# Data Revolutions in Sociology

**Barbara Saracino**

**Abstract** Following the French Revolution, society becomes the leading dimension of existence, more prominent than the state and government. According to the "hygienic party", who promoted health and education, statistics can contribute to an understanding not only of the main causes of disease and death, but also of crime and unrest; scientific ground can be provided for social policy. All kinds of social phenomena are classified, measured, and made public, and large-scale statistical recurrences are identified. In the "avalanche of numbers" published, the first sociologists recognized mass regularities in order to identify laws of human behaviour. Nowadays, the availability of Big Data paves the way to new epistemological challenges. Will the new data revolution lead to a new paradigm in Sociology?

**Keywords** History of statistics · Big Data · Data deluge · Computational turn · Sociological knowledge

"The era of Big Data has begun. Computer scientists, physicists, economists, mathematicians, political scientists, bio-informaticists, sociologists, and other scholars are clamouring for access to the massive quantities of information produced by and about people, things, and their interactions. Diverse groups argue about the potential benefits and costs of analysing genetic sequences, social media interactions, health records, phone logs, government records, and other digital traces left by people. Significant questions emerge" (Boyd and Crawford 2012, 662).

The purpose of this paper is to analyse the historical period that witnessed the rise of social mechanics and the birth of sociology as specialized discipline and thus to contribute to the reflection on possible present and future developments, consequences, and implications of the use of Big Data for the production and development of sociological knowledge.

B. Saracino (✉)
Department of Social Sciences, University of Naples Federico II,
Vico Monte Della Pietà 1, Naples, Italy
e-mail: barbara.saracino@unina.it

In the seventeenth century, the idea that social phenomena could undergo quantitative analysis gained momentum. Demographic problems were the first to be discussed systematically, as the wide-spreading insurance systems required accurate numerical basis, and the size of the population was considered a crucial element for state power and wealth.

The study of social statistics started in 1660, and since that time for around a century and a half, it was known as "political arithmetic". Its purpose was the promotion of upright and well-documented state policies. In his *Observations upon the Bills of Mortality,* John Graunt wrote: "That whereas the Art of Governing, and the true Politicks, is how to preserve the Subject in Peace and Plenty; that men study only that part of it which teacheth how to supplant and reach one another, and how, not by fair out-running but tripping up each other's heels, to win the Prize. Now, the Foundation or Elements of this honest harmless Policy is to understand the Land, and the Hands of the Territory, to be governed according to all their intrinsic and accidental differences" (Porter 1986, 18). Policies had to rest on a profound understanding of the territory and its inhabitants, on concrete knowledge expressed in terms of numbers, weights, and measures.

According to Hacking (1990), Graunt and the English started the public use of statistical data, and Italian philosophers were the creators of the modern notion of state, but it was German thinkers and statesmen who were the first to gain awareness of the importance of data collecting. Instead of leaving it to personal initiative, a nation state needed to establish a specific organization, a central statistical office, in charge of collecting all the data necessary to define its own size and power. Leibniz was the spiritual father of Prussian official statistics. In 1685, he affirmed that a Prussian state had to be established, that the measure of the power of a state was its population, and that a state needed a central statistical office in order to know its own power.

According to Leibniz, this office had to be at disposal of all the administration branches and had the task to maintain a central register of deaths, baptisms, and marriages, in order to allow the estimation of population size. In those days, a general population census was deemed impracticable, as the numerousness of the population of a country, unlike a walled city or a colony, was not a measurable quantity. Only the establishment of designated institutions would finally allow it.

Leibniz was extremely interested in statistical questions of all sorts and pursued a rich correspondence about many issues of public health and demography. Prince Frederick II of Prussia wanted to be king of a united Brandenburg and Prussia, and Leibniz urged his case. Frederick's opponents argued that Prussia could provide only a limited contribution to unification with Brandenburg, so that the king should not be Prussian. But that was a mistake, according to Leibniz, as the real measure of the power of a kingdom was the number of its subjects: 65.400 children were born every year in the whole region, and 22.680 were Prussian, so Prussia was vital. Leibniz wrote these notes in 1700; the following year, the kingdom of Brandenburg-Prussia was established. Some years later, court officers created a system to register births, deaths, and weddings in the four main cities of the

kingdom. In 1733, the data about the population became a state secret.[1] During the Seven Years' War, a third of the population was decimated and colonization was required in order to restore ravished farmland. During the reign of Frederick, the list of the things that were counted extended up to seven pages.

