerico, difficile da definire senza porre confini troppo angusti o, al contrario, lasciare eccessivi spazi interpretativi. Se questo è vero per esperti e ricercatore, quale può essere la percezione di ‘digital humanities’ per i non addetti ai lavori? Che significato assume questa espressione in chi non se ne occupa professionalmente e a che ambiti viene ricollegata?

Attraverso un’indagine linguistica qualitativa che ha coinvolto 618 studenti liceali, arriviamo a comprendere come le ‘digital humanities’ siano assimilate nell’immaginario collettivo al concetto e alle pratiche di ‘e-learning’: un risultato che sottolinea la necessità di una riflessione sull’importanza della divulgazione delle digital humanities anche in settori e ambiti non strettamente accademici.

## PAROLE CHIAVE

digital humanities, percezione, definizione, divulgazione

## INTERVENTO

Usando una metafora astronomica, credo che potremmo descrivere il campo delle Digital Humanities come una galassia. In una galassia, infatti, non esistono confini netti e, soprattutto, se vi si sta immersi è quasi impossibile riconoscerne l’articolazione. Ma, se si guarda da fuori (e da lontano), ecco che una morfologia e un confine emerge chiaramente [3].

Con questa poetica quanto vaga espressione Fabio Ciotti, nome ben noto nel panorama delle Digital Humanities, sottolineava efficacemente le difficoltà che possiamo riscontrare nel definire e limitare questo ambito di ricerca, un problema sentito quanto dibattuto dalla comunità scientifica [1; 4; 5; 10]. Meno interesse. Invece, sembra aver suscitato quale sia la percezione delle ‘digital humanities’ in un contesto non professionale: cosa pensa una persona qualsiasi, con un livello medio di istruzione, quando sente parlare di ‘digital humanities’? Quali esempi potrebbero venirgli in mente? Che significato ha per lei questa espressione, che cosa richiama?

Si tratta di domande importanti, ricollegabili non solo alla necessità di ottenere finanziamenti o di attirare i migliori studenti e ricercatori [7; 8] ma anche e soprattutto alla necessità di diffondere cultura, lottando per una scienza aperta, condivisa e davvero inclusiva: una battaglia che non si deve limitare all’open access o a scontri sul diritto d’autore ma che risiede anche, e soprattutto, nella disponibilità a un colloquio che aspiri a coinvolgere l’intera comunità [2; 6].



Un recente studio condotto dalla dottoressa Sciumbata [9] ha portato all'attenzione dell'AIUCD i risultati di una indagine sulle conoscenze digitali di studenti universitari nativi digitali iscritti a corsi in materie umanistiche, mettendo in luce non solo una serie preoccupante di carenze ma anche la propensione nei soggetti coinvolti a sopravvalutare le proprie competenze e abilità, considerandosi quindi adeguatamente formati, se non addirittura esperti.

Come si vedrà dallo studio proposto, una situazione simile (se non addirittura più drammatica) si ritrova nell'ambito delle 'digital humanities'. Da un questionario sottoposto con il supporto di un'associazione culturale a un campione di 618 studenti di varie scuole superiori italiane è emerso che la stragrande maggioranza degli intervistati non solo non era in grado di definire, anche solo in linea generale, cosa potessero essere le 'digital humanities' ma che tale mancanza non era assolutamente percepita dai soggetti coinvolti, fermamente convinti di aver risposto bene alle domande interessate.

L'analisi ha coinvolto studenti nati fra il 1999 e il 2004, equamente suddivisi per genere e iscritti a licei classici, scientifici, linguistici e delle scienze umane. È stato permesso di partecipare solamente a coloro che a un colloquio preliminare avevano dichiarato di essere (almeno potenzialmente) interessati a un percorso accademico e/o lavorativo in ambito umanistico; la scelta, opinabile, è stata fatta per massimizzare i risultati positivi, partendo dall'assunto, idealmente condivisibile, che chi desideri intraprendere una determinata carriera sia più propenso a informarsi sui temi e sugli ambiti di ricerca interessati da questa.

