

BIBLIOTECA DI SC. STATISTICHE DEM. E ATTUARIALI  
SERVIZIO BIBLIOTECARIO NAZIONALE

BID PUR 0533009 CID \_\_\_\_\_  
ACQ. 367 / 00 INV. 98704  
COLL. 5 COLL WP CLASS. \_\_\_\_\_  
2000/1

**DATA ACCESS VERSUS PRIVACY: AN  
ANALYTICAL USER'S PERSPECTIVE**

U. Trivellato

2000.1

**Dipartimento di Scienze Statistiche  
Università degli Studi  
Via S. Francesco, 33  
35121 Padova**

Marzo 2000

UNITED STATES DEPARTMENT OF JUSTICE  
FEDERAL BUREAU OF INVESTIGATION  
WASHINGTON, D. C. 20535  
ADVISOR: [illegible]  
DATE: [illegible]

TO: [illegible]

FROM: [illegible]

[illegible]

Director, FBI  
Washington, D. C. 20535

[illegible]

**DATA ACCESS *VERSUS* PRIVACY: AN ANALYTICAL USER'S PERSPECTIVE**

Ugo Trivellato

*Dipartimento di Scienze Statistiche, Università di Padova,  
Via San Francesco, 33 – 35121 Padova, Italy  
E-mail: [trivell@stat.unipd.it](mailto:trivell@stat.unipd.it)*

First draft: May 1999.

This version: March 2000.

# DATA ACCESS *VERSUS* PRIVACY: AN ANALYTICAL USER'S PERSPECTIVE

Ugo Trivellato \*

*Multae utilitates impedirentur si omnia peccata districte prohiberentur*  
*[If all sins were strictly forbidden, then many useful things would be hampered]*

Thomas Aquinas (*S. Th.* II-II, q. 78, a. 1 ad 3)<sup>1</sup>

## 1. INTRODUCTION

The issue of privacy *versus* data access is multifaceted. In fact, it is at the crossroad of two streams of considerations: (i) the traditional concern of official statistical agencies for confidentiality, both for ethical reasons and as a guarantee for maintaining respondents collaboration; (ii) the more recent and broader concern about the legal protection of individuals with regard to the processing of personal data, that extends to all kinds of personal data and of their uses. Besides, intensity and focus of these concerns and ways of handling them vary considerably across countries, as they are rooted in each country's cultural and institutional settings<sup>2</sup>.

---

\* This is a revised and slightly enlarged version of a paper I presented at the CEIES Seminar on "Innovation in provision and production of statistics: the importance of new technologies", Helsinki, 20-21 January 2000, and previously at the ESF European Research Conference "Socio-Economic Research & Geographic Information Systems: Geographic information and analysis in the social sciences", Espinho (Portugal), 22-27 May 1999. I would like to thank Lucia Buzzigoli, Angela Dale, Jean-Pierre Grandjean, Cristina Martelli, Franco Peracchi, Charles Raab and ESF and CEIES seminar participants for comments and suggestions on earlier drafts. Needless to say, the views expressed are my personal ones.

<sup>1</sup> The general theme of question 78 is the sin of usury. In that context, Thomas Aquinas comments upon the quotation as follows: "*Wherefore human law has permitted usury, not that it looks upon usury as harmonising with justice, but lest the advantage of many should be hindered*". Patently, the quotation does not mechanically apply to our case, *i.e.*, the contrast between privacy and data access for research purposes. It merely serves as a loose analogy, to exemplify the trade-off between 'prohibitionist' rules and the many potential benefits from a liberal attitude. As it will be clear from the sequel, in our case the argument should rather be reformulated in the following terms: if the mere possibility of a sin (the potential infringement of privacy, which by the way is not challenged by proper statistical and research activities) were to be strictly forbidden, then many useful things (the advancement of science and its beneficial effects for human life and the society) would be hampered.

<sup>2</sup> See Als (1996) for an informative and penetrating review. I give just two examples, relating to France and the Netherlands. In France, data on individuals and families are entitled to a strong protection, while this is not the case for "*individual information of an economic and financial nature [on firms, which] cannot be used for the purpose of revenue control or for the purpose of economic repression*", but can be used for other purposes (Decree No. 84/628, Article 6: see Buzzigoli, Martelli and Torelli, 1998, pp. 22-27). The option taken by France is not shared by most EU countries: they do not make that distinction or take an even opposite attitude, with higher protection given to data on firms. As for the Netherlands, the vivid sketch provided by Als (1996, p.11), while commenting on concerns for confidentiality and the comparatively poor survey responses of Dutch people,

The debate on the topic is very alive and controversial<sup>3</sup>, especially as a consequence of recent international recommendations or regulations on data protection: the Recommendation No. (97) 18 of the Council of Europe “concerning the protection of personal data collected and processed for statistical purposes” (from now on, simply CE Recommendation)<sup>4</sup>, and the Directive 95/46/CE “on the protection of individuals with regard to the processing of personal data and on the free movement of such data”<sup>5</sup>. EU Member States have to comply with the Directive, bringing into force national laws and secondary legislation. In several countries the process is still going on, in general or specifically with respect to the processing of micro data<sup>6</sup> for statistical and scientific purposes.

I will approach the issue from a distinct perspective, characterised by two aspects: (i) I will take the point of view of an analytical user, *i.e.*, of a person aiming at (collecting and chiefly) processing micro data for statistical/research purposes; (ii) I will focus on some basic features of the issue and on their policy implications, rather than on technical aspects and methodological nuances.

The outline of the paper is as follows. In Section 2, the societal role of statistics and research and their intrinsic needs are considered. Indeed, they are the basis for motivating a credit of confidentiality when personal data are (collected and) processed for statistical/research purposes. Some basic principles and guidelines for regulating this credit of confidentiality, as they are set out in the CE Recommendation and in the EU legislation, are then summarised (Section 3). Section 4 cursorily reviews statistical, technological and legal devices, and related practices, for protecting confidentiality while allowing the processing of micro data for statistical/research purposes. The state of the affairs at the level of the EU is finally discussed, as regards both rules and practices followed by Eurostat and draft legislation (Section 5).

Concluding remarks stress the need for additional efforts to make micro databases widely available to researchers.

---

is illuminating about the role of cultural factors: “*Take a walk the evening in the streets of any small Dutch town. You will be able to observe family life – no shutters, no curtains; yet these same people who literally exhibit themselves will not accept that their sex and their date of birth should form part of an individual identifier! And they reject population censuses.*”

<sup>3</sup> As an example, an issue of “The Economist”, 1<sup>st</sup>-7<sup>th</sup> May 1999, devoted the cover, the leading article and the special section to *The end of privacy*.

<sup>4</sup> See Council of Europe (1997b). This is simply a recommendation to Member States, but it is quite interesting because it deals specifically with data protection for statistical and research purposes.

<sup>5</sup> See European Parliament and Council of the EU (1995).

<sup>6</sup> Some clarification about terminology is perhaps useful (more in the sequel). ‘Micro’ data are unit record data, *i.e.*, data pertaining to ‘individuals’ (persons, firms, *etc.*). ‘Personal’ or ‘confidential’ data refers to unit record data relating to an identified or identifiable individual. ‘Anonymous’ data designates micro data pertaining to a non-identifiable individual.