Beyond state officers and private citizens collecting demographic data, such as Süssmilch, and beyond political arithmetic, Germany witnessed the development of "university statistics". The work of university statisticians was almost never quantitative. They feared that the indiscriminate use of data would add a materialistic character to the comparative study of states, ending up undermining educational and social value of their teaching. University statisticians believed it was necessary to distinguish between two sciences: a descriptive and non-numerical science, which was theirs, and another science that was heir of English political arithmetic (Lazarsfeld 1961).

At the beginning of nineteenth century, in Great Britain and France, political arithmetic was replaced by statistics. The change was not only terminological, but reflected a substantive transformation. Numerical statistics inherited an extraordinarily large field of application, from geography to climate, from commerce to population and culture. Statisticians started to investigate about all kinds of institutions and to collect data about commerce, industrial progress, work, poverty, education, health care, and crime. The extension of the field of numerical surveys is combined with an important change in the conception of their purpose. That may appear clearer through the comparison between two famous scholars, Süssmilch and Malthus, respectively, before and after the French Revolution (Porter 1986).

Before the French Revolution, Süssmilch, starting from the premise that population growth was the main aim of every ruler, devoted his work to show what the prince could do to promote demographic increase. After the French Revolution, Malthus argued that high density of population was the major cause of misery and poor health in a country. Population was not something flexible that could be manipulated, but the product of persistent customs and natural laws. Government could not dominate society, for it was itself conditioned by it. Malthus believed that society was a dynamic and potentially instable force, a source of trouble. Through statistical surveys, political leaders could have the chance to know the people and attempt to avoid disorders, introducing public education and informing about the true causes of poverty.

Collini (1980, 203–204) has written that in the first half of the nineteenth century, European intellectuals started to consider the dimension of society more central than state and government. Society was seen both as source of progress, constituting labour force for industrialization, and as cause of instability, symbolized by the French Revolution and by the incessant troubles in all Europe. "The emphasis upon the priority of the social came to be closely bound up with two characteristic features of nineteenth-century thought in general. The first was a

---

[1]Statistical data were treated as strictly confidential due to their potential military value. Later, revolutionary governments would proclaim the necessity to publicize them.

widely ramifying historicism […] The second was a profound commitment to a conception of the methods of natural science as man's only reliable cognitive relation to the world, and hence as the model for the study of human behaviour. Taken together, these beliefs constituted a character for a science of society, the project of discovering the natural laws which governed social development, and upon which political prescription and action were alike dependent".

During the nineteenth century, a growing number of scientists began to search for mass regularities and to overlook the causes of single events. They gradually realized with astonishment that single disordered, chaotic, or irrational phenomena showed unexpected regularities on large scale, and thus, they established a new type of law, the statistical law.[2] According to Hacking (1990), to ascertain the existence of statistical laws, both observation of regularities on large scale and a "right kind of readers" were necessary. Regularities became visible when social phenomena were classified, quantified, and publicized, that is to say after the "avalanche of numbers" published at the beginning of the nineteenth century. The right kind of readers, ready to find analogies between laws of society and laws of nature, were Western European intellectuals.[3]

The period defined by Westergaard (1932) as the "era of enthusiasm" for statistics started in the first decades of the nineteenth century in France and then developed with the Victorian statistical movement.

As Coleman (1982) has pointed out, starting from the 1820s in France, some "defenders of public health", especially military doctors on retirement after Napoleonic wars, took the initiative and conducted quantitative investigations. Their general interests were health and education, and collecting data could help them to understand causes of disease and death, criminality, and revolt. It seemed thus possible to obtain a scientific basis for a social policy.