Il questionario è stato somministrato attraverso computer; ogni domanda occupava una pagina ed erano presenti dei cursori che permettevano di muoversi liberamente fra le varie pagine, eventualmente anche tornando indietro a una risposta precedente per integrarla e/o modificarla.

Veniva richiesto nella prima pagina di definire i seguenti sette termini con un limite teorico di 20 parole ciascuno: antropologia, digital humanities, diplomatica, drammaturgia, filologia, iconologia, letterature comparate.

La seconda pagina richiedeva di fornire un esempio concreto di una qualche attività per ciascuno dei lemmi precedenti, fornendo un esempio per rendere più chiaro come rispondere (TERMINE: 'pedagogia' – DEFINIZIONE: 'disciplina che studia problemi e tematiche relative all'educazione e alla formazione umana' – ESEMPIO: 'una ricerca che studi la funzione educativa dei giochi da tavola per i bambini fino dagli 8 ai 12 anni')

Per lo svolgimento della prova erano concessi in tutto venti minuti; quando mancavano circa cinque minuti alla conclusione ciascun gruppo di studenti (i questionari erano somministrati nell'aula informatica, normalmente occupata da due classi alla volta) veniva avvisato che, quale incentivo, si era deciso di offrire un premio pari a € 200,00 in buoni spesa elettronici per colui o colei che avesse risposto meglio alle domande poste. Agli studenti inoltre era stato detto prima della prova che era proibito accedere a internet per cercare le risposte; per questo motivo era stato richiesto di consegnare i propri cellulari. Tuttavia in nessuno dei computer era stato inibito l'accesso alla rete per cui un numero considerevole di studenti (11% prima dell'annuncio riguardante il premio, a cui si è aggiunto un ulteriore 37% dopo la comunicazione) ha deciso di aiutarsi nella compilazione del questionario attraverso il web. Quello però che gli studenti non sapevano era che su ogni macchina era stato installato un software che registrava ogni attività dell'utente, dalle digitazioni fino alle pagine internet visitate o al tempo trascorso davanti a ogni singola schermata.

In questo modo è stato possibile raccogliere non solo le risposte finali inviate dai ragazzi ma anche, attraverso un lungo e paziente lavoro di consultazione dei log e di visualizzazione di filmati, di ricostruire ogni passaggio precedente all'invio delle stesse. Ogni singola risposta è stata così arricchita da una serie di 'metadati' particolarmente significativi, utili per identificare eventuali zone d'ombra e di dubbio (una definizione fornita in pochi secondi denota una maggiore sicurezza rispetto a una riscritta più volte, corretta e di fronte alla quale si è arrabattati per due o tre minuti), nonché per raffrontare le definizioni originali con quelle successive, arricchite dai contributi estrapolati dal web da parte di chi avesse 'barato'.

Per quanto riguarda le 'digital humanities' il quadro, come già preannunciato, appare desolante: oltre l'88% degli studenti coinvolti le considera 'formazione a distanza', definendole idealmente come 'strumenti digitali per formare'. 'E-learning' è l'esempio di attività più comune, presentato da quasi il 30% dei partecipanti, mentre un altro 12% prova a specificare meglio indicando 'Google Classroom'. Preoccupa ancora di più quanto riscontrato controllando i metadati. Il tempo dedicato a definire e esemplificare le 'digital humanities' è pari a 1'41" minore rispetto a quello dedicato a 'diplomazia' 'filologia' 'iconologia' e 'antropologia', mostrando quindi che gli studenti sembrano non avere dubbi sulle loro risposte. Ciò è confermato anche da chi ha usato internet per correggere e integrare le proprie risposte: non solo i 'bari' hanno cercato maggiormente 'filologia' (91%) e 'diplomazia' (75%) rispetto a 'digital humanities' (54%) ma, mentre 'diplomazia' è stata la voce più emendata (la stragrande maggioranza dei partecipanti l'aveva definita in precedenza come la 'scienza delle relazioni internazionali'), 'digital humanities' conta poche correzioni (26%), quasi tutte radicali (eliminazione di quanto scritto in precedenza per copiare e incollare la voce presente su Wikipedia).

