Matthew Fontaine Maury

The 19th-Century Forerunner of Big Data

Massimo Guarnieri

Computer speed and web interconnectivity are allowing us to manage and process data at unprecedented levels. In the past two decades, we have discovered new paradigms for extracting valuable information from such data, and it is very likely that this process has only just begun. These paradigms have resulted in market and financial forecasts of huge economic importance. The companies positioned to exploit them are creating new, powerful empires that navigate in often unexplored seas of opportunities in which regulations are weak and governmental actions against unconventional management techniques are frequent.

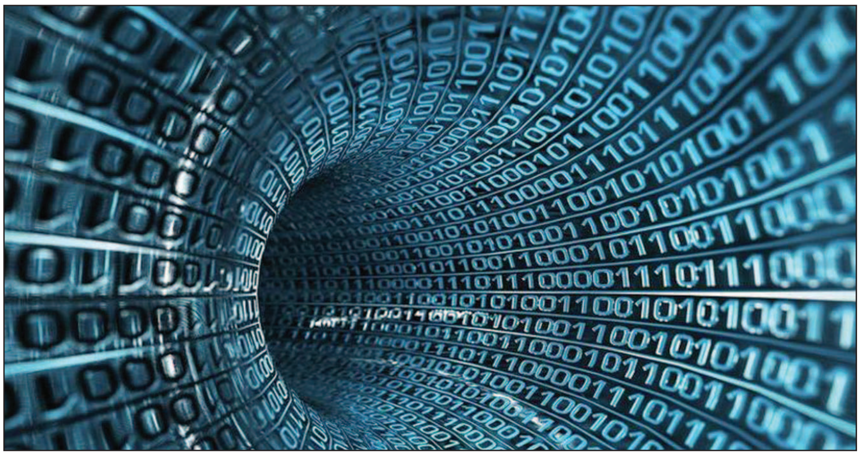

FIGURE 1 – An allegorical view of countless data representing the big data concept. (Image courtesy of Victoria Striker.)

Traditionally, data are collected in a restricted environment such as a laboratory experimental campaign, or by the proper sampling of a population, e.g., market surveys or from election projections. New possibilities have been discovered from the availability of huge amounts of data coming from sensors, devices, video/audio streams, networks, log files, transactional applications, and social media, much of which is generated in real time and on a large scale. This has now become known as *big data* (Figure 1). Today, big data is a term applied to data sets whose size or type is beyond the ability of traditional relational databases to capture, manage, and process information with low latency. They are characterized by one or more attributes identified by the letter "v," and three are considered most important [1]. Volume, which indicates the quantity of generated and stored data, is pivotal since data size determines its value and potential insight. Velocity refers to the speed at which the data are generated and processed to meet the required demands and challenges. Variety is related to the type and nature of the data, which are instrumental in effectively using the resulting insight. These "v" attributes are complemented by two more: variability highlights the consistency of the data set needed to handle and manage data, and veracity refers to data reliability.

If, on the one hand, big data offers the capacity to collect and analyze data at/on an unprecedented breadth, depth, and scale [2], on the other hand, it is notable not only because of its size but also because of its relation to other data. It is fundamentally networked, and its value is increased due to the patterns that can be derived from the connections between pieces of data [3]. On this basis, big data has emerged as a system of knowledge that is already changing the objects of knowledge, while having the power to inform on how we understand human networks and communities [4]. In short, it is producing a radical shift in how we think about research.

Big data offers new capabilities for accessing and connecting information, suggesting new patterns and models of interpretation. Analyzing big data allows analysts, researchers, and business professionals to make better and faster decisions. Using advanced analytic techniques, such as text analytics, machine learning, predictive analytics, data mining, statistics, and natural language processing, businesses can analyze previously untapped data sources to gain new insights, resulting in more efficient decisions. Massive quantities of information on people, things, and their interactions have become available to scientists researching in the fields of computer science, physics, economics, mathematics, biology, political science, and sociology. New technologies have allowed a wide range of varied operators including humanities and social science academics, marketers, governmental organizations, educational institutions, and motivated individuals [5], [6] to produce, share, interact with, and organize data.

In the next two years, the creation of 40 ZB of data is expected, an amount so large that there is no useful framing exercise to demonstrate its size. It is increasingly easy to produce massive amounts of data, and the combination and reuse of data from different origins can lead to new, unexpected results since the insights generated through analytics and combinations of different data sets are able to generate real values [7]. As long as privacy and security are maintained, the movement of data across borders has the potential to generate global economic gains. The World Economic Forum considers big data to be a new class of economic asset akin to human capital and natural resources, e.g., oil and gold. Big data is becoming increasingly important to major chief financial officers who can put strategies and solutions into place to ensure that investments in big data projects meet management expectations and deliver real value [8].

