Quality & Quantity
https://doi.org/10.1007/s11135-022-01351-7

Sociology between big data and research frontiers, a challenge for educational policies and skills

Stefania Capogna[1]

Accepted: 9 February 2022
© The Author(s), under exclusive licence to Springer Nature B.V. 2022

Abstract
The paper focuses on the challenges posed within sociology and social research by the transformations created by the "data society". To this end, the paper outlines some of the most significant elements for new frontier research which sociology is forced to confront also in relation to the challenges for educational policies and skills. While leading literature decries the need to promote alphabetising data, otherwise defined as *data literacy*, the idea proposed here is that it is necessary to work towards *understanding data*. An understanding that highlights the role sociology has firmly set itself since its foundation, that of studying and explaining the complexity of the relationships that characterize social life in every context and period.

Keywords Methodology · Sociology · Big data · Mixed methods · Social research

1 Introduction[1]

The paper focuses on the challenges posed within sociology and social research by the transformations created by the "data society". With the so-called *datification* process (Van Dijck, 2014) made possible by the digital revolution, it is possible to observe the growing power of platforms, digital infrastructures (public or private) which, in consideration of their ability to capture, assemble and manage unimaginable shares of *data* are able to increasingly guide the culture of measurement, *ranking*, forecasting and evaluation. This trend has significant repercussions on: policy systems, increasingly oriented towards *auditing* and *accountability logics* (Strathern, 2000); economic circles, involved in the race to control data for its commercial applications (for example, the so-called GAFAM: Google, Amazon, Face Book,

[1] On the same topic, see also Capogna (2021) about the framing effect

✉ Stefania Capogna
s.capogna@unilink.it

[1] Digital Technologies, Education & Society Research Centre, Link Campus University, Via del Casale di San Pio V, 44, 00165 Rome, Italy

Published online: 05 March 2022

Apple, Microsoft (Srnicek, 2017); political circles, that do not defer from exploiting *social* platforms to make targeted predictions and propaganda (e.g. Facebook, Cambridge Analytics); military control, in order to maintain supremacy, security and for field applications (Chin, 2019); the *inebriation of forecasting*, fuelled by the illusion of control over the world, made possible by scientific advances in predicting outcomes (Schoen, Avello et. al., 2013).

All of this inevitably influences research, with several differences in terms of priorities and opportunities for access to funding between the so-called hard sciences, which can have immediate returns and are easily measurable, and the social sciences, characterized by long term impacts and which are difficult to pigeon hole using rigid evaluation procedures.

With these premises in mind, we intend to reflect on the new frontiers of sociological research. In a society overwhelmed by the pervasiveness of data which, at all levels and across all sectors and contexts, requires an increased and widespread ability to work with the *language of numbers*, a certain difficulty in understanding its presuppositions and repercussions can be observed.

While leading literature decries the need to promote *alphabetising data*, otherwise also defined as *data literacy*, the proposal we wish to make is that it is necessary to work towards understanding the trends that the data aim to outline. An understanding that highlights the responsibilities of sociology in assuming the decisive role that it has set itself since its foundation; a role aimed at understanding, in the Weberian sense, the complexity of relationships and their evolution over time and space. Other frontiers of the social sciences have long started a fruitful discussion on these issues giving life to new fertile research fields such as those connected to the development of computational social science (Gloor, 2007; Abbington, 2019; Alvarez, 2020), the SAM approach (Cheung, Jak, 2016), research in the field of human resources (Zhang et al., 2021) and Humanities (Chen, Yu, 2018); while sociology, especially in Italy, despite the mobilisation of more recent times, struggles to claim its active role in this field, under the weight of the age-old, and never completely resolved, quantity-quality clash.

There are two crucial issues around which this reflection unfolds:

a) the implications for sociology that the digital transformation presents;
b) overcoming of contra-oppositions to the method, for reorganisation purposes from the standpoint of study.

The question that guides this reasoning can be summarised as follows: what methodological implications can be delineated for sociological research?

On the basis of these brief considerations, the aim of this essay is to explore the areas of transformation that requires this field of study to face the "threat" of new advances, other areas of knowledge apprised by the language of numbers (mathematics, statistics, information technology, economics) with the new type of data, to discuss in a renewed and dialectic way with other fields of knowledge, and to continue to exercise that transversal and interdisciplinary view which is absolutely necessary for interpreting contemporary complexities.

Therefore, the goal is to trace the most significant elements for new frontiers of sociological research. Starting from a reflection on the concept of *big data*, we briefly reconstruct its most salient stages (§ 1); the need to form new areas of competence for understanding contemporary complexities (§ 2); to reason, ultimately, on method implications (§ 3) and on new research prospects (§ 4). The conclusions reflect on the implications that affect edu-

cational policies which are faced with a twofold problem: the acquisition and management of data useful to support decision-making processes; the formation of sociological skills at different levels of mastery and for different professional areas that have nothing to do with traditional research.

2 A brief history of *big data*

Although it is not possible to identify with certainty who coined the term *Big Data*, its diffusion is linked to the analysis of the McKinsey Global Institute (2011) which illustrates its potential for application in terms of *business.*

Compared to previous periods, *big data* introduce some relevant new elements that have established themselves as challengers to research in all fields.

It is possible to distinguish, in synthesis, three essential stages of development, which correspond to a progressive fusion between different domains of knowledge and the expansion of applicative spaces.