Sebbene il campione analizzato non sia statisticamente significativo sorgono comunque spontanee alcune riflessioni che mi sento di condividere. Il risultato appare eclatante: la maggioranza degli studenti liceali interessati all'ambito umanistico non ha la minima idea di cosa siano le 'digital humanities' e tende a relegarle al proprio ambito esperienziale, identificandole come 'e-learning' e 'didattica a distanza'. Questo deriva sicuramente dalla complessità della definizione in sé, già ricordata all'inizio, ma anche da una chiusura del mondo accademico verso l'esterno. Non esistono, ad oggi, siti internet o altri prodotti in lingua italiana che cerchino di spiegare, anche solo per somme linee, le 'digital humanities' ai non addetti ai lavori nonostante il fatto che la loro eterogeneità potrebbero essere, a mio parere, di estremo interesse anche per il vasto pubblico quale connubio della tradizione umanistica con l'informatica, riconoscendo a questa una nuova dignità e sottraendola al ruolo meramente ancillare che il pensiero comune le ha attribuito.

Si tratta di un cammino impervio, sicuramente, ma fondamentale: in un mondo che si affida sempre più passivamente alla tecnologia le 'digital humanities' rappresentano non solo una concreta e pragmatica prospettiva di lavoro ma anche, e soprattutto, un modo per riappropriarsi del proprio retaggio umanistico, attraverso la rielaborazione critica del presente.

## BIBLIOGRAFIA

1. Berry, David M. *Understanding Digital Humanities*. Londra: Palgrave Macmillan, 2012.
2. Caso, Roberto. *La rivoluzione incompiuta. La scienza aperta tra diritto d'autore e proprietà intellettuale*. Milano: Ledizioni, 2020.
3. Ciotti, Fabio. «Oltre la galassia delle Digital Humanities». In *AIUCD 2019 – Book of Abstract*. Firenze: AIUCD, 2019.
4. Gold, Matthew K. *Debates in the Digital Humanities*. Minneapolis: University of Minnesota Press, 2012.
5. Gold, Matthew K., e Lauren F. Klein. *Debates in the Digital Humanities 2016*. Minneapolis: University of Minnesota Press, 2016.
6. Hecker, Susanne. *Citizen Science: Innovation in Open Science, Society and Policy*. London: UCL Press, 2018.
7. Hicks, Diana, e J. Sylvan Katz. «Equity and Excellence in Research Funding». *Minerva* 49, n. 2 (1 giugno 2011): 137–51. <https://doi.org/10.1007/s11024-011-9170-6>.
8. Lok, Corie. «Science Funding: Science for the Masses». *Nature* 465, n. 7297 (2010): 416–18. <https://doi.org/10.1038/465416a>.
9. Sciumbata, Floriana C. «Students of humanities and digital skills: a survey on Italian university students». *Umanistica Digitale*, 2020.
10. Svensson, Patrik. *Big Digital Humanities: Imagining a Meeting Place for the Humanities and the Digital*. Ann Arbor: University of Michigan Press, 2016.

# Musei e digitale durante la pandemia

Chiara Zuanni

Institut Zentrum für Informationsmodellierung – [chiara.zuanni@uni-graz.at](mailto:chiara.zuanni@uni-graz.at)

## SINTESI

Questo poster presenterà i risultati di una ricerca volta ad analizzare l'impatto della pandemia sulla trasformazione digitale dei musei. Il poster si soffermerà in particolare su tre aspetti: 1) la mappatura dei dati concernenti le attività digitali dei musei durante la primavera ed estate del 2020; 2) l'analisi dei dati dai social media; 3) i primi risultati dell'analisi sull'impatto a medio-/lungo-termine volta ad osservare nuove strategie digitali, nuove forme di organizzazione del lavoro digitale, e nuove tecnologie e conoscenze digitali emerse durante la pandemia. Il poster presenta perciò i primi risultati di un lavoro ancora in corso, e si concentrerà su esempi provenienti dal mondo della gestione dei beni culturali in Italia.