## 2. STATISTICS AND RESEARCH: THEIR SOCIETAL ROLE AND THEIR INTRINSIC NEEDS

In the context of the legislation on data protection, the trade-off and the balancing is between privacy and the fundamental right to freedom of expression. This right includes explicitly the freedom to receive, and impart information (Article 10 of Convention for the Protection of Human Rights and Fundamental Freedoms, 1950), where the freedom to receive information is considered as implying the freedom to seek information.

The balance is modified when statistics and research come into play, because of their societal function. This function is stated as follows in the CE Recommendation: "*The needs in both the public and private sectors for reliable statistics for analysis and understanding of contemporary society, and for defining policies and strategies for making arrangements in practically all aspects of daily life*"<sup>7</sup>. It is precisely because of this function that specific regulations are set out regarding the protection of personal data collected and processed for statistical and research purposes.

Three points deserve attention. First, it has to be stressed that the distinctive characteristic of a statistical purpose is the collective use of micro data. This means that individuals are the necessary medium for the background information, but that they are not regarded as significant in their own right. In fact, starting with the basic material represented by individual information about many different people, the statistician elaborates results designed to "*characterise a collective phenomenon*" (CE Recommendation, Article 1)<sup>8</sup>. In other words, the statistical result separates the information from the person: personal data are collected and processed with a view of producing consolidated and anonymous information. From this perspective, it is also clear that protecting privacy is in the interest of statisticians, in order to maintain the confidence of respondents and the public and to avoid prejudicing future data supply.

Second, from the standpoint of data protection, (a large part of) scientific research is similar to, and indistinguishable from, statistics<sup>9</sup>. This statement holds both for basic research and for research supporting policies. This point is clearly spelled out in the CE Recommendation (Explanatory memorandum, paragraph 14) with respect to strictly scientific purposes: "*Scientific research uses statistics as one of a variety of means of promoting the advance of knowledge. Indeed, scientific knowledge consists in establishing permanent*

<sup>7</sup> See also Jowell (1981) and Reynolds (1993), among many others. Jowell convincingly argues that "*the individual's 'right to privacy' needs always to be balanced against society's 'right to know'*" and elaborates on this "*important dilemma for research*" (Jowell, 1981, p. 175).

<sup>8</sup> This point was made quite clear already in 1977, by the American Statistical Association *Ad hoc* Committee on Privacy and Confidentiality: "*An administrative record is collected and maintained for the purpose of taking action on or controlling actions of an individual person or other entity. ... Contrastingly, the statistical record has an entirely different purpose. The statistical purpose is to augment general knowledge: to learn the dimensions, trends and relationships of collectives of persons or other entities ... The very essence of statistical analysis is that the identity of individual units ... is immaterial*" (American Statistical Association, 1977, p. 62; emphasis added).

<sup>9</sup> I do not consider here that part of research, especially in the medical and psychological sciences, that involves personalised feedback. In this area, personalised intervention is basic to research (even though statistical analysis may come into play at a later stage), and this calls for specific ethic and legal rules. On that, see Council of Europe (1997a).

*principles, laws of behaviour or patterns of causality which transcend all the individuals to whom they apply. Thus it is aimed at characterising collective phenomena, this being the very definition of statistical results. It could be said, therefore, that research becomes statistical at a certain stage in its development.*" Similar arguments apply to research supporting policies: their design, monitoring and evaluation. In this respect too, the relevant information always relates to collective phenomena and cannot, therefore, under any circumstances entail direct or individualised consequences for individuals.

It is important to add that scientific research asks for an increasing use of micro data. Various reasons act in this direction<sup>10</sup>. To be brief, I will just mention the growing attention paid to individual agents (persons, households, firms), to the heterogeneity among them, to micro-dynamics and interdependencies; and for purposes of assistance to policy and decision-making, the focus on distributive features and on programmes targeted to specific groups of agents<sup>11</sup>.

The third, crucial point has to do with the access to micro databases. 'Free' – *i.e.*, reasonably open and equitable – access to micro data is essential to science, as well as to the functioning of a democratic society. Science is an incremental process that relies on open discussion and on competition among alternative explanations. This holds both for basic research and for research supporting policies<sup>12</sup>. Thus, replication studies are fundamental to science. To this end, it is vital that free access to micro databases will be allowed to qualified researchers who are willing to analyse them.

Arguments about the role of official statistics in a democratic society point to the same direction. The principle of impartiality, one of the "fundamental principles of official statistics" adopted by the UN Statistical Commission as well as of the principles for Community statistics set out in Council regulation 322/97 (the so-called UE "Statistical law")<sup>13</sup>, implies that statistical information should be made accessible to all on a fair basis. Strictly speaking this principle, as well as the other ones, applies to aggregate statistics produced by an official statistical agency, not to micro data. However, it is fully reasonable to suitably extend it to access to micro data for research purposes<sup>14</sup>, while adopting appropriate measures to protect respondent confidentiality.

---

<sup>10</sup> Indeed, the impressive development of micro databases – both on households/individuals and on firms/establishments and more recently on linked employer-employee data, from surveys as well as from administrative sources – is also 'supply' driven, thanks to the enormous advances in data collection, storage and processing allowed by the computer revolution. I will look only on the 'demand' side.

<sup>11</sup> I elaborate more on these points in Trivellato (1999). Various contributions to the subject were presented at the Eurostat-Istat Conference on "Economic and social challenges in the 21<sup>st</sup> century: statistical implications", Bologna, 5-7 February 1996. See particularly Atkinson (1996) and Malinvaud (1997).

<sup>12</sup> Heckman and Singer (1995, p. 93) convincingly argue that "*evaluations build on cumulative knowledge*". See also Rettore and Trivellato (1999).

<sup>13</sup> See UN Statistical Commission (1994) and Council of the EU (1997).

<sup>14</sup> For an extended discussion on the topic, with arguments based on the nature of statistical information as a quasi-public good, the distinction between centralised funding and centralised production, and the theory of bureaucracy and stimulus of competition, see Behringer, Seufert and Wagner (1998).

*Inter alia*, this perspective entails that statistical information collected by official statistical agencies should be treated largely as a public good. Official statistics share some features with public goods, and in addition collective fixed costs have a dominant role in producing it (Malinvaud, 1987, pp. 197-198). However, it is not a public good *per se*: it is patently possible to discriminate among users, both through pricing and through selective, discriminatory access<sup>15</sup>. Characterising official statistical information essentially as a public good is a normative issue: the result of a choice in a democratic society. The main implication for our case is that access to micro data to all *bona fide* researchers – possibly subject to registration and appropriate undertakings, if requested for confidentiality reasons – should be at no cost or at the marginal cost.

### 3. WHERE TO STOP THE PENDULUM? SOME BASIC RULES AND THEIR INTERPRETATION

I was quite long on these matters. But sound perception of guiding principles counts. In my opinion, they persuasively motivate why a *bona fide* analytical user should be given a sort of ‘credit of confidentiality’ in access to micro data.

Clearly, this credit should be reasonable, that is to say sparing in its dimension and regulated with attention to the technical aspects involved, and accompanied by adequate guarantees. Indeed, as long as data are used to produce statistical results and the results themselves are impersonal, there is no threat of infringement of confidentiality. There are, however, risks of disclosure, from disseminated statistical information and when processing the micro data. As a consequence the data might be used for non-statistical purposes, particularly to take decisions or measures in respect to a specific individual. It is in this connection that technical precautions are stipulated and legal guarantees (up to penalties) are laid down.