The first phase (1963-2000) is influenced by the rapid evolution that affects the statistical sciences and is oriented to the development of data storage, extraction and optimisation systems in *Relational Database Management Systems* (RDBMS). This phase, governed primarily by skills and professionalism related to the statistical/mathematical domain, provides the basis of modern data analysis, as we know it today, through query techniques, known as *database queries*, analytical processing *online* and standard reporting tools. This can be recognised as the initial phase in which the foundations are laid for all subsequent developments, which are increasingly guided towards considering the value of statistical evidence with a view to interpreting the world and prefiguring development scenarios.

The second phase starts from the 2000s, with the advent of the Internet and the spread of the *web* on a large scale and the consequent possibility of collecting information in a completely new way. *Web* traffic, based on the *HTTP* code, has introduced an enormous increase in semi-structured, unstructured and standardised data, which offers innovative storage solutions to effectively manage these informative resources, paving the way for their extensive use in social, commercial and academic settings. The development and spread of the *social media* has greatly enhanced this phenomenon. During this phase, the relevance of computer skills and neighbouring fields (cybernetics, artificial intelligence, *machine learning*, etc.) is overwhelmingly affirmed within this area of tumultuous expansion, where knowing how to collect, organise and extract significant information from these unstructured information systems is becoming increasingly important.

The third phase (still in progress) is identified by some as the *datification society* (Van Dijck, 2014) which sees alongside the enormous, multi-faceted developments in digital technology, the push towards the capitalization of unstructured *web*-based content, with the ability to retrieve and monetize information collected from digital devices and platforms; information which is directly, freely and/or involuntarily released by subjects, or automatically collected by digital devices. Today, for example, there is great emphasis on the plethora of *mobile* devices that offer the possibility of: analysing behaviour; storing and analysing location-based data (GPS); tracing movements; and studying physical behaviour and health (e.g. pedometer). These functions open up completely new work areas, ranging from the possibility of rethinking, monitoring and intervening in: transport, city plan-

ning, health care, democratic participation, political engagement, training, work, and so on. This is accompanied by the whirlwind increase in internet-enabled, sensor-based devices. An evolution that contributes to increasing data generation like never before. Notoriously coined as the *Internet of Things* (IoT), millions of televisions, thermostats, wearables and appliances generate an indefinable amount of *zettabytes* of this traffic every day. In this phase, new epistemic environments appear in a world of opportunities opened up by *big data*: economics, finance, communication and marketing, neuroscience, etc.

Over the last few years, unprecedented connections and collaborations have flourished, in an endeavour to create a unified transversal collaboration between these fields. The greatest impact has increasingly emerged from the triangulation between state-of-the-art statistical methods, the developments of *computational science* and innovative theories in different fields of application (King, 2016). Yet, on being asked to understand the unique theoretical-empirical contribution that it can make in this modernized framework of complexity, sociology's contribution still appears reticent. However, the growing development of data repositories for open science (Davidson et al. 2019), the evolution of Data Analytics (DA) (Rawat, Sood, 2021), of processors for natural language analysis (NLP) (Franzosi, 2021) and for Text Mining (Greco et al., 2020) increasingly highlights the need to reconstruct a qualitative and quantitative perspective to overcome an ever-present risk inherent in social research based on partial and self-referential concepts, that of sterility and discontinuity (Smith 1988: 28).

To start with a reflection on *big data*, a definition, albeit brief, is required. In general, it is agreed that they are "disordered data." A fact that, by itself, is enough to undermine the social sciences and sociological research, which right from their outset have been based on "ordered" data *formats* that are clearly defined and precisely controlled and reproducible, according to the logic of the "scientific method" (Popper, 2002).

Today, *big data* commonly are defined in terms of the so-called 3Vs, according to a well-known definition illustrated by Doug Laney in 2001:

- *volume* as they are able to collect huge amounts of data in real time and without any human intervention;
- *variety* because they come from sources (platforms), diversified and generally not communicable with each other;
- *velocity*, in the sense that they self-produce by different ways and methods, at a whirlwind pace.

But the use of *big data* poses a series of problems in terms of:

- *veracity* which makes us question the trustworthiness of the information, considering that it can be altered in a variety of ways [(e.g. *digital divide*, mystification (Marr, 2014)];

• *value* because, as in the era of mining exploration, in the race for colonisation and political-economic-military exploitation, the value of *data* can be capitalized upon through *business analysis* processes (Marr, 2014);

• *variability*, in the sense that these data undergo fluctuations linked to different *trends* for example those determined by the time period, by seasonal changes, by the fact that they

come from different sources that must be connected, "cleaned", transformed, contextualised (McNulty, 2014), and "historicised", as Weber (1964) already explained in an attempt to clarify the fine line that unites and distinguishes *historical sociology* and *sociological history*.

Finally, *big data* are characterised by the fact that they are:

- *unique* in the sense that each micro-data is uniquely identifiable (Dodge, Kitchin, 2005), thanks to the traceability guaranteed by digital technology. This element is moving the frontier of exploitation further forward, giving rise to new phenomena (markets) such as that of *cryptocurrencies*, *bitcoins* or, to remain in the educational field, of *openBadges* and *Blockcerts*[2];
- *relational* (Boyd & Crawford, 2011) as they can be linked with data from other sources;
- *extensible* in the sense that information can be added and/or changed easily;
- *scalable* because, thanks to their structure, they can be expanded rapidly and without additional costs (Marz, Warren, 2012).