More recently, blockchain has emerged as a new information paradigm. Invented by unknown people in 2008 and formally attributed to a fantasy character named Satoshi Nakamoto [9], it consists of a continuously growing list of data blocks linked to each other and secured by means of cryptography. Since variations require a consensus of the network's majority, a blockchain is inherently resilient to change, resulting in an ideal data structure behind bitcoin and other cryptocurrencies. Additionally, its use has been proposed in other fields, such as identity management, the Internet of Things, food traceability, and even voting [10].

A wide-ranging debate is under way among diverse opinion groups about the benefits and risks of analyzing information from spaces where large amounts of data are deposited, such as Google, Facebook, Twitter, Verizon, 23andMe, and Wikipedia. Significant questions are being proposed regarding big data's effects on democracy, freedom, and people control. These questions lay far beyond the limits of this article, but another question is worth considering: Are there aspects of these new scientific paradigms that can be traced back to the past? Specifically, before computers were invented and started changing the way we manage information? I believe there is.

Matthew Fontaine Maury (1806–1873) was a U.S. Naval officer who left an indelible mark in oceanography and naval meteorology [11] (Figure 2). In 1825, at age 19, he embarked as a midshipman on the frigate *USS Brandywine* and started sailing the Atlantic Ocean. The following year, once in the Pacific, Maury embarked on the *USS Vincennes* and sailed on a four-year circumnavigation of the globe [12]. During his service, Maury, who was soon very intrigued by then-unexplained marine effects, started collecting navigational data. During his next cruise of four years beginning in 1831, Maury conceived the idea of new current and wind charts. One of his earliest observations was that when navigating Cape Horn, and depending on the time of the year, an inshore or a southern route had to be taken due to the impracticability of either route. He reported his findings in one of his earliest papers [13]. In subsequent years, he published other scientific papers and books on his nautical observations, often with great success [14]. In 1839, his career as an on-board officer came to an abrupt end, the result of a bad leg fracture from a stagecoach accident while traveling from his hometown, Fredericksburg, Virginia, to New York. Forced to give up sea service, Maury was assigned to the U.S. Naval Observatory in Washington D.C. in 1842. From this office, he was appointed head of the Depot of Charts and Instruments. It was here that he discovered a huge number of dusty log books of every trip made by American ships, crammed there since the American Revolution. Navigation data had been recorded faithfully in the ships' log books, but after their journeys, the log books had been stored and the registered data was forgotten. Maury realized the value of the information regarding currents, winds, and positions on seas and oceans broken down by the days of the year. The availability of such huge data inspired him to study navigation, meteorology, winds, and currents, and he drew from the records (instead of from direct observation) in a way and to an extent that had never been done before. He pored over the documents to collect information on winds, calms, and currents for all seas in all seasons and decided to assemble all of that data into a systematic structure (Figure 3). Realizing that many of the log books had been reported in different years (but on the same day, and in the same place with the same climatic conditions), he sensed that as the Earth annually made its revolution around the sun, even currents

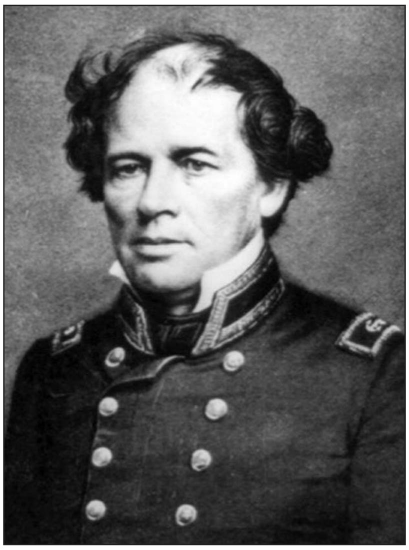

FIGURE 2 – Matthew Fontaine Maury as a U.S. Navy lieutenant, shortly before leaving for Brussels, Belgium. (Image courtesy of Wikimedia Commons.)

DURING HIS NEXT CRUISE OF FOUR YEARS BEGINNING IN 1831, MAURY CONCEIVED THE IDEA OF NEW CURRENT AND WIND CHARTS.