The picture is rather intricate, because various pieces of provisions intersect and overlap: international and national; pertaining to the general issue of data protection or specifically to the processing of personal data for statistical and research purposes (or just to the processing of such data by official statistical agencies); simple recommendations, laws and secondary legislation. In an effort of simplification, I will focus on some basic rules relevant to our specific topic, chiefly as they emerge from the discipline at the international level – the CE Recommendation on the one hand and the Directive 95/46/CE, complemented by the Council Regulation 322/97, on the other<sup>16</sup>.

<sup>15</sup> A remarkable, fortunately unsuccessful, example of such a perspective is provided by the so-called ‘Rayner doctrine’. Some twenty years ago, Sir Derek Rayner was asked to prepare a report to the UK Prime Minister, consisting of a review of the Government Statistical Services with proposals for restructuring, and retrenching, them. A paramount statement of the report was: “[Statistical] information should not be collected primarily for publication. It should be collected primarily because government needs it for its own business” (Rayner, 1980). For a presentation of the Rayner Report and of the severe criticisms expressed within the Royal Statistical Society, see Hoinville and Smith (1982). The UK Government progressively abandoned the Rayner doctrine, up to its drastic reversal with the 1998 “Green Paper” (HM Government, 1998).

<sup>16</sup> I leave aside broader considerations on the overall impact of the EU data protection directive. In passing, I just note that it gives individuals a property right in information about themselves, thus an unprecedented control over such information. Various concerns have been expressed in this respect: the directive has a “frightening ...



A preliminary clarification has to do with of the very same notion of 'personal (or confidential) data' and the complementary notion of 'anonymous data'. According to the CE Recommendation, "*personal data' means any information relating to an identified or identifiable individual .... An individual shall not be regarded as 'identifiable' if identification requires an unreasonable amount of time and manpower. Where an individual is not identifiable, data are said to be anonymous.*" The point is clear and quite important. Essentially, it means that (i) a reasonably low risk of identification in anonymous data sets is accepted, and that (ii) provisions for data protection do not apply to anonymous data. However, the notion of identifiability is phrased differently across the various pieces of legislation<sup>17</sup>; and what is perhaps more important, the ways of putting these indications into practice rest largely open. So, implementation is crucial.

As for the provisions given for processing personal data for statistical purposes, a crude, but useful distinction can be made between positive principles specifically relevant to the domain and derogations to some general norms of data protection. Among the first category, it is worth mentioning:

- The principle of lawful use of personal data, for statistical purposes only: "*Personal data collected and processed for statistical purposes shall serve only those purposes. They shall not be used to take a decision or measure in respect of the data subject, not to supplement or correct files containing personal data which are processed for non statistical purposes*" (CE Recommendation, principle 4.1).
- A sort of principle of parsimony, which implies that personal data should be rendered anonymous and identification data (*i.e.*, those personal data that allow direct identification of the individual) should be separated from the data used to produce the statistical results as soon as it is reasonable.
- A set of indications about measures to be taken to ensure the security of personal data.
- A firm statement about publication: "*Statistical results shall be published or made accessible to third parties only if measures are taken to ensure that the data subjects are no longer identifiable on the basis of these results, unless dissemination or publication manifestly presents no risk of infringing the privacy of the data subjects*" (CE Recommendation, principle 14.1).

Exemptions to general provisions for data protection, when personal data are processed for statistical purposes, include:

- The possibility of processing data for statistical purposes that were originally collected for non-statistical purposes (chiefly, for administrative uses), with partial derogations to the

---

*bureaucratic orientation, an orientation which consists in regulating everything down to the slightest detail*" (Als, 1996, p. 20); it is doubtful whether the directive can be applied in practice, if too many people try to use it; "*broadly enforced, such a property right would be antithetical to an open society*" ("The Economist", May 1<sup>st</sup>-7<sup>th</sup> 1999, p. 13).

<sup>17</sup> For instance, the formulation of the Council Regulation 322/97 alludes to a somewhat wider notion of identifiability ("*To determine whether a statistical unit is identifiable, account should be taken of all the means that might reasonably be used by a third party to identify*" it: Article 13; emphasis added). On the contrary, the formulation of the UK Data Protection Act 1998 is decidedly more liberal, because it refers to identification "*from those data and other information which is in the possession of, or is likely to come into the possession of, the data controller*" (Section 1 (1); emphasis added).

obligation to inform the persons involved.

- The conservation of personal data, which can be extended “*for longer periods [than is necessary for the purposes for which the data were collected or ... further processed]*” (Directive 95/46/EC, Article 6)<sup>18</sup>.
- Restrictions to the right of access and rectification of any person about personal data concerning him/her.

What are the operational implications of these principles for an analytical user? First, one has to consider that they are formulated in fairly general terms. Besides, the Directive 95/46/CE, the only act binding for Member States<sup>19</sup>, essentially confines the discipline on data processing for “*historical, statistical or research purposes*” to exemptions to general provisions on data protection, to be laid down by national legislation subject to adequate safeguards. Finally, both the CE Recommendation and the Directive 95/46/CE are wisely silent on technical and organisational measures for assuring anonymity, confidentiality and secure processing, thus allowing adaptations to advances in computing and statistics.

The overall consequence is that considerable room is left to different attitudes across the various countries. Harmonisation proceeds, but at a moderate rate. Member States provisions rest quite diversified. National legislation and even more national practices do matter<sup>20</sup>. A look at practices is useful.

#### 4. SAFE DATA, SAFE SETTINGS AND THE WEB: DEVICES AND PRACTICES FOR PRESERVING CONFIDENTIALITY WHILE ALLOWING ACCESS TO MICRO DATA

Strategies for assuring disclosure avoidance while enhancing free data access are often classified under the headings of ‘safe data’ and ‘safe settings’. I will start from that, but I will argue that due consideration should be given to the most up-to-date technology, including the

---

<sup>18</sup> An interesting specification of reasons for the conservation identification data is given in the CE Recommendation, principle 11.1: “*a) for the collection, checking and matching of the data; or b) to ensure the representativeness of the survey; or c) to repeat the survey with the same people*”. The UK Data Protection Act 1998, Section 33 (3), is particularly liberal in this respect: “*Personal data which are processed on for research purposes ... may be kept indefinitely*”.

<sup>19</sup> Member States were asked to adopt the legislative and administrative measures necessary to comply with the Directive within three years from its adoption (Article 32). In some countries, however, the process is still under way.

<sup>20</sup> I base this statement, and part of the considerations of subsequent Section 4, on a cursory survey of national legislation and practices in selected countries. For reviews, see Als (1996) on UE countries; Motohashi (1998) on selected OECD countries, especially on longitudinal micro databases on firms/establishments; Buzzigoli, Martelli and Torelli (1998), who survey Australia, Canada, France, Germany, The Netherlands, US; Bodin (1999b), who examines Sweden, The Netherlands, UK and the US. For the US, see also Duncan, Jabine and deWolf (1993) and Stevens (1998). I directly consulted actual or draft national legislation for Belgium (Draft Law on official statistics, January 2000), Finland (Statistics Act 62/1994, with subsequent amendments up to 1998), France (see CNIS, 1999), Italy (Law No. 675/96 on data protection; Decree No. 281/1999 on provisions for personal data processed for historical, statistical and scientific purposes) and UK (Data Protection Act 1998; see also Lloyd, 1998).

Word Wide Web, and to the overall system adopted to handle data distribution matters<sup>21</sup>.

#### 4.1. Safe data

The production of 'safe' micro-data sets – *i.e.*, data for which factual anonymity is assured – involves a variety of disclosure control measures, applied to the statistical units (sampling and sub-sampling, micro-aggregation, masking, *etc.*) and/or to the variables (variables suppression, aggregation of modalities and bottom-/top-coding, strategies for injecting error, *etc.*).