3 From data *'literacy'* to *understanding* data

Through this whirlwind development of digital technologies, accompanied by the evolution of theoretical-interpretative models in every sphere of knowledge and the maturation of new fields of expertise, a large current of ideas which attributes *value* to *data*, has established itself. The growing power of technical-digital infrastructures that accumulate and process the perceived *data*, like the *information* in the 90s (and reminiscent of the speculative bubble that developed around the idea of the *new economy*), see it as a precious resource to be exploited to produce new wealth in the 21st century. In this way, the nascent *data science*, commonly summarised by the label *digital literacy*, broadens the horizon of a thriving market, defined by promising fields of application. Added to this is the acceleration caused by the ordeal of the pandemic crisis over the last two years which accentuates the need to coincide the digital transition with the formation of a culture aimed at understanding and mastering both the numerical and natural language which, in the "space-time" of the web, create exponential relational and social dynamics, bringing to light new challenges in every field, primarily the ethical-social one. An acceleration made even more urgent by the Next Generation EU plan (NGEU), a European initiative to provide financial support to all member states to help them to recover from the adverse effects of the COVID-19 pandemic, relaunching all the countries and outlining the lines of collective development for the next few years by digital and green transition.

Returning to the topic of the essay, another element of acceleration can be found in the evolution that has already marked developments in Artificial Intelligence (AI) since the 1970s, now an integral part of the business model of entire sectors (medicine, education, work organisation, safety) at global level (AA.VV., 2021). This invites sociology to intervene more decisively regarding these issues, to make its indispensable and unique contribution to the interpretation and training of specific skills to overcome the logic inspired by

[2] The term *Blockcerts* refers to a digital certificate in addition to the physical degree certificate, which can remain accessible and verifiable over time and directly certifies its authenticity, and therefore is especially beneficial in.student and work-related mobility processes. The dissemination of these certificates is supported by the Bestr platform: https://bestr.it/about.

mere *alphabetising of data.* By this we mean the ability to read, understand, create and communicate data (a case in point being how an education in finance is considered extremely important in contemporary society) to direct thought towards prioritizing *understanding*.

The dominant debate, in fact, tends to explain this alphabetising (*digital literacy*) as the result of the combination of the following three areas of competence:

- *information literacy* which means "*The ability to access, evaluate, organise and use information in order to learn, problem-solve, make decisions—in formal and informal learning contexts, at work, at home and in educational settings*" (National Forum for Information Literacy);
- *technical skills* that refer to the set of skills or technical knowledge used to carry out practical tasks in the fields of science, arts, technology, engineering, and mathematics, the so-called hard sciences;
- and *statistical skills* to which we refer the ability to know how to manipulate, organise, explore, work with data and communicate.

From this perspective, the social outlook underlying these processes is totally absent. As if the analysis and management of such data were a mere technical and technological exercise, a myopic interpretation that loses sight of the human and humanistic dimension that hides behind the *language of numbers* and the symbolic violence[3] that, sometimes too simplistically, accompanies interpretations.

This absence of the necessary interpretation of relational dynamics from the dominant debate on the centrality of *data literacy skills*, typical of the epistemological frameworks that apprise this discipline, confirms the reticence and delay of sociology in intervening in this debate and field of application. While *big data* increasingly assume a value capable of "*knowing capitalism*", for many sociologists they remain an open question (Goldthorpe, 2016: 80-81).

3.1 The sociological perspective

In the face of the race for the new *digital Far West*, a timid awareness of sociology highlights the need to preside over this field of study, too long considered to be relevant to of other disciplines. Suffice to say that the first scientific article introducing the term *digital sociology* was written by Wynn in 2009. Before him, Savage & Burrows (2007) had addressed these issues, albeit without coining a specific term.

In fact, *big data* cannot be trivially conceived as a numerical representation/coding of individual behaviour (e.g. behaviour of consumers/users of the network) as social reality itself is radically transformed by digital processes, which are never neutral, coming to completely reconfigure the *fields of action* (Lewin, 1936) and the sphere of *social action* (Weber, 1922) that directs individual and collective action.

Through the process of *commodification* of *data*, which transforms personal information into goods, the subject is reduced to an *object.* Following Lukacs (1967), who picks up and

[3] Bourdieu (1977) sees the naturalisation of power relations, and the resulting symbolic violence, as the result of two correlated processes "dehistoricisation" and "universalisation", by means of which social inequalities and injustice are normalised, and no longer seen as problems to be fought against but as natural and inevitable conditions.

expresses the concept of *reification* present in the first volume of' the famous book, *Das Kapital*, by Marx(1867), the subject comes to be opposed to what he himself voluntarily and/or involuntarily produced about himself on the net. All this gradually becomes independent of him until he remains dominated by it, through autonomous laws that are foreign to him and on which the shadow of "cognitive capitalism" is grafted, revealing new forms of exploitation of labour, and of human and natural resources (Gorz, 2004). Unwittingly, individuals come to base their activities on the network through a mere collection/exchange of data *(like, followers*, stories), scattered and disconnected, with no obligation to any *relationship* that would imply otherwise, the encounter with the *other self* (Mead, 1934), recognizing and identifying the *person* who is inside and behind the *bit*, through that process of empathic recognition (Stein, 1958) which is embodied in the communicative relationship.