Cullen (1975) has written that research in England, from the first half of Victorian age, had the main scope to celebrate industrial progress, blaming other causes, such as alcohol, moral degradation, and urbanization, for social unrest. Ignorance and dirt were considered responsible for diffusion of diseases, growth of criminality, and risk of national disorder within the working class. Statistical investigation would provide the empirical support for the necessary reforms.

According to Funkhouser (1937, 291), "an interesting development in the history of statistics is that of the gradual merging of political arithmetic and the theory of probabilities into a science of statistics in the first part of the nineteenth century. The students of political arithmetic had the urge for the scientific study of anthropological and political questions and were slowly improving their data in

---

[2]In 1889, Galton affirmed that the law of error "reigns with serenity and in complete self-effacement amidst the wildest confusion. The huger the mob and the greater the apparent anarchy, the more perfect is its sway".

[3]In *The Taming of Chance,* Hacking (1990) compared Prussian (and Eastern European) attitude towards numerical data with the position of scholars in Great Britain, France, and the other countries of Western Europe. It was the West, where libertarian, individualistic, and atomistic notions of person and state were spreading, to start to formulate social laws based on data.

quantity and quality, but they lacked a powerful enough tool to handle their problems. This tool was provided in the theory of probabilities". One of the most committed supporters of the application of probability theory to the study of social phenomena was Adolphe Quetelet.

Influenced by the works of Laplace and Fourier, Quetelet started to believe, around 1830, in the possibility of applying the methods of physical and natural sciences to human activities, going beyond the mere collection of data that was so fashionable at the time. Therefore, he starts to dig out statistical laws from the avalanche of published data, finds the cause of the observed demographic regularities in forces acting within society, and combines statistical interests with astronomical and mathematical instruments. In his opinion, mathematics can bring order out of the apparent social chaos and can give the chance to dominate scientifically seemingly uncontrolled social phenomena. The application of the law of errors to the distribution of human characteristics permits to confirm the hypothesis of social physics and to demonstrate that concepts and instruments of astronomy are the most adequate to catch the essential characters of the human being, the only entity that, until that time, was deemed impermeable to science. Quetelet thinks that persons are imperfect copies of the average man and that their growth is influenced by a large series of accidental causes and errors, just like the exact observation of an astronomical object or event. The function of the errors is useful to define a "type" and to identify that single cause of phenomena that is usually obscured by the action of disturbing causes. The first positivist sociologists considered Queletet's researches about humankind as a valid support to their idea of normality. The Belgian author conceived normality as optimal status to achieve; his idea was echoed by Durkheim who went as far as to oppose normal to pathological status, meant as deviation from the norm.

Nearly two centuries after the era of enthusiasm for statistics and Quetelet's works, we are discussing the pervasive character of the phenomenon of the so-called datification. The debate concerns the proliferation of data and information in the contemporary society of knowledge—founded on the diffusion of digital society, computers, informatics culture, and importance of Internet and made possible thanks to "information infrastructures" such as databases, networks, and interfaces (Lorenzet 2015)—but is also about the growing importance of quantification processes in organization contexts and of "governance by numbers" (Supiot 2015).

Due to the ongoing increasing digitalization, we are now witnessing a process of acceleration that is causing not only quantitative but also qualitative effects on the ways to realize and produce knowledge. One of the most distinctive traits of the current attention dedicated to Big Data[4] is perhaps the growing trust in machines for

---

[4]With Big Data, I intend here extensively all the data and information that are capable of meeting the requirements of all the characteristics that have been identified in the literature, such as the three "Vs": Volume, Velocity, and Variety (Giardullo 2015, 2).

the production of knowledge, combined with the shift from prevailing mechanical–analog technology to digital–algorithmic devices (Neresini 2015).

"To mediate an object, a digital or computational device requires that this object be translated into the digital code that it can understand. This minimal transformation is effected through the input mechanism of a socio-technical device within which a model or image is stabilised and attended to. It is then internally transformed, depending on a number of interventions, processes or filters, and eventually displayed as a final calculation, usually in a visual form. […] In other words, a computer requires that everything is transformed from the continuous flow of our everyday reality into a grid of numbers that can be stored as a representation of reality which can then be manipulated using algorithms" (Berry 2011, 1–2).