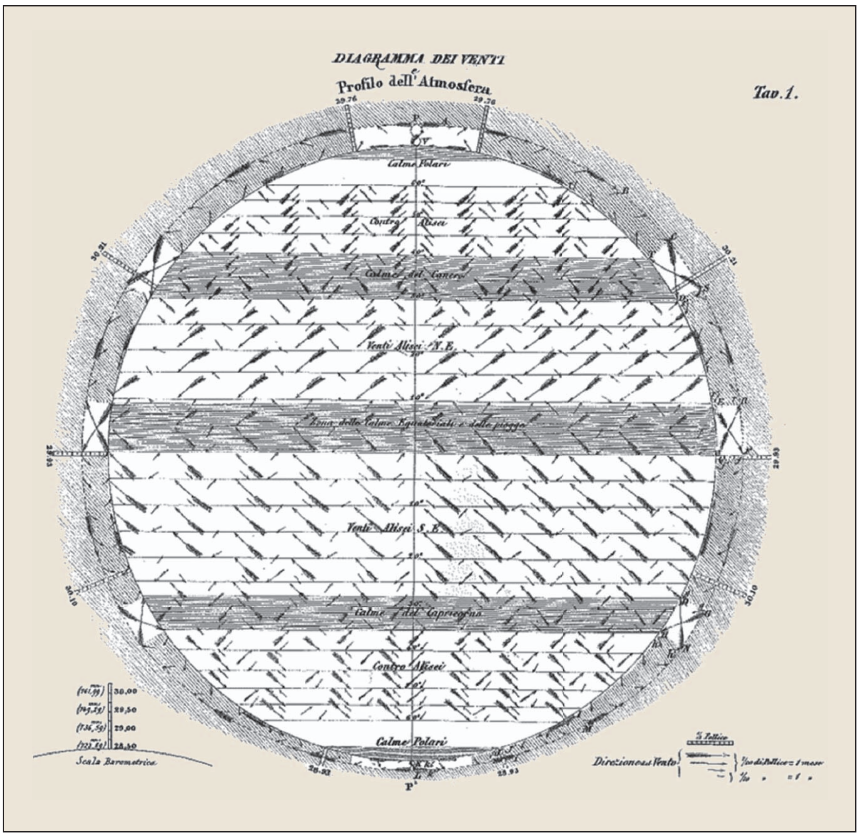

FIGURE 3 – A diagram of the wind from *The Physical Geography of the Sea*, by Matthew Fontaine Maury. (Image courtesy of Wikimedia Commons.)

and winds circulated and reappeared equally at regular time intervals.

His goal was to put this information in the hands of all the captains of all vessels [15]. In 1847, he published the results of his mapping work in the Wind and Current Chart of the North Atlantic Ocean. With information provided in the book, navigators could, based on the time of the year, trace new routes across the ocean much faster than those traditionally used, which were based on centuries of sailing experience. The first ship that followed Maury's instructions sailed the return trip from Baltimore, Maryland, to Rio de Janeiro, Brazil, in 35 days fewer than the usual 110 days that it took [16]. In 1853, Maury introduced an efficient representation of his data. He divided the maps into rectangular tessellations, wide 5° in latitude, and 5° in longitude. Each rectangle was designed with four concentric circles divided into sectors, with each containing wind and calm data for the four seasons. This system of representing synoptic oceanographic data was adopted by several countries and used to develop charts for all of the major trade routes.

New routes based on Maury's atlas resulted in a drastic reduction of ocean-crossing times, along with huge economical and time advantages. From 1855, Maury's currents mapping, based on 46,000 days of navigational records, helped to define the steamers' routes across the Atlantic, enabling coal consumption minimization. The travels of the clippers, the fast cargo sailing ships of the late 19th century, benefited significantly from Maury's findings.

Maury also promoted the institution of a reporting system of sea conditions and observations among the nation's shipmasters to gather additional data on sea conditions and observations, further enhancing and widening his information base. Moreover, convinced that improved scientific knowledge of the sea could be derived from better international cooperation, he urged the United States to promote a conference to establish his universal system of meteorology. At this pioneering conference held in Brussels, Belgium, in 1853, his proposal gained the full consensus of those in attendance. As a result of the conference, a large number of nations, including opposing countries, agreed to cooperate in the sharing of land and sea weather data using uniform standards. Within a few years, nations possessing three quarters of the world's vessels were sending their oceanographic observations to Maury's office, where the information was collected and processed and the results distributed worldwide. Maury's *Physical Geography of the Seas* and subsequent editions remain as classics of sea navigation around the world [17] (Figure 4).

By hand-processing huge amounts of data, Maury transformed the oceans from trackless hazards into a network of highways marked by dependable winds and currents and showed shipmasters how to save weeks and months of navigation time. In a world interconnected by maritime commerce, Maury's work was vitally important not just to America, but to all nations.