There is considerable experience in several countries with the production and release of anonymised data sets, for a variety of social domains. The most prominent examples within the EU are perhaps the data sets of some general purpose household panel surveys: the German Socio-Economic Panel (GSOEP), the British Household Panel Survey (BHPS), and the European Community Household Panel (ECHP)<sup>22</sup>.

The release of 'safe' data sets deserves some consideration. Not surprisingly, it varies with the risk of respondent re-identification (and with its obvious counterpart: reduction in the information content of the data set), as well as with the fact that data set does or does not include sensitive information<sup>23</sup>. In some cases, when the risk is taken to be practically nil, the data are distributed as 'public use micro-data files', released to everybody. In several other cases, when the risk is judged to be small (indeed, the data are considered anonymous) but not negligible, they are released with some restrictions, mainly based on role and licensing. These restrictions are of a varying degree: at one extreme, "*universal access ... for all 'bona fide' non-commercial users (subject to registration and standard undertakings of non-abuse), at no [or at the marginal] cost*" (Jenkins, 1999, p. 81), as it is the case for BHPS and GSOEP; at the other extreme quite restricted release procedures, based on examination and approval of a research project by an *ad hoc* board, subject to detailed agreements, expensive.

Overall, my appreciation is that variability in these release policies is moderately related to the risks of disclosure associated to the micro-data sets involved. Rather, it largely reflects differences in attitudes – liberal *versus* restrictive, to be concise – across countries and statistical agencies.

---

<sup>21</sup> The literature on the subject is enormous. A few general references are Eurostat (1996), Willenborg and de Waal (1996), Fienberg and Willenborg (1999). Useful, updated papers were presented at several *ad hoc* Seminars organised by Eurostat: the 3<sup>rd</sup> International Seminar on Statistical Confidentiality, Bled (Slovenia), 2-4 October 1996 (Statistical Office of the Republic of Slovenia and Eurostat, 1996); the Conference on Statistical Data Protection, Lisbon, 25-27 March 1998 (Eurostat, 1999); the Joint UN ECE/Eurostat Work Session on Statistical Data Confidentiality, Thessaloniki, 8-10 May 1999. The papers presented at this last meeting can be found at <http://www.unece.org/stats/documents/1999.03.confidentiality.htm>.

<sup>22</sup> Extensive information on the first two databases can be found at <http://www.diw-berlin.de/soep/e.faltblat.html> and <http://www.iser.essex.ac.uk/bhps> respectively. On the ECHP, see Section 5.

<sup>23</sup> By sensitive data we mean those data called "*special categories of data*" by the Directive 95/46/CE. They basically comprise data referring to the racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership, physical or mental health, and sexual life. However, the definition of what constitutes sensitive data varies somewhat across countries: for a fairly broad definition see, *e.g.*, the UK Data Protection Act 1998, Section 2. Sensitive data deserve of special protection, and their processing is subject to more extensive prohibitions and requirements than is the case with other forms of data.

#### 4.2. *Safe settings*

The polar approach consists of granting permission for external researchers to have access to micro data held within the 'safe setting' of a secure data storage and working area under the control of the official statistical agency (and usually on its premises).

To our purposes, it is interesting because it combines a high level of protection, relying on physical and logical restrictions to data access, with the opportunity for the researcher to work on confidential data. On the other hand, it is quite clear that a safe setting implies selective entry (because of restricted access rules, capacity limitations, heavy administrative monitoring, substantial direct and indirect costs to the user), and might impose constraints to the principle of free access for research purposes.

Well-known examples of safe settings for micro databases on firms/establishments include the Center for Economic Studies (CES) of the US Bureau of the Census and the CeReM at Statistics Netherlands<sup>24</sup>.

It is worth noting that it would be improper to identify safe settings as areas where disclosure avoidance is assured by physical and logical restrictions only. Indeed, ethical and legal guarantees are already used jointly. As an example, a potential outside user at the CES must (i) obtain a special sworn status by taking a legal oath not to disclose confidential data, and (ii) process the data at the designated secure site.

#### 4.3. *The Web*

The scene is changing remarkably, however. New technological developments in communications, computing and statistical software open up radically new opportunities. One point has to be stressed. Developments in information technology do not act simply as a threat to confidentiality. Cryptography, database protection, audit systems for statistical databases, statistical data confidentiality methods and software<sup>25</sup> are experiencing impressive progress. Their combined use allows to design a system with the capability of (i) allowing an analytical user to access remote statistical micro databases, over the Internet, and to process them for research purposes, while (ii) providing secure statistical confidentiality control.

Interesting examples of electronic micro-data access/dissemination already exist. See, among others, the Data Liberation Initiative (DLI) developed by Statistics Canada, the Data Access and Dissemination System (DADS) project of the US Census Bureau<sup>26</sup>, Nomis in the UK<sup>27</sup>. But the potential of this approach has still to be fully explored. Indeed, given its

---

<sup>24</sup> See McGuckin and Pascoe (1998) and Balk (1998) respectively. Note that the CES has already established several Research Data Centers (this is the name given to its safe settings) scattered in the country.

<sup>25</sup> Many papers presented at the Conference on Statistical Data Protection, Lisbon, 25-27 March 1998, dealt with these topics: see Eurostat (1999). See also McClean (1998).

<sup>26</sup> See Statistics Canada (1998), US Census Bureau (1997) and Buzzigoli, Martelli and Torelli (1998), pp. 18-20 and 65-66. As far as I am informed, the micro-data sets distributed through the two initiatives consist only of public use micro-data files.

<sup>27</sup> Nomis is an official labour market online database for the Office of National Statistics, which disseminates geo-statistical information to 800 customers. Properly, it is a Geographical Information System (GIS) database, not a

capabilities and flexibility, it is likely that in the future it will become the dominant approach.

In this context, a 'safe setting' (or a 'safe network', as we should perhaps call it) will simply designate a set of rules allowing the processing of (potentially) confidential data within a secure environment, with no reference to a physical location. Rules will consist of technological devices, logical and statistical procedures as well as of ethical and legal guarantees.

#### 4.4. *The organisation system for micro-data access for research purposes*

An additional, crucial point deserves attention. In data distribution matters for research purposes, the issue is not confined to raw micro-data access/dissemination. Some other features are essential, or useful:

- extensive documentation on the data, with questionnaire, code-books, and other meta-data informing about the data source, data quality, *etc.*;
- information and training, especially when new technologies and software are introduced or new micro-data sets are made available;
- significant involvement and feedback among analytical users and from them to data producers and handlers (via, *e.g.*, user groups, scientific boards of advisors or *ad hoc* institutions). This dialogue has positive, cumulative effects on two directions: (i) it provides "*significant value added to basic data (new compatible and comparable derived variables and data structures), ... with these derived variables deposited and [ ... further] distributed*" along with the basic data (Jenkins, 1999, p. 78), as well as opportunities for exchanges of views about how best to use the data; (ii) it supplies the data producing agency with useful information for improvements in data collection design, production and distribution.

Such a broader organisation system demands a joint effort by the data producing agencies and institutions or associations from the scientific community, often using some *ad hoc* intermediary agency. I do not elaborate more on this point. It has been persuasively illustrated by Jenkins (1999), with reference to the UK experience with the Data Archive: an agency independent from the various data producers that efficiently handles data distribution matters, coupling it with free access to *bona fide* analytical users and with much work on documentation, training, networking, *etc.*<sup>28</sup>.