Digital devices have enabled a strong acceleration of the tendency towards mathematization which characterizes modern sciences and thus the way in which we describe, analyse, and intervene in reality to modify it. The so-called computational turn has enormously amplified the importance of one of the fundamental principles of the possibility of accumulating and producing knowledge, that is to say the capability of "acting at a distance".

According to Latour (1987), it is not possible to describe knowledge in itself, by opposing it to ignorance or to belief, as its very sense raises only when taking into consideration an entire cycle of accumulation. To accumulate means to acquire a familiarity with distant things, events, and people. How can you act at a distance on unfamiliar objects? By bringing them home, somehow. How can you do it, given that they are distant? By inventing devices that render them mobile, combinable, and stable. This mixture of mobility and combinability permits to dominate at a distance. Inscriptions (formulas, tables, and charts) accelerate the accumulation movement. To dominate at a distance, several actions are necessary: firstly, to translate the world in order to make it enter in "centres of calculation"; secondly, many elements must be moved from a distance without being really introduced inside to avoid their flooding in centres of calculation; and thirdly, new codes must be invented to hold maximum information in minimum space. Operating on the centres through a series of subsequent representations permits to obtain and keep an advantage: it is possible to obtain representations of nth-order that are combinable with other representations of nth-order giving the same mathematical structure to every element. The process of abstraction thus enables the gathering of the maximum amount of information in one single place.

If nothing is more stable, mobile, and combinable of a number in digital format, digitalization has brought to the utmost extreme the processes of abstraction, standardization, and action at a distance, with the double effect both of generating big quantities of data in numerical form and of increasing their agency.[5] The

---

[5]"Perhaps the most important element distinguishing Big Data from other huge collections of data, that is, census data, is the fast and automatic generation of a high volume of information, which means delegating data collection to an automatic device. In fact, huge databases are 'populated' through specific scripts that are nested in servers and types of counter machinery" (Giardullo 2015, 2).

production and the treatment of big quantities of data have made the role of algorithms everyday more necessary in the process of construction of knowledge.

Technically speaking, an algorithm is a codified procedure to transform an input into an output, "but as we have embraced computational tools as our primary media of expression, and have made all information digital, we are subjecting human discourse and knowledge to these procedural logics that undergird all computation. And there are specific implications when we use algorithms to select what is most relevant from a corpus of data composed of traces of our activities, preferences, and expressions". Gillispie (2014, 167–168) has considered algorithms that manage information not only as codes lines but also as "new knowledge logic", and he has identified different characteristics of their unprecedented "public relevance". Their public relevance includes the choice about what to include or exclude in the preparation of data; the implications of the attempts to know and predict algorithms' users; the criteria by which what is relevant is determined; the "algorithmic objectivity", that is to say how to position the algorithm in the face of controversy as an assurance of impartiality due to its technical character; the reshaping of users' practices in response to the algorithms they depend on; and the production of "calculated publics".

The expression "data deluge" was created within the Human Genome Project at the beginning of the 1990s to indicate the enormous quantity of data that molecular biology was starting to deal with. In most recent years, data deluge is overwhelming even the work of social scientists.

The increasing availability of Big Data for research (in private or public institutions) and the design of interventions (from marketing to public administration) have divided researchers into two opposite sides: on the one hand, the sceptics, who fundamentally question the legitimacy of the use of such data on the basis of privacy issues and other ethical concerns and on the other hand, the enthusiasts, who focus on the transformational impact of having more information than ever before (Gonzáles-Bailón 2013, 148).

As argued in *Nature* (2007, 637–638), "For a certain sort of social scientist, the traffic patterns of millions of e-mails look like manna from heaven. Such data sets allow them to map formal and informal networks and pecking orders, to see how interactions affect an organization's function, and to watch these elements evolve over time. They are emblematic of the vast amounts of structured information opening up new ways to study communities and societies. Such research could provide much-needed insight into some of the most pressing issues of our day, from the functioning of religious fundamentalism to the way behaviour influences epidemics […] But for such research to flourish, it must engender that which it seeks to describe […] Any data on human subjects inevitably raise privacy issues, and the real risks of abuse of such data are difficult to quantify".