## PAROLE CHIAVE

museologia digitale, analisi dei social media, beni culturali, COVID-19

## INTERVENTO

### 1 INTRODUZIONE

A partire dal tardo inverno 2020, con l'avanzare della pandemia COVID-19 e i lockdown in diversi paesi, anche i musei si sono ritrovati chiusi e senza la possibilità di continuare i propri programmi per il pubblico. Secondo un sondaggio ICOM (International Council of Museums) condotto durante la scorsa primavera e che ha ottenuto oltre 1600 risposte da 107 paesi, le attività digitali sono aumentate del 15% nei musei di tutto il mondo, sebbene in modo diseguale<sup>2</sup>. Nei paesi in cui si è assistito ad un incremento delle attività digitali, queste hanno anche catturato l'attenzione della stampa e di diverse piattaforme online, che hanno compilato liste di risorse ed hanno discusso il successo di diverse iniziative digitali lanciate in questo periodo.

---

2 <https://icom.museum/en/news/museums-museum-professionals-and-covid-19-survey-results/>

## 2 IL PROGETTO ‘MAPPING MUSEUMS DIGITAL INITIATIVES DURING THE PANDEMIC’

A partire da marzo 2020, l'autrice ha iniziato a raccogliere dati sulle varie iniziative digitali proposte dai musei durante questo periodo; sulle attività sui social media; e sul dibattito nella comunità museale. Un primo passo verso l'analisi e la successiva presentazione di questo lavoro è costituito da una mappa interattiva che presenta queste iniziative e alcuni dei dati da Twitter. I dati raccolti dall'autrice sono stati compilati in un semplice foglio di calcolo di Google: Google Sheets è stato scelto anche perché lo strumento per collezionare dati da Twitter usato in questa ricerca usa anch'esso un foglio di calcolo Google per archiviare i dati, e perché vi era compatibilità pure con i dati che potevano venire raccolti tramite un questionario di Google (Google Forms).

Al momento attuale (fine novembre 2020), l'autrice ha raccolto circa 700 esempi, anche grazie a risorse rese disponibili da varie associazioni museali (Irish Museum Association; Austrian Museum Association; Museum Computer Network), mentre all'incirca altri 150 esempi sono stati aggiunti direttamente dai musei tramite un questionario su Google (che funge, in questo caso, da base per l'attività di crowdsourcing). In aggiunta, dati da Twitter sono stati raccolti per alcuni degli hashtag più popolari (#MuseumsFromHome; #MuseumsUnlocked; #GettyChallenge; #MuseumWeek e i rispettivi hashtags; etc.): questi dataset al momento contengono oltre un milione di Tweets, di cui soltanto poco più di un centinaio sono stati resi visibili sulla mappa (per evitare sovraffollamento di dati). Un semplice Google Script, che sfrutta l'API di GoogleMaps, ha permesso di geocodificare velocemente gli indirizzi dei musei e le città degli utenti di Twitter (come dichiarate sul loro profilo). Infine, i dati vengono trasformati in GeoJSON ed aggiunti ad una mappa Leaflet, con marcatori diversi a seconda del tipo di attività proposto dai musei (per esempio tour virtuali piuttosto che streaming sui social media; etc.) e con la possibilità di filtrare le varie categorie.