While one aspect – distribution to users allocated to a specialised agency – is somewhat peculiar to the UK, the basic features of the organisation set up just described characterise several leading experiences in various countries. Among these are the already mentioned Canadian DLI<sup>29</sup>, and the Dutch programme based on an agreement between the Central Bureau

---

micro database. However, it is relevant for our purposes, because the geographical resolution is down to the electoral wards (some 10,000 units), with a rather skew distribution of the resident population, employment and unemployment and potential disclosure. See Blackmore (1998).

<sup>28</sup> Extensive documentation on the UK Data Archive can be obtained from <http://daww.essex.ac.uk>.

<sup>29</sup> In the DLI, the partners to Statistics Canada are essentially scientific associations and associations of university libraries; the technical and administrative support is provided directly by Statistics Canada; data dissemination, via Internet, is mainly to academic institutions, which then make the data available to professors and students for non commercial uses (Statistics Canada, 1998; Buzzigoli, Martelli and Torelli, 1998, pp. 18-20).

of Statistics and WSA/NWO (a specialised agency founded by the Netherlands Organisation for Scientific Research with the specific “aim of making the CBS micro data files available for scientific research at reduced cost” and at standard, simplified undertakings about confidentiality<sup>30</sup>).

Somewhat different organisation frameworks and policies are needed for the case when confidential micro databases kept within the official statistical agency. In this respect, the US experience is illuminating. A distinct feature of that experience is precisely the long standing practice of financing on-site research on micro-data files by external senior scientists. The American Statistical Association/National Science Foundation/Bureau of Labor Statistics/Bureau of the Census Fellowship Program is perhaps the best example of a systematic effort of supporting research on micro data, in order to “bridge the gap between government and academic social science” – as the 1999-2000 programme states. Not everything is entirely satisfactory with the programme<sup>31</sup>. However, its beneficial effects both on data collection methodologies – from concept development to information dissemination – and on substantive research areas are quite apparent. As far as I know, there is no similar programme within the EU or its member countries.

Summing up, some of the best practices in free access to micro data for statistical purposes result from a combination of various ingredients: (i) a reasonably liberal, flexible legislation, relying more on self-responsibility and codes of conduct than on extensive administrative monitoring; (ii) a sound organisational framework for handling data distribution matters; (iii) an advanced system of secure data access, based on the Web; (iv) as for the case of confidential micro databases kept within the official statistical agency, financial support to on-site research by external scientists.

There are also important strategic consequences from these good practices. External analytical users become resolute supporters of the surveys (and possibly of other, register-based data collection modes) feeding the micro databases, and an important constituency who will advocate continuation and extension funding to them.

My perception is that, among the EU countries, the most favourable mixture of these ingredients can be found in the UK<sup>32</sup>. It is also interesting to consider the case of the

---

<sup>30</sup> WSA/NWO plays an intermediary role, while the release of micro data is handled directly by CBS, by means of a “model contract for external research institutes concerning the multiple use of microdata sets” (quotations are from this contract). The research institution then asks the individual researcher to sign a confidentiality undertaking. See <http://129.125.158.28/wsahomuk.html>.

<sup>31</sup> Some inconveniences, largely arising from the condition for the Research Fellow to conduct his/her research on the premises of the relevant statistical agency, were pointed out by Duncan and Pearson (1991). Recent attempts to overcome part of these inconveniences essentially consisted of making the duration of the fellowship appointment flexible (from 6 to 12 months, with additional opportunities for extensions to appointments, split into two separate terms, and part-time fellowships).

<sup>32</sup> See Jenkins (1999) for an illustration of its positive results with reference to micro-data sets from large-scale household surveys. It is also interesting to mention how the issue of confidentiality is addressed within the online GIS database Nomis (see previously footnote 27). Essentially, it goes through processes of: “1. Formal licensing of users for sensitive data series. 2. Affirmative statements about confidentiality on all outputs where data owners agree the statements. 3. Automating confidentiality rules so that confidential items are clearly flagged. That importantly allows researchers to analyse the data but not to publish or to pass on those cells. 4. Maintaining a full audit trail of all extractions....5. Developing a mutually beneficial partnership between data owners and

Netherlands. It exemplifies how considerable efforts on points (ii) and (iii) can partly compensate a rather strict legislation on (and a strong public concern about) confidentiality. Countries relying heavily on detailed legislative provisions and administrative monitoring tend to lag behind. In some of them however, e.g., in France and Italy, the legislative process is still on the way and the issue of access to micro data for statistical and research purposes is presently given due consideration.

## 5. THE STATE OF THE ART WITHIN THE EU

I come now to the state of affairs with the EU and Eurostat. Council Regulation 322/97 has basically two provisions relevant to micro data access and confidentiality.

- (a) Article 17 states that access to confidential data transmitted to Eurostat may be granted by Eurostat itself, provided that it is for scientific purposes and that explicit approval for such access has been given from the Member State which supplied the data.
- (b) Article 20, while establishing compulsory assistance by the Committee on Statistics Confidentiality<sup>33</sup>, affirms that the Commission should adopt measures “*designed to assure that all the national authorities and [... Eurostat] apply the same principles and minimum standards for avoiding disclosure of confidentiality*”.

On the one hand, the Commission and Eurostat are asked to play an active role for the harmonisation of confidentiality rules and “*minimum standards*”. On the other hand, for the release of harmonised anonymous micro-data sets at the EU level Eurostat has to comply with the approval of all the Member States – each one for its own national data set. This implies that Eurostat might be induced by the harmonisation criterion to adopt the standards set up by the stricter State, that is to say to embrace what could rather be called ‘maximum standards’. I see an ambiguity, a sort of vicious circle here. It might hamper a positive action by Eurostat in promoting and implementing open-minded harmonised rules about confidentiality. A clear, firm commitment to a policy of liberal access to the scientific community is needed to overcome this risk. This was not the case, at least until recently.

As for the actual data access policy by Eurostat (or, in any case, the policy in operation up to the end of 1999<sup>34</sup>), it is still quite restrictive. As far as I am informed:

- In general, access to micro data is allowed only at the safe setting of the Eurostat secure system, mostly for research activities in the interest of the Commission, by Eurostat’s personnel or by consultants contracted by Eurostat.
- Dissemination of anonymised micro databases is confined to just one survey, the ECHP (admittedly one of the major advances in the European statistical system).
- The rules for data dissemination rest on research contracts, stipulating strict conditions for data access and use, and on fairly high prices (Marlier, 1999a).

---

users” (Blackmore, 1998, p. 2).

<sup>33</sup> The Committee was set up by Article 7 of Euratom/EEC Council Regulation 1588/90 on the transmission of data subject to statistical confidentiality to the Statistical Office of the European Communities.

<sup>34</sup> Appreciable changes in the price setting policy for the ECHP user’s database have been introduced, starting from 1 January 2000. I will comment on these innovations shortly.

This distribution strategy has been seriously criticised by Jenkins (1999), who vividly contrasts it to the more open and fruitful practices followed in the UK system. I largely share his views. I will just add two comments. First, the data dissemination rules can hardly be justified only on confidentiality grounds (one should keep in mind that we are dealing with anonymised data sets, to which data protection legislation does not apply). Second, the price system has some unpleasant features: the data are quite expensive; the very articulated set of prices for various types of users resembles more a protectionist tariff system than a mean for recovering the marginal costs. A propensity for restricted access and extensive administrative control still appears to be at work, with likely adverse consequences for science and policy advice<sup>35</sup>.