When the *data* is treated as an *object*, the subjective, contextual, relational and situational nature that contributes to its construction, management, action and interpretation is overlooked. We also lose sight of the relative field of forces where relational dynamics are always asymmetrical, therefore, subject to the influence of *power*. A power, that in the transition from domestic to global made possible by digital technology, is exponentially large, bringing to light even cultural implications Gluesing (1988).

In this way, it helps to duplicate the irreconcilable polarity between *objectivism* and *subjectivism*, widely debated by the harbingers of the discipline from its outset, with Durkheim (1897) the bearer of a *positivist* outlook of reality, due to his attention to *facts*, and Weber (1922) the inspiration behind *interpretative sociology* through his concept of "social action endowed with meaning".

Faced with such transformations, social inquiry is faced with a double paradox.

The first actually relates to the theme of research, opening the way to two opposing scenarios. On the one hand, a large amount of data is available, relatively accessible, more or less pre-packaged to facilitate interpretation by means of tools *data visualisation*, which can be diversely used with regard to the methodological rigor that needs to guide the analysis; on the other hand, sociology loses spaces of legitimacy, competence and capacity for action in the overall framework of the areas of scientific knowledge that make use of this type of investigation. And so, we witness incredulously, and perhaps unaware of, the "plunder" of sociology's own conceptual tools of, for the use and consumption of neighbouring areas of knowledge.

The second involves the field of training and skills that are useful and necessary to carry out sociological research in context and for diversified purposes. In a world with a high degree of complexity and systemic and inter-systemic turbulence, which requires an ever greater capacity for critical analysis, transversal to the various sectors, roles and professions, we find ourselves impoverished of the necessary *sociological imagination* (Mills, 1959) essential to *plow the field of research*. In other words, the greater the demand for widespread critical thinking in every sector (*education*, work, health, politics, globalisation, technology, public administration, welfare, etc.), where sociology could play a leading role, the more difficulty the discipline seems to have in recognizing and maintaining its own specific area for action and training, in comparison with other fields of scientific knowledge.

Sociology, which arises from the understanding of social complexity introduced by modernity, by an interdisciplinary lens and a holistic approach today seems to suffer from a sort of lack of *data literacy* that limits its power of influence in the public arena; the ability to attract research resources; the possibility of recognizing itself as a scientific community

around a common language, numerical-natural/symbolic, which is functional to overcoming any ideological contrast between quantity and quality.

4 Methodological implications

In the past, the problem of social inquiry was confronted with the difficulty/need to collect data. Today, on the contrary, it is confronted with the double miracle of the digital age (Franzosi (2021), the exponential growth of big data and the development of tools for Natural Language Processing (NLP) and, at the same time, with the hegemony of Western scientific knowledge. A hegemony determined both by the emergence of English as a lingua franca in the scientific field, and by the current research evaluation model powered by *open science technology platforms*, the most important of which at a global level are of Anglophone origin, coming in this way to colonize the scientific debate.

Among the most relevant difficulties introduced by *big data* and the ease of access/use of these information resources, we can list some which are certainly not exhaustive, but which are of particular interest to socialogical research, especially the *education* sector.

The concept of *representativeness* is the basis of any serious research aimed at explaining any social phenomenon. However, the lack of representativeness on a statistical level does not prevent the elaboration of meaningful analyses of phenomena that can take on a thematic value, sound out original and/or emerging research pathways, bring out hidden needs, and penetrate hidden dynamics that require alternative approaches. In addition, social research that favours the use of digital infrastructures clashes, *primarily*, with the issue of the *digital divide*, which we know influences at different levels, creating a vast plethora of excluded/invisible people who cannot be reached through *online* research.

The concept of *reliability*, strictly connected to the issue of veracity, in consideration of the fact that *big data* feeds on the following two main channels. (a) Self-production linked to movements carried out on the *web*, by virtue of the traceability guaranteed by digital technology and through which all our actions, consuming, asking, doing, buying, and seeing, leaves clearly unequivocal "traces" (Gray, 2010). And since the digital presence is increasingly subtle, pervasive and ubiquitous, we mostly do not realize the pervasiveness of digital technology in our daily actions. (b) The tendency to transfer personal information knowingly/unknowingly (e.g. *cookies*, *privacy* management, *likes*, etc.) or by necessity, through accessing both public and private sites and/or resources.

Another element strictly connected to the issue of the *veracity* of data is confronted with the *arbitrariness* that characterizes the provision of personal information, which can be easily falsified (*bots*[4] and *trolls*[5] come to mind), but even excluding the topic of intentional falsification when information is released voluntarily (as in the case of social media *Facebook, Instagram, Linkedin, Tik Tok* etc.), it is evident that they are filtered by what the subject wants to show and share at that moment, that is, from the image you want to show of yourself. In this case, we speak of *self-glazing* by means of that virtual over-exposure which, by retro-acting on identity formation (Mead, 1934), contributes to directing the rela-

[4] The *bots* are complex artificial intelligence devices with natural language processing (NLP) in order to communicate with humans (e.g. Alexa, Cortana, etc.)