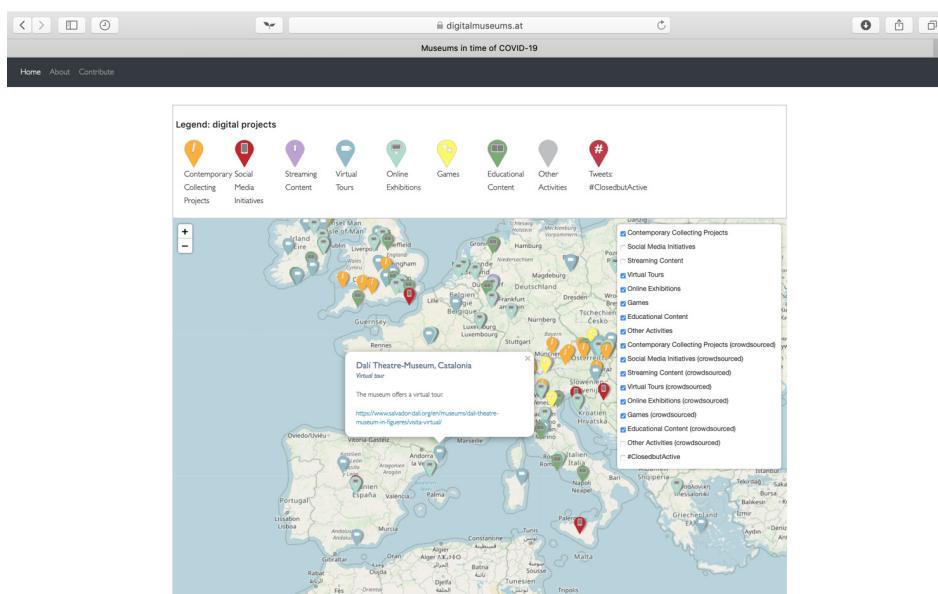


Figura 1 - Mappa di progetti digitali promossi durante la pandemia

### **3 ANALISI E FUTURI SVILUPPI**

Questa mappa rappresenta il primo passaggio verso lo sviluppo di un'analisi più dettagliata delle attività digitali proposte dai musei durante la pandemia. Fino all'autunno 2020, l'attenzione si è concentrata sulla raccolta dati e lo sviluppo di una rete di partner culturali intorno al progetto. Nel frattempo, si è iniziato a trasferire i dati da documenti Google ad un archivio più duraturo, ed in questo caso si sta lavorando con metodi di web archiving e sulla modellazione dei dati (per garantire la possibilità di lavorare coi dati in future ricerche).

In parallelo, si è sviluppata una collaborazione con colleghi nel dipartimento di educazione nell'università di Graz, per osservare l'uso di risorse digitali prodotte dai musei nelle scuole, durante le attività di distance learning. Altre aree della ricerca, ancora in corso, si focalizzano da un lato su questioni di politiche culturali e sviluppo di strategie, teams, e competenze digitali nel settore museale; e dall'altro lato, sulle attività di collezione di memorie della pandemia iniziate da musei, archivi, ed altre istituzioni (e le conseguenze per la gestione di queste raccolte, che includono born-digital data).

### **4 CONCLUSIONE**

Il poster presenterà questo insieme di progetti volti a contribuire alla discussione sulla trasformazione digitale dei musei, allo sviluppo di un archivio legato alla situazione dei beni culturali nella pandemia, e alla ricerca di standard per la conservazione di born-digital data testimoni della vita quotidiana nel 2020.

### **5 RINGRAZIAMENTI**

Un ringraziamento particolare va a tutti i musei e le istituzioni che hanno creduto in questo progetto, contribuendo alla sua promozione ed alla raccolta dati.

### **ULTERIORI RIFERIMENTI**

UNESCO. 'Museums around the World in the Face of COVID-19.', 2020.

<https://unesdoc.unesco.org/ark:/48223/pf0000373530>











ISBN: 9788894253559

Copyright ©2021 AIUCD  
Associazione per l'Informatica Umanistica e la Cultura Digitale



Volume realizzato presso l'Università di Pisa nel mese di gennaio 2021

**aiucd.segreteria [at] aiucd.org**

Il presente volume e tutti i contributi sono rilasciati sotto licenza  
Creative Commons Attribution Share-Alike 4.0 International license ([CC-BY-SA 4.0](https://creativecommons.org/licenses/by-sa/4.0/)).  
Ogni altro diritto rimane in capo ai singoli autori.

This volume and all contributions are released under the  
Creative Commons Attribution Share-Alike 4.0 International license ([CC-BY-SA 4.0](https://creativecommons.org/licenses/by-sa/4.0/)).  
All other rights retained by the legal owners.