It is fortunate that recently important steps were taken by Eurostat for reconsidering practices and legislation. As far as I know, the topic entered the agenda of the Statistical Programme Committee in March 1999, under the stimulus of the need to access micro, possibly confidential data by research teams selected within the Targeted Socio-Economic Research Program (TSER)<sup>36</sup>. A Task Force on Access to Confidential Data for Research Purposes was set up, and a broad, sound approach was adopted. It is well illustrated by one of the key principles developed to guide the work: *"The key role of statistical institutes and authorities is to release the maximum amount of information without breaching the confidentiality of the individual respondent. Data have been collected at public expenses, often entailing considerable response burdens. It is the duty of official statisticians to ensure that the best possible (safe) use is made of the data"*<sup>37</sup>.

Notable developments already came about, in two directions: innovations in dissemination practices and the drafting of a new regulation.

As for dissemination practices, the price setting policy for ECHP micro-data sets has been revised, starting from 1 January 2000 (Marlier, 1999b). The good news is that prices for the ECHP users' data base have been reduced significantly, roughly by 50%. Unfortunately, however, the segmented set of prices for various types of analytical users has not been reconsidered.

A promising discussion began also about the possibility of disseminating an anonymised micro database from the Labour Force Survey (Franco, 1999)<sup>38</sup>.

<sup>35</sup> Patently, this is a conjecture. But it is also a legitimate concern. Evidence about the effects of legislation on data protection upon research is scarce. (Indeed, the legislation is recent, and its effects critically depend on the ways it is implemented and enforced.). I am aware of just one study in the medical field, carried out in the US, aimed at determining the effects of state legislation requiring patient informed consent prior to medical record abstraction by external researchers. Its conclusions are the following: *"Legislation requiring patient informed consent to gain access to medical records for a specific study was associated with low participation and increased time to complete the observational study. Efforts to protect patient privacy may come into conflict with the ability to produce timely and valid research to safeguard and improve public health"* (McCarthy et al., 1999, p. 417).

<sup>36</sup> As it was noted, it is paradoxical that research teams selected by EU authorities for carrying out TSER in the interest of the EU encountered tremendous difficulties in getting authorised access to the relevant data sets.

<sup>37</sup> Task Force on Access to Confidential Data for Research Purposes (1999, Explanatory notes), p. 1.

<sup>38</sup> The database comprises micro-data from the Labour Force Surveys conducted in all the Member States of the EU and EFTA, as well as in a number of applicant countries.



The most significant piece of work, however, consists of the drafting of a Commission Regulation on "access to confidential data for scientific purposes"<sup>39</sup>. Moving from some key principles, among which the one just quoted, the Draft Regulation establishes some important rules:

- (a) data access will be granted via two approaches: (a1) release of 'safe', anonymised micro-data, and (a2) on-site access to confidential data held within a 'safe setting' on Eurostat premises<sup>40</sup>;
- (b) a substantial set of surveys or statistical data sources is listed, which can be accessed under one or another approach<sup>41</sup>;
- (c) fairly restricted licensing procedures are maintained for both approaches;
- (d) a provision is made about prices, which vaguely echoes the criterion of marginal costs<sup>42</sup>, but stipulates also that "*they should not lead to unfair competition with the national authorities*".

First, deserved credit should be given to Eurostat for the work done and the more open attitude taken on the issue. But the drafting process offers also opportunities for an informed debate, and hopefully for improvements. My opinion is that these opportunities should be taken at best, and should involve the wider scientific community.

As a commencement of such a debate, I offer a few comments. They move from two basic considerations. It would be convenient to adopt a medium-term perspective, flexible enough so it will leave room for developments in the procedures that could be implemented. Besides, strategies for preserving confidentiality should rely significantly on self-responsibility and codes of conduct<sup>43</sup>. Compared with these standpoints, the route taken by the Draft Resolution is open to the risk of being too narrow and short-term oriented. In other words, the conceived solutions are likely to be rigid and might become rapidly obsolete. Some specific remarks and suggestions follow.

---

<sup>39</sup> See Task Force on Access to Confidential Data for Research Purposes (1999).

<sup>40</sup> 'Safe settings' can also be established on the premises of Member State national statistical institutes, under appropriate conditions.

<sup>41</sup> Note that in the case of on-site access to confidential data, data on firms/establishments are subject to stricter conditions than those relating to households and natural persons.

<sup>42</sup> "*Costs related to the use of the Commission facilities and to the data accessed shall be borne by the researcher*" (Article 9). To me, the notion of "*costs related to data accessed*" is essentially indeterminate.

<sup>43</sup> The Directive 95/94/EC explicitly deals with codes of conduct (Article 27). On the contrary, both the Council Regulation 322/97 and the Draft Regulation ignore them. Good examples of general purpose codes of conduct for statisticians are: the Declaration of Professional Ethics, adopted by the International Statistical Institute in 1985 (International Statistical Institute, 1986); the *Code de déontologie statistique*, adopted in 1984 by the French Association des Administrateurs de l'INSEE; the Code of Practice for Official Statistics, approved in the UK in 1996 (Bodin, 1999a); the Ethical Guidelines for Statistical Practice, recently approved by the Board of Directors of the American Statistical Association (American Statistical Association, 1999). More focused on confidentiality issues are various rules of conduct set up within the UK official statistical system, chiefly the GSS Code of Practice on the Handling of Data Obtained from Statistical Inquiries, issued in 1991, and the statement from the Office for National Statistics on Maintaining the Confidentiality of Data, made on 1 April 1996 (see Office for National Statistics, 1996, and Bodin, 1999b, pp. 5-10).

- Attention should be given to more liberal licensing procedures. Basically, free access (subject to registration, standard undertakings and codes of conduct) should be granted to anonymised data sets. Legislation on data protection is not an obstacle in this respect. And I do not see any convincing reasons for ruling out *a priori* this possibility.
- The Draft Resolution states precisely the approaches that can be used for granting data access: safe data with restricted licensing and safe settings. Why to be so circumstantial? There are two potential shortcomings in that. Rules about the release of anonymised micro-data would tend to be unduly uniform (*inter alia*, they exclude the release of public use micro-data files). And what is more, the detailed provision of two approaches precludes developments in the most promising direction – secure data access based on the Web.
- Along the same vein, I have some doubts on the listing of the databases, for which safe release or on-site access is granted, by their name. Clearly, it would be better to specify just the types of surveys and the categories of data sources, from which micro-data sets could be obtained.
- The document is silent about the organisational framework for handling data distribution matters. It will be wise to introduce some guidelines about how to develop a beneficial partnership between Eurostat and the wider scientific community.
- A clarification on the price setting policy, definitely oriented to the criterion of marginal costs and to uniform prices for analytical users<sup>44</sup>, is highly needed.

## 6. TO CONCLUDE

A further, co-operative effort is needed among official statisticians and the wider scientific community. It has to be addressed to the preparatory work for the drafting of liberal laws and regulations, as well as to the design and the implementation of sound organisation systems for micro-data access for research purposes.

But it has also to be targeted at more general aims, relating to the perception of these issues by the public. We need to broaden and deepen public understanding of the societal role of statistics and research: for the society's well-being as well as for its democratic life. We need also to extend public understanding of our technical capability and ethical responsibility to maintain confidentiality.