[5] The term *troll* in Internet jargon means an artificial intelligence designed to simulate a person interacting with others through provocative messages.

tional dynamics in the virtual/real contraposition. It is not coincidental that we talk about an *image-based society.*

Another essential element is found in the *reproducibility* of the result, which we know to be a fundamental prerequisite of the scientific method (Popper, 2002). This is a central point, as any digital platform/resource/technology must be considered as a *social artifact*, therefore, the expression of a specific vision of the world. As such, it is never neutral, it is rather a *black box* that always incorporates an asymmetry relationship understood by the fact that the user, even when a researcher, does not know and does not have access to the logic, objectives, constraints and to the different rationalities that have determined certain choices. We are always confronted with an unavoidable *opacity* that clashes with the cardinal principle of *reproducibility* and *control*, along the whole chain of *theoretical-methodological choices* that are the basis of data collection. In other words, the new "worlds of data" pose an irremediable problem of insight and understanding the profound reasons that inform and contribute to their acquisition and relative diffusion.

Within this centrifugal tension, social research, even in *education*, *loses control* over data formats, construction processes, possible distortions and/or other sources of error in their production. Social researchers do not have control over *the data chain* and they do not play the role of designers of collection tools (Salganik 2018). The process of operationalizing the dimensions of constructs in variables is not in their hands as in classical surveys, but is managed by engineers, computer scientists, statisticians and other professions involved in the daily practices of *hardware* implementation, or even by sensors and programs (*Apps*, algorithms) designed to automatically feed the survey platform.

The theme of *validity* introduces the relationship between *theory* and *empiricism*, which is expressed through the coherence between the concept and the ability of the instrument to measure what it is intended to detect. A common criticism is that the development of research methods based on *big data* direct *data-driven* (Kar, Dwivedi, 2020) rather than *theory-driven* research, with the risk of working on superficial information, without having sufficient awareness of the influences determined by the context, the environment, the same data generation tool and from the specific *fields of action* within which it is built.

Desrosières (2015, 2016) explores the theme by suggesting the need to distinguish between *measurement* and *quantification*. In his opinion, *quantification* comes first and indicates the defining process that leads to the elaboration of univocal, standard concepts, the elaboration of classification and measurement procedures. The defining process develops along a continuous pathway of negotiation, coordination and critical review which involves different actors, belonging to different disciplinary and professional fields and which leads to the construction of a system of shared theoretical-methodological conventions. This means that *quantification* is anchored to a specific justification logic that legitimises the system of detection and *measurement* of the phenomenon under consideration. These conventional logics are guided by value principles that direct the different perspectives, through which it is possible to problematise the social world. From this, we can deduce that the *measurement* system incorporated in the digital platform/device reproduces this outlook, and in a new way, the *false consciousness* of adorned memory (Adorno, 1956). Considering *measurement* and *quantification* in this way allows us to penetrate the *black-box* of *big data*, to reflect on: what moves behind the conventions of measurement (Salais 2016); the processes and rationalities underlying the construction of automatic detection systems; the logic of *open data*; the internal coherence of the theoretical-methodological structure; the methods

of data communication and consultation; the limits and approximations that are inevitably connected to the numerical language, just as to the alphabetic language; the values that inform the platform; the stakes of the actors who contribute to the design, implementation, management, supply of the data sets and their related release.

Moving on to another list of considerations, it should be remembered that, in general, due to the pressure of speed, economy and the possibilities offered by computing power, research *with* and *through* digital platforms tends to focus on its structural characteristics, assuming for example in the network analysis, "that the network structure by itself virtually determines the action" (Uzzi, 1997: 63). This hypothesis often underpins research on career advancement and business opportunity networks, starting with the recognition of an advantage for those "closely connected in social networks" (Collins, 98: 44-45). A perspective of this type is the expression of an overdetermined vision that does not take into account the transformative and emancipatory power inherent in the *subjective agency*.

Overcoming a naive vision in the use of these resources would help in understanding the interpretative, negotiating and implementational perspective of the actors involved, of the different rationalities that set up the processes and their implications for collective action and the common good.

Although the results of these regulatory process, however useful for research, will always remain more or less explicit and/or accessible also through codes or methodological guidelines. Salais recalls (2016: 119) that most of the negotiation and definition processes remain invisible in all subsequent phases and we tend to forget them as the platform/device becomes autonomous from its creators/programmers. In this way, the tools gradually take on their own strength, they are treated as non-contestable, non-modifiable (for example, because the historical series and/or the reasons for the initial choices that inspired the process are lost, causing inevitable drifts in all subsequent applications). The domestication process (Jedlosky, 2005) does the rest, transforming them into habit, making them forget their invisible prescriptive force, normalizing their presence and use within daily practices.

Some authors (Thévenot 1983; Diaz-Bone 2016/a, 2016/b) have introduced the notion of a *statistical chain* to indicate that different actors and situations are involved in the process that leads to the proliferation of these information resources. The aim is to make sense of how *dating* and *quantification* can lead to invalid data when even one of the actors involved changes rules along the chain. These rules are not only relevant for guaranteeing methodological consistency but, if properly constructed, communicated and observed, they can represent a form of guarantee with respect to the principle of transparency that every democratic society and public office should be made aware of.