However, even the enthusiasts have divided into two groups. As the availability of big quantities of data has grown, researchers have engaged in a debate about the opposition between hypothesis-driven research and data-driven research. They first believe that Big Data will radically change the way in which we make sense of the world: the data speak for themselves, and theoretical interpretative models are not

necessary. The main argument of those who proclaim the "end of theory" (Anderson 2008) is that the most measured and recorded age in history demands a different approach to data: in other words, being able to track human behaviour with unprecedented fidelity and precision is more powerful than imperfect models of why people behave the way they do. Alternatively, there are those who believe the exact opposite: that theory and interpretation are more necessary than ever before if we are to find the appropriate layer of information, to disentangle signal from noise, to identify meaningful correlations, and to discard those that are unsubstantial (Gonzáles-Bailón 2013, 148).

In the "avalanche of numbers" published at the beginning of the nineteenth century, the first sociologists recognized mass regularities in order to identify laws of human behaviour. Nowadays, the availability of Big Data paves the way to new epistemological challenges. Will the new data revolution lead to a new paradigm in Sociology?

# References

Anderson, C. (2008). The end of theory: The data deluge makes the scientific method obsolete. *Wired Magazine*.

Berry, D. M. (2011). The computational turn. Thinking about the digital humanities. *Culture Machine, 12,* 1–22.

Boyd, D. & Crawford, K. (2012). Critical questions for big data. *Information Communication & Society*, 15(5), 662–679.

Coleman, W. (1982). *Death is a social disease. public health and political economy in early industrial france*. Madison: The University of Wisconsin Press.

Collini, S. (1980). Political theory and the 'science of society' in Victorian Britain. *The Historical Journal, 23*(1), 203–231.

Cullen, M. (1975). *The statistical movement in early Victorian Britain: The foundation of empirical social research*. Hassocks: Harvester.

Funkhouser, H. G. (1937). Historical development of the graphical representation of statistical data. *Osiris, 3,* 269–404.

Galton, F. (1889). *Natural inheritance*. London: Macmillan.

Giardullo, P. (2015). Does "bigger" mean "better"? Pitfalls and shortcuts associated with big data for social research. *Quality and Quantity*, 1–19.

Gillespie, T. (2014). The relevance of algorithms. In T. Gillispie, P. Broczkowski, & K. Foot (Eds.), *Media technologies: Essays on communication, materiality and society* (pp. 167–194). Cambridge, MA: MIT Press.

Gonzáles-Bailón, S. (2013). Social science in the era of big data. *Policy and Internet, 5*(2), 147–160.

Hacking, I. (1990). *The timing of chance*. Cambridge: University Press.

Latour, B. (1987). *Science in action: How to follow scientists and engineers through society*. Harvard: University Press.

Lazarsfeld, P. F. (1961). Notes on the history of quantification in sociology: Trends. *Sources and Problems. Isis, 52*(2), 277–333.

Lorenzet, A. (2015). Le « scene mediali » della datificazione. La controversia tecnoscientifica sulla privacy nei quotidiani italiani. *Rassegna italiana di sociologia*, LVI(3–4), 581–607.

Nature (2007, 11 October). A matter of trust. *Nature*, 449, 637–638.

Neresini, F. (2015). Quando i numeri diventano grandi: che cosa possiamo imparare dalla scienza. *Rassegna italiana di sociologia*, LVI(3–4), 405–431.

Porter, T. M. (1986). *The rise of statistical thinking. 1820–1900*. Princeton: University Press.

Supiot, A. (2015). De L'harmonie par le calcul à la gouvernance par les nombres. *Rassegna italiana di sociologia*, LVI(3–4), 455–464.

Westergaard, H. (1932). *Contributions to the history of statistics*. London: P.S. King & Sons.