Maury did not restrict his ocean investigations to wind and currents. In search of major correlations to his analyses, he started deep-sea sounding in the Northern Atlantic Ocean in 1849. The results of this campaign revealed the existence of the Atlantic plateau that could facilitate the laying of a telegraph cable. Such information was pivotal to the first transatlantic cables that were laid by Cyrus W. Field's companies in 1858 and 1866 [18]. It is beyond the scope of this article to describe the rich service and political life of Maury, as well as his contribution as a Virginia native, to the Confederate cause during the American Civil War. At the very least, it is worth remembering that he was also an astronomer, historian, meteorologist, author, geologist, and a science promoter. As I have shown, his work was a forerunner to present big data concepts with regard to several points. It is a nontrivial consideration; it happened roughly 150 years ago, at a

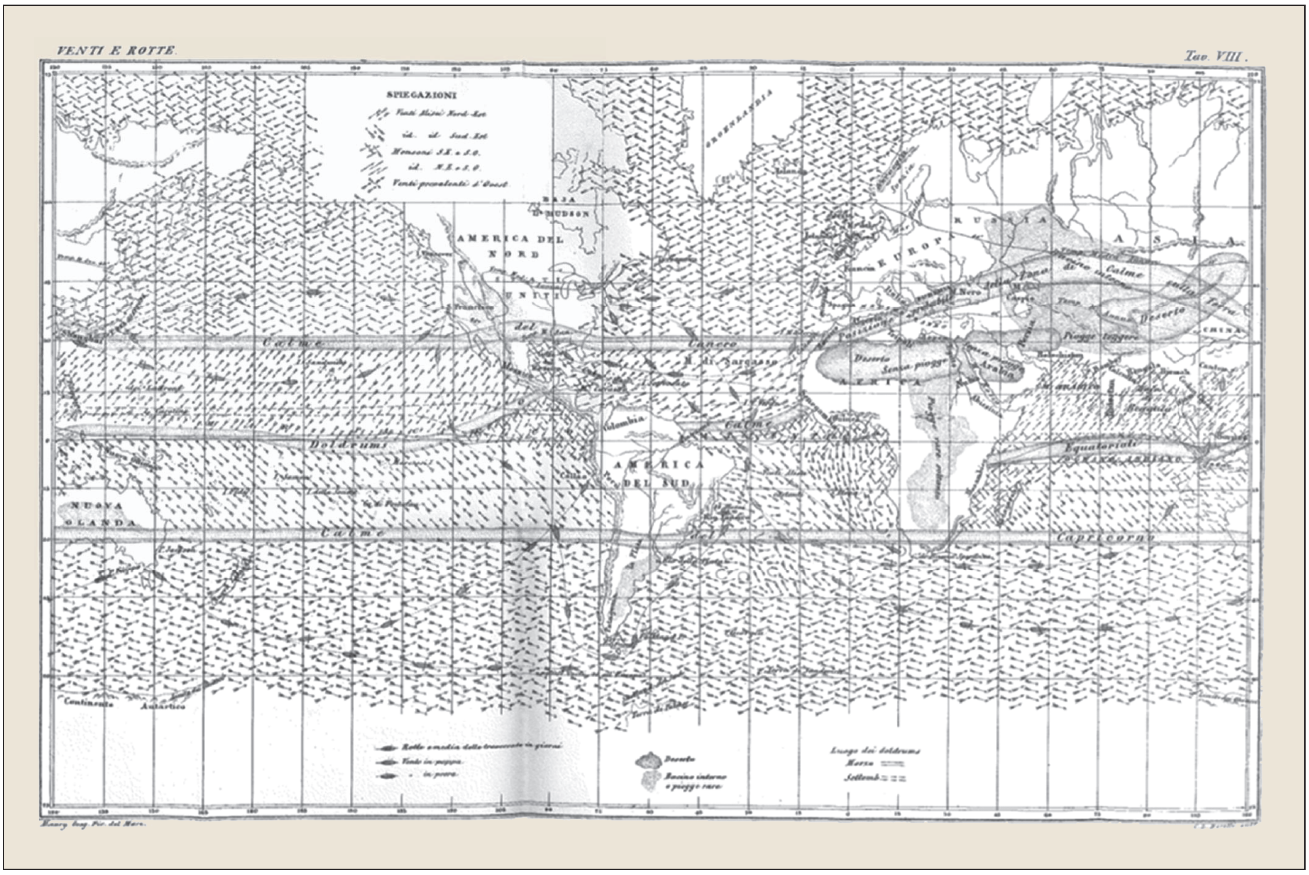

FIGURE 4 – A wind diagram from *The Physical Geography of the Sea*. These tables are the result of 1,159,353 distinct observations concerning the strength and direction of the wind, and, roughly, 100,000 observations on the barometeric measurements at sea. (Image courtesy of Wikimedia Commons.)

time when cooperation among nations was not commonplace.