A further interpretation key (van Dijck 2014; Gray et al. 2018) shifts attention to the role played by the *infrastructures of* collection, to bring to light their function as an information/communication/strategic coordination *interface* between different sectors (political, bureaucratic, and state circles; economic and managerial circles) and daily epistemic practices. These *infrastructures* are regulated and institutionalized resources that guide and organize the *datification* of social phenomena. An example can be found in school, university, research and third mission evaluation systems. An essential feature of these infrastructures is that they provide pre-configured panels, the so-called dashboards and/or graphic representations, useful for providing a snapshot of reality to inform policy and/or strategic choices. These *frameworks* develop over time, through a negotiation process that involves and mediates between the interests of those who contribute to the design of the digital infra-

structure, representing the outcome of a "building of knowledge" activity (Berger, Luckmann, 1966) which provides premeditated schemes and criteria for the interpretation and evaluation of the data offered. In this way, this system of infrastructures, relatively public and/or accessible, comes to "objectify", to build up as "worlds of data", meaning specific and complex sets of interactions between actors, institutions and practices that they impose themselves on the reading and interpretation of reality and of organizational and social processes [6], exercising a prescriptive and binding force that slowly becomes autonomous.

Faced with this state of affairs, some authors speak of a "crisis" in sociology (Frade, 2016; Savage, Burrows, 2007). It is a discipline that must face a high capacity of measurement, traceability and data collection, and with those who believe that "with sufficient data, the numbers speak for themselves" (Anderson, 2008).

Within this framework, a significant movement of ideas has developed which denounces the fact that: the main institutional context of innovation and methodological competence has moved from academic social sciences to the world of private companies (van Dijck 2014; Diaz-Bone 2020; Zuboff 2015), with evident distortions with respect to the reasons for studying it; the role of quantitative data and information has shifted from informing (national) policy making, and understanding social phenomena and transitions, to using numbers for *benchmarking*; for the individualized "surveillance" and the construction of forecasting models (Zuboff 2015) for the evaluation of productivity; the ownership of data and the responsibility for the definition and development of the research has shifted from the public to the private sector, suggesting a sort of subjection of the political-institutional system to one of a speculative economic-financial and/or technocratic order (Srnicek, 2017); it is necessary to enter this black box that imposes limited knowledge, at times incorrectly and ethically disturbing (Kitchin, 2014; Kitchin, McArdle, 2016); *big data* pose on sociology new methodological challenges to which we cannot escape (DiMaggio, 2015; Lee, Martin, 2015; Marres, Gerlitz, 2016; Williams et al., 2017).

Among the emerging methodologies, for example, we find that of Social Network Analysis (SNA) applied to the measurement of the popularity of a brand is highly developed, starting from the assumption that the Internet Economy is nothing more than the mirror of the real world (Gloor, 2007). Another significant methodology is the Semantic Brand Score (SBS) which measures the importance of a brand based on the analysis of the texts. This methodology combines social network analysis and semantic analysis but can be applied to diversified fields (Fronzetti, Colladon (2018). Furthermore, the integration of social science with computer science and engineering fields has produced a new area of study: computational social science. Although many of the latest techniques for dealing with digital data are in development, there is still no established protocol on how to use this new wealth of *sources* for sociological research.

In this era of transitions and uncertainty, four challenges have been identified by Lupton (2015: 6) for a renewal of sociology in the digital society.

Training digital practice in professions, i.e. training in the responsible and conscious use of digital tools for professional purposes (building networks, learning to manage digital identity and *web reputation*, learning to manage the issue/challenge of data sharing, etc.).

Develop sociological analyses on digital use to understand the ways in which people employ digital technology to shape their sense of self, their embodiment, their social and professional relationships.

[6] This is also seen in Diaz-Bone et al. (2020)

Analysing digital data through mixed approaches capable of overcoming traditional contrapositions, also because some processes and phenomena can be studied and understood only in the context of the best traditions of qualitative research (Garcia, Gluesing, 2013).

Develop a critical sociology of digital information capable, on the one hand, of undertaking a reflective analysis, informed by social and cultural theory, while on the other, capable of examining digital information as part of a broader and more complex ecosystem in order to understand its interactions and repercussions on other segments of the same system such as, for example, the ecological one, in terms of fallout for pollution.

5 Research path ways

5.1 Critical perspective

The adoption of a sociological perspective to the analysis of *big data* requires maintaining a critical outlook both with respect to their information capacity and in relation to their collection and distribution platforms. Any analysis that wants to be based on these sources, regardless of their nature, requires the exercise of lateral thinking, which is pragmatic and not naive, in order to understand their possibilities and limits.

Some scholars (Gehl, 2015) suggest avoiding the uncritical assumption of the results obtained from these survey systems, to examine the sociotechnical processes involved along the "data building chain". An appropriate attention is needed especially when we refer to systems that contribute to their historicization, through real-time and self-powered updates, which guide the system of understanding the world and the lines of action taken by many. As already mentioned, in fact, historicization and comparison represent some of the cornerstones of Weberian sociology (Weber, 1964), aimed at understanding the force of events in the tension between individual motivation and their historicization.

This means *biases*, identifying blind spots, where information is missing. It is also necessary to approach a situational perspective aimed at contextualizing the process of affirmation and evolution of the *data*, and be willing to explore its meaning, thinking creatively about how to deal with this epochal and paradigmatic change in the methodology of social research.