UGO TRIVELLATO

*Dipartimento di Scienze Statistiche,  
Università di Padova*

---

<sup>44</sup> With respect to costs and prices, I find the indication to avoid “*unfair competition with national authorities*”, given by the Draft Resolution, not quite appropriate. Indeed, a national statistical agency has some sort of monopolistic control over the information it produces, then some discretion on setting prices. The risks to be avoided are likely to be rather different from – or at least more diversified than – a threat to competition.

## REFERENCES

- G. ALS (1996), *Statistical confidentiality in the 15 Member States of the European Union: a critical comparison*, in Statistical Office of the Republic of Slovenia and Eurostat, "Third International Seminar on statistical confidentiality, Bled - Slovenia, 2-4 October 1996. Collection of papers", Ljubljana, Statistical Office of the Republic of Slovenia, pp. 9-29 (mimeo).
- AMERICAN STATISTICAL ASSOCIATION (1977), *Report of the Ad hoc Committee on privacy and confidentiality*, "The American Statistician", 31 (2), pp. 59-78.
- AMERICAN STATISTICAL ASSOCIATION (1999), *ASA issues: Ethical guidelines*, prepared by the Committee on Professional Ethics and approved by the Board of Directors, "Amstat News", No. 269, November 1999, pp. 9-15.
- A.B. ATKINSON (1996), *Social and economic change: implications for statistics in the 21<sup>st</sup> century*, invited paper at the Eurostat-Istat Conference on 'Economic and social challenges in the 21<sup>st</sup> century: statistical implications', Bologna, 5-7 February 1996 (mimeo.).
- B.M. BALK (1998), *Establishing a Center for research of economic micro data at Statistics Netherlands*, in "Proceedings of the International Symposium on linked employer-employee data", 21-22 May 1988, Arlington (VA), U.S. Bureau of the Census (CD-ROM).
- J.-L. BODIN (1999a), *Etat des réflexions sur les principes fondamentaux de la statistique publique*, Série Etudes n° 2, Paris, AFRISTAT.
- J.-L. BODIN (1999b), *Réglementation en vigueur et pratiques en usage: aux Pays Bas, au Royaume Uni, en Suède, aux Etats-Unis*, Groupe de travail entre statisticiens publics et statisticiens privés pour réfléchir à la transposition en droit français de la directive 95/46 du Parlement européen et du Conseil, Paris (mimeo).
- F. BEHRINGER F., W. SEUFERT, G.G. WAGNER (1998), *Problems and examples of dissemination of 'scientific use of micro data files' in Germany and elsewhere*, in "Proceedings of the International Symposium on linked employer-employee data", 21-22 May 1988, Arlington (VA), U.S. Bureau of the Census (CD-ROM).
- M. BLAKEMORE (1998), *Customer-driven solutions in disseminating confidential employer and labour market data*, in "Proceedings of the International Symposium on linked employer-employee data", 21-22 May 1988, Arlington (VA), US Bureau of the Census (CD-ROM).
- L. BUZZIGOLI, C. MARTELLI, N. TORELLI (1998), *Accesso ai dati statistici individuali: l'esperienza di altri Paesi*, Rapporto di ricerca per la Commissione per la Garanzia dell'Informazione Statistica, Roma (mimeo).
- CNIS (Conseil national de l'information statistique) – Groupe de concertation sur la transposition en droit français de la directive européenne 95/46/CE (1999), *Rapport au Ministre de l'économie, des finances et de l'industrie*, Paris, Septembre 1999 (mimeo).
- COUNCIL OF EUROPE (1997a), *Recommendation No. R (97) 5 concerning the protection of medical data*, adopted by the Committee of Ministers on 13 February 1997 at the 584<sup>th</sup> meeting of the Ministers' Deputies, Strasbourg [with Explanatory memorandum].
- COUNCIL OF EUROPE (1997b), *Recommendation No. R (97) 18 concerning the protection of personal data collected and processed for statistical purposes*, adopted by the Committee of Ministers on 30 September 1997 at the 602<sup>nd</sup> meeting of the Ministers' Deputies, Strasbourg [with Explanatory memorandum].
- COUNCIL OF THE EU (1997), *Council regulation (EC) No. 322/97 of 17 February 1997 on Community Statistics*, "Official Journal of the European Communities", 22.02.1997, No. L 52/1-7.
- G.T. DUNCAN, T.H. JABINE, W.A. DE WOLFS (Eds.) (1993), *Private lives and public policies: confidentiality and accessibility of Government statistics*, National Academy of Sciences, Washington, D.C., National Academy Press.
- EUROPEAN PARLIAMENT AND COUNCIL OF THE EU (1995), *Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data*, "Official Journal of the European Communities", 23.11.1995, No. L 281/31-50.
- EUROSTAT (1996), *Manual on disclosure control methods*, Luxembourg, Office for Official Publications of the European Communities.
- EUROSTAT (1999), *Statistical data protection. Proceedings of the conference, Lisbon, 25 to 27 March 1998*, Luxembourg, Office for Official Publications of the European Communities.

- S.E. FIENBERG, L.C.R.J. WILLENBORG (Eds.) (1999), *Disclosure limitation methods for promoting the confidentiality of statistical data*, special issue of the "Journal of Official Statistics", 14 (4).
- A. FRANCO (1999), *Individual data in the Labour Force Survey: dissemination policy*, Working Party on the Dissemination of statistical information, Doc. No. DWP/04/99-III-7-EN (mimeo)
- J.J. HECKMAN, J.A. SMITH (1995), *Assessing the case for social experiments*, "Journal of Economic Perspectives", 9 (2), pp. 85-110.
- HM GOVERNMENT (1998), *Statistics: A matter of trust*, Presented to Parliament by the Economic Secretary to the Treasury, London, HMSO [known as the "Green Paper"].
- G. HOINVILLE, T.M.F. SMITH (1982), *The Rayner review of Government Statistical Services*, "Journal of the Royal Statistical Society", A, 145, Part 2, pp. 195-207.
- INTERNATIONAL STATISTICAL INSTITUTE (1986), *International Statistical Institute Declaration of Professional Ethics*, "International Statistical Review", 54 (2), pp. 227-242.
- S.P. JENKINS (1999), *Measurement of the income distribution: an academic user's view*, in CEIES, "Proceedings of the seventh seminar: Income distribution and different sources of income, Cologne-Germany, 10<sup>th</sup> -11<sup>th</sup> May 1999", Eurostat, Theme 0 - Miscellaneous, Luxembourg, Office for Official Publications of the European Communities, pp. 75-84.
- R. JOWELL, (1981), *A professional code for statisticians? Some ethical and technical conflicts*, in "Bulletin of the International Statistical Institute. Proceedings of the 43rd Session", Buenos Aires, Volume XLIX, Book 1, pp. 165-209 (with discussion).
- I. LLOYD (1998), *A guide to the Data Protection Act 1998*, London, Butterworths.
- E. MALINVAUD (1987), *Production statistique et progrès de la connaissance*, in "Atti del Convegno sull'informazione statistica e i processi decisionali", Roma, 11-12 Dicembre 1986, Annali di Statistica, Serie IX, Vol. 7, Roma, Istat, pp. 193-216.
- E. MALINVAUD (1997), *Effet des evolutions techniques et des changements de la spécialisation internationale sur les marchés du travail et les systèmes productifs: comment les statisticiens pourraient-ils relever les défis?*, "International Statistical Review", 65 (1), pp. 97-109.
- E. MARLIER (1999a), *The EC Household Panel Newsletter. Editorial*, "Statistics in Focus", Theme 3 - 2/1999, pp. 1-3.
- E. MARLIER (1999b), *The EC Household Panel Newsletter (3/99). New pricing policy*, "Statistics in Focus", Theme 3 - 16/1999, pp. 2-3.
- D.B. MCCARTHY *et al.*(1999), *Medical records and privacy: empirical effects of legislation*, "Health Services Research", 34 (1), pp. 417-425.
- S. MCCLEAN (1998), *Statistical microdata, macrodata and metadata on the Web: perspectives*, in "Proceedings of the 21<sup>st</sup> SCORUS Conference", Belfast, 8-11 June 1998, Conference Compendium, University of Ulster, pp. 4.5.1-4.5.8.
- R.H. MCGUCKIN, G. PASCOE (1998), *The Longitudinal Research Database (LRD): status and usefulness*, "Survey of Current Business", November 1998, pp. 30-37.
- K. MOTOHASHI (1998), *Institutional arrangements for access to confidential micro-level data in OECD countries*, in "Proceedings of the International Symposium on linked employer-employee data", 21-22 May 1988, Arlington (VA), U.S. Bureau of the Census (CD-ROM).
- OFFICE FOR NATIONAL STATISTICS (1996), *Maintaining the confidentiality of data*, London, HMSO.
- D. RAYNER (1980), *Review of government statistical services: Report to the Prime Minister*, London, HMSO.
- E. RETTORE, U. TRIVELLATO (1999), *Come disegnare e valutare politiche attive del lavoro*, "Il Mulino", 48 (385), 1999, pp. 891-904.
- P.D. REYNOLDS (1993), *Privacy and advances in social and political science: balancing present costs and future gains*, "Journal of Official Statistics", 9 (2), pp. 275-312.
- STATISTICAL OFFICE OF THE REPUBLIC OF SLOVENIA, EUROSTAT (1996), *Third International Seminar on statistical confidentiality, Bled - Slovenia, 2-4 October 1996. Collection of papers*, Ljubljana, Statistical Office of the Republic of Slovenia, pp. 9-29 (mimeo).
- STATISTICS CANADA (1998), *Report to the Data Liberation Board on management of the second year of operation of the Data Liberation Initiative*, Ottawa, Statistics Canada (mimeo).