Acknowledgment

I am grateful to Dr. Nicola Apolloni, whose master's dissertation, "Mastering the Oceans—Vita e avventure di Matthew Fountaine Maury, fondatore della moderna oceanografia," (Life and adventures of Matthew Fountaine Maury, Founder of Modern Oceanography) University of Padua, Italy, 2016–2017, was of significant help in writing this article.

References

[1] A. De Mauro, M. Greco, and M. Grimaldi. (2016). A formal definition of big data based on its essential features. *Library Rev.* [Online]. *65(3)*, pp. 122–135. Available: https://www.emeraldinsight.com/doi/full/10.1108/LR-06-2015-0061

[2] D. Lazer, A. Pentland, L. Adamic, S. Aral, A. Barabási, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann, T. Jebara, G. King, M. Macy, D. Roy, and M. Van Alstyne, "Social science: Computational social science," *Science*, vol. 323, no. 5915, pp. 721–372, 2009.

[3] D. Boyd and K. Crawford, "Six provocations for big data," in *Proc. Decade in Internet Time: Symp. Dynamics Internet and Society*, 2011, pp. 1–17.

[4] H. Chen, R. H. L. Chiang, and V. C. Storey. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly: Manag. Inf. Syst.* [Online]. *36(4)*, 1165–1188. Available: https://arizona.pure.elsevier.com/en/publications/business-intelligence-and-analytics-from-big-data-to-big-impact

[5] X. Wu, X. Zhu, G.-Q. Wu, and W. Ding, "Data mining with big data," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 1, pp. 97–107, 2014.

[6] M. Chen, S. Mao, and Y. Liu, "Big data: A survey," *Mobile Netw. Appl.*, vol. 19, no. 2, pp. 171–209, 2014.

[7] A. Schlosser. (2018, Jan. 10). You may have heard data is the new oil. It's not. World Economic Forum. [Online]. Available: https://www.weforum.org/agenda/2018/01/data-is-not-the-new-oil

[8] FSN Newswire. (2012, Dec. 18). Mastering big data: CFO strategies to transform insight into opportunity. [Online]. Available: http://www.fsn.co.uk/channel_bi_bpm_cpm/mastering_big_data_cfo_strategies_to_transform_insight_into_opportunity#.WoFK84JG3OR

[9] A. Narayanan, J. Bonneau, E. Felten, A. Miller, and S. Goldfeder, *Bitcoin and Cryptocurrency Technologies: A Comprehensive Introduction.* Princeton, NJ: Princeton Univ. Press, 2016.

[10] K. Christidis and M. Devetsikiotis. (2016, May). Blockchains and smart contracts for the Internet of Things. *IEEE Access.* [Online]. *4*, pp. 2292–2303. Available: https://ieeexplore.ieee.org/document/7467408/

[11] Wikisource. (2018, May 9). Sketch of Matthew Fountaine Maury. [Online]. Available: https://en.wikisource.org/wiki/Popular_Science_Monthly/Volume_37/July_1890/Sketch_of_Matthew_Fontaine_Maury

[12] C. L. Lewis, *Matthew Fontaine Maury, the Pathfinder of the Seas*. Annapolis, MD: U.S. Naval Institute, 1927.

[13] J. M. Lewis. (1995, Dec. 26). On the navigation of Cape Horn. American Meteorological Society. Boston, MA. [Online]. Available: https://journals.ametsoc.org/doi/pdf/10.1175/1520-0477%281996%29077%3C0935:WOTWSM%3E2.0.CO;2

[14] M. F. Maury, *A New Theoretical and Practical Treatise on Navigation*. Philadelphia, PA: E. C. & J. Biddle, 1845.

[15] M. F. Maury, *The Physical Geography of the Seas*. New York: Harper & Brothers, 1855.

[16] M. F. Maury, *Explanations and Sailing Directions to Accompany the Wind and Current Charts*. Philadelphia, PA: E. C. & J. Biddle, 1854.

[17] C. G. Hearn, *Tracks in the Sea: Matthew Fontaine Maury and the Mapping of the Oceans*. Crawfordsville, IN: McGraw-Hill, 2002.

[18] M. Guarnieri, "The conquest of the Atlantic," *IEEE Ind. Electron. Mag.*, vol. 8, no. 1, pp. 53–67, 2014.

[19] USS Maury AGS-16 Association. (2018). Matthew Fontaine Maury: Pathfinder of the seas. [Online]. Available: http://www.ussmauryags16.org/mathew_fontaine_maury.html