This requires the researcher to have the ability to intervene in the *cleaning* to remove any semantic ambiguity in the information acquired, correct any distortions, clarify the interpretation and try to problematize what remains in the background, what is behind, before and within the so easily accessible *data.* To do this *cleaning* and comprehension work, taking inspiration from one of the most classic explanatory models of the communication process (Lasswell, 1936), and when dealing with these kinds of sources, they should be examined within context:

- *Who* manages the data sources and for *whom*, *who* releases the data? These subjects may in fact not converge and have different interests;
- *What do the data say*, and *what do they say about what is unsaid?*
- Which *means/channel* conveys the data, with which other infrastructures do they communicate or not communicate with, what requirements do they guarantee and/or ignore (security, transparency, interoperability, sharing, transmission, etc.);

- *Who* can access it, *how* and to whom is it priced? What do they have access to and what do not have access to?
- *What effects* are produced, including unexpected, involuntary, unexpected ones?

5.2 The circularity of research

If we assume that sociology is analysis and understanding of social phenomena, and that as social research it is made up of *theory* and *empiricism* that does not see fractures and contrasts, we can recognise the principle of circularity between these two dimensions of sociological action in its effort to connect the particular to the general, and to explain the individual within the perspectives of the universal. Moreover, this effort is emblematically represented by the famous book on *Suicide* (Durkheim, 1897), which many recognize as the cornerstone of sociology. There are different styles of thinking and methods of investigation that can be pursued for this purpose.

Deduction is based on a sequence of statements by means of which it is intended to demonstrate, by means of demonstrative logic and the truth/falsity of an assumption. It is a reasoning based on the condition that can be summarized in the formula "if … then". Within this logic, according to Bruschi (1999: 517) "true premises correspond to true conclusions".

In the case of *inductive logic*, we enter the field of discovery: "the truth of the premises does not necessarily correspond to the truth of the conclusion" (Bruschi, 1999: 549). The strength of inductive logic depends on the accumulation of the observations made. In induction, we proceed by accumulating facts of a homogeneous nature according to a comparative methodology to increase the strength of the argument. The greater the number of comparative facts confirming the hypothesis, the greater the strength of the proposed argument.

The *abductive reasoning* focuses on the interaction between data, *method* and *theory.* In this method of investigation, concepts, theories and methodological skills guide the knowledge discovery process. The data is used to direct further investigation as well as the interpretation process and theoretical development. The *abductive* reasoning involves tools and data in a circular critical process. The focus shifts from demonstrating preconceived hypotheses to understanding how to structure the mountain of available data "into meaningful categories of knowledge" (Goldberg, 2015: 3).

In this way, social research through *big data* comes to act a *symphonic approach* through the conscious and iterative assembly of data, method and theory, description and interpretation, within a spiral process where theory and empiricism, description and interpretation interpenetrate and reinforce each other, creating new knowledge informed by data.

5.3 Interdisciplinarity: threat or resource

Faced with the phenomenon of *datafication* (Van Dijck, 2014) in society, the need to promote across sectors and roles a more mature interpretation of *literacy* and critical understanding of data (*data literacy*) is felt in many areas. As mentioned, however, many initiatives aimed at this purpose focus on developing computational and statistical techniques and skills for working with predefined *datasets*. In view of what has been said so far, it is considered important to encourage a critical discussion around the infrastructures through which data are created, used and shared, the objectives pursued through this activity, the public and

private responsibilities and the ethical and deontological implications affected by these developments. This means promoting cross-fertilization between different epistemic apparatuses where sociology can take on the role of the driving force in the diffusion of a new *data culture*[7], a *sociology of data* capable of informing the *policy* processes, the choices of collective interest at all levels[6] and to lead to a full understanding of what lies *behind*, *in* and *before* the data and, finally, that goes far beyond mere information/access to information content.

5.4 Mixed methods

The recent emergence of digital data as a new source for understanding the *online* world is promising, but analysing it alone is not enough for sociological research intent on exploring broader aspects of society and how behaviours have changed following the emergence of *the digital society* (Lupton, 2015). To analyse this complexity, it is necessary to deal with inhomogeneous data types and establish research projects from an integrated perspective. Mixed methods thus emerge as a valid alternative and opportunity for sociologists. In traditional sociology, quantitative and qualitative methodologies differ in the value of research orientation. As Hamberg summarizes (Hamberg et al. 1994, p. 178), quantitative research focuses on generalizing hypotheses by examining internal validity, while qualitative research focuses on persuading reliability and credibility to provide transferability. The two perspectives are not conflicting and self-excluding even if they are not in a relationship of reciprocal substitution. Up to now, communication between the two methodologies has been lacking due to the division into the "two sub-cultures". The challenge of big data makes it necessary to overcome this gap in favour of research and training pathways capable of promoting the adoption of integrated research methodologies, inspired by the symphonic metaphor, capable of making the variety of tempos, methods and instruments its valuable and distinctive trait. A brilliant example of this combination can be seen in research applications that triangulate different methodologies by using mixed ethnographical and automated data mining to analyse communication networks in global organisations (Gluesing et al., 2014).

According to Creswell and Clark (2017: 5) the principles of the mixed methods approach can be summarized in four essential *steps*. Collect and analyse qualitative and quantitative data rigorously in response to research questions and hypotheses. Integrate, mix or combine the two data forms and their results. Organize these procedures into specific research designs that provide the rationale and procedures for conducting studies. To frame these processes within a precise epistemological and methodological framework.