- D.W. STEVENS (1998), *Confidentiality revisited: motives and consequences*, in "Proceedings of the International Symposium on linked employer-employee data", 21-22 May 1988, Arlington (VA), U.S. Bureau of the Census (CD-ROM).
- TASK FORCE ON ACCESS TO CONFIDENTIAL DATA FOR RESEARCH PURPOSES (1999), *Draft Commission Regulation on Access to Confidential Data for Research Purposes. (i) Draft Commission Regulations. (ii) Explanatory notes*, Working Party on Statistical Confidentiality, Documents Eurostat/A4/SS/21 and Eurostat/A4/SS/22 (mimeo).
- (1999), *The end of privacy. The surveillance society*, "The Economist", May 1<sup>st</sup>-7<sup>th</sup> 1999, pp. 13-14 and 19-23.
- U. TRIVELLATO (1999), *Progettare un'informazione statistica pertinente*, in "Atti della Quarta Conferenza Nazionale di Statistica, Roma 11-12-13 novembre 1998", Roma, Sistan-Istat, 1999, Tomo 1, pp. 49-72.
- UN STATISTICAL COMMISSION (1994), *Fundamental principles of official statistics*, adopted at the Statistical Commission's 473<sup>rd</sup> meeting on 14 April 1994, UN Statistical Commission Paper E/CN.3/1993/26, New York, United Nations.
- US BUREAU OF THE CENSUS (1997), *IT operation plan. Part I. Data Access and Dissemination System*, CB-DR-97-02-N, Washington, US Department of Commerce.
- L. WILLENBORG, T. DE WALL (1996), *Statistical disclosure control in practice*, New York, Springer.

## SUMMARY

### *Data access versus privacy: an analytical user's perspective*

The issue of privacy *versus* data access is at the crossroad of two streams of considerations: (i) the traditional concern of official statistical agencies for confidentiality, both for ethical reasons and as a guarantee for maintaining respondents collaboration; (ii) the more recent and broader concern about data protection, that extends to all kinds of personal data and of their uses. The paper takes the point of view of an analytical user, *i.e.*, of a person aiming at (collecting and chiefly) processing micro data for statistical/research purposes. It focuses on some basic features of the issue and on their policy implications. First of all, the societal role of statistics and research and their intrinsic needs are considered (Section 2). Indeed, they are the basis for motivating a credit of confidentiality when personal data are processed for statistical/research purposes. Some basic principles and guidelines for regulating this credit of confidentiality, as they are set out in recent international recommendations and regulations, are then summarised (Section 3). Statistical, technological and legal devices, and related practices, for protecting confidentiality while allowing the processing of micro data for statistical/research purposes are cursorily reviewed (Section 4). The state of the affairs at the level of the EU is finally discussed, as regards both rules and practices followed by Eurostat and draft legislation (Section 5). Concluding remarks stress the need for additional efforts to make micro databases widely available to researchers.

UNIVERSITÀ DI PADOVA  
BIBLIOTECA DI SCIENZE ECONOMICHE E ATTUALITÀ  
Via del Seminario 1, 35100 PADOVA

## RIASSUNTO

### *Accesso ai dati 'versus' riservatezza: il punto di vista di un ricercatore*

La tensione fra accesso ai dati e tutela della riservatezza è al crocevia di due ordini di considerazioni: (i) la tradizionale cura dedicata dalle agenzie statistiche ufficiali alla tutela della confidenzialità, per ragioni etiche e per mantenere la collaborazione dei rispondenti; (ii) la più recente e ampia preoccupazione per la protezione dei dati, che si estende a tutti i dati personali e all'insieme dei loro usi. Il lavoro affronta la questione dal punto di vista di un utente 'analista', cioè a dire di una persona interessata a (raccolgere e soprattutto) trattare dati individuali per scopi statistici e di ricerca. Inoltre, l'interesse è portato su aspetti basilari della tematica e sulle implicazioni che essi hanno per politiche di accesso ai/diffusione dei dati. Innanzitutto, si evidenzia la funzione sociale della statistica e della ricerca scientifica, con le connesse necessità in materia di accesso ai dati (sezione 2). Sono queste, infatti, le motivazioni alla base dell'assegnazione al ricercatore di un credito di riservatezza nel trattamento di dati personali. Si richiamano quindi alcuni basilari principi e linee-guida in materia, quali risultano da recenti raccomandazioni e direttive internazionali (sezione 3). Si passano poi in rapida rassegna gli strumenti statistici, tecnologici e normativi – e le connesse pratiche –, mirati a proteggere la riservatezza e nel contempo a consentire il trattamento di dati individuali, possibilmente anonimi e se necessario personali, per scopi statistici e di ricerca (sezione 4). Viene quindi esaminato lo stato della situazione a livello dell'Unione Europea, con riguardo sia a regole e pratiche seguite dall'Eurostat sia a innovazioni in corso di definizione (sezione 5). Le considerazioni conclusive sottolineano la necessità di ulteriori passi per rendere le basi di dati individuali più ampiamente disponibili a ricercatori.

UNIVERSITA' DI PADOVA  
BIBLIOTECA DI SC. STATISTICHE DEM. E ATTUARIALI  
Via del Santo, 22 - 35123 PADOVA