6 Conclusion

To sum up the reasoning expounded so far, we will try to rethink what the sociology's contribution of can be in the complexity outlined above.

First, there is the question of the uses and usefulness of the data, and the relative access to the sources of their generation, including their costs and ownership. Unlike the engineering

[7] For an in-depth analysis on the subject of data culture, information and infrastructures, see, among others, Bowker et al. (2000); Edwards et al. (2011).

and natural sciences, with their multiple opportunities to collaborate with large private companies, there are few possibilities for academic social research, in particular those related to *education*, to acquire funding to support an infrastructure of dedicated research. The crisis of the public debate in directing research towards elements of collective interest can also be seen in the progressive reduction of funding aimed at supporting research free from measurable economic effects, to the advantage of speculative research, both public and private, aimed at colonizing new data mines.

The issue, therefore, arises of the ability of sociological research to inform *policies* to avoid/reduce the polarization of phenomena that widen previous inequalities and areas of social injustice and, at the same time, guide the development of new social models oriented towards a new humanism and a conscious and sustainable use of technology.

Similarly, few public initiatives and state organizations support the new needs of social research, aimed at building data structures comparable to those offered by digital platforms, as is the case in other highly profitable fields. In the educational field, the proliferation of such infrastructures is driven, on the one hand, by needs of control and centralized *accountability*, where digital supremacy is often *outsourced* to technological consortia and, on the other, by the *digital edu business* market which is tempting to the giants of the web, and which is progressively taking away ground and legitimacy from formal institutions.

This state of affairs produces at least two perverse effects (Boudon, 1977) at different levels. (a) The propensity to change the institutional mandate of those with directional duties (for example, the evaluation agencies of schools, universities, research and Third Missions), showing a growing attention to society as a whole and the institutions involved in the presentation of data, which become tools of legitimation in public forums, functional to attempting to acquire more resources and self-legitimize. (b) The tendency is to influence the time, ways and opportunities of those who carry out social research in this field, to the point of determining the accessible and dominant themes, defining a sort of *research setting* that guides those who investigate in these fields, pre-structuring the very possibilities of doing research and querying the data. And this is because the social researcher, the sociologist, is also subject to the logic of *accountability* that can lead one to study what is available, rather than direct one's gaze on what remains under the surface. In other words, the *platform society* (Srnicek, 2017)) comes to address not only the narrative, but also the very way in which this narrative can be realized. The task of sociology, today more than ever, in particular in the field of *education*, which has always been sensitive to the themes of inequalities and symbolic violence (Bourdieu, 2003), is called upon to propose a counter-narrative on these issues, to counter all forms of exclusion, which are radically modified by the introduction of digital technology and its applications.

Secondly, it must be emphasized that, as in all sectors, this too is confronted with the proliferation of data generated by differentiated sources:

- - public and institutional ones which, while responding to the logic of *open data*, for obvious reasons of *privacy*, do not allow access to *datasets* that can be queried at the level of the information unit, with obvious repercussions on the possibility of statistical sampling and data control and, often, the impossibility to overcome the problem of *dark data*[7] or carry out diachronic and/or targeted research;
- - organizational ones related to the traceability of internal processes, for example, the efficacy of data linked to the incoming, ongoing and outgoing population, or those

connected to the traceability of learning, supported and/or mediated by technology. These data require precise and widespread construction, processing and analysis skills for different professional and/or public/target uses;

- - those deriving from the aforementioned continually expanding *business edu platforms* that are outlining alternative learning channels which are configured as new conquest markets in the *panel* of credit recognition, *micro-credentials*, certifications and skills.

This variety of data leads, at different levels, to the need to forge widespread sociological skills present in a multiplicity of roles that go beyond doing research by profession: school principals, teachers, administrators, academic bodies, coordination figures, evaluation experts, communicators. This requires sociology to rethink its ability to position and train adequate skills, within a plurality of study courses characterized by different learning and professional objectives, to educate subjects capable of exercising an effective capacity for critical analysis and *understanding data* (beyond mere *data literacy*) in diverse contexts and work situations. There is also the need to define how to guide in-service training processes and pathways of social and organizational change within the entire supply chain that is the subject of attention, namely the macro integrated education-training-work system, teaching-research-innovation, vigorously compelled by the digital transition and the threat of sustainability. Furthermore, there is the need to know how to steer the capacity of sociological research to be an agent of change and a significant partner in the social co-construction of data, useful for understanding today's challenges.

Today, more than ever, in the face of the risks associated with complexity, sociology is called upon to provide useful tools for analysing and understanding reality, starting from overcoming the contrast between sociology and applied research to explore new frontiers of investigation that are capable of overcoming contemporary fragmentation through experimentation, fusion and evaluation of the results of one's work. This is possible only through a triangulation between theoretical perspectives, methodologies and tools, in order to enhance the unique sociological contribution in understanding and managing the current transformations.

Decades of ideological confrontation concerning method, quantitative *versus* qualitative, system *versus* actor, object *versus* subject, macro *versus* micro, theory *versus* practice, deduction *versus* induction, academia *versus* practice, have, in fact, betrayed and impoverished the characteristic mission of sociology: interpretation of the social reality and, possibly, helping to understand it. The task today is to push for the dominant view on the use/application of *big data* in the real world, insisting on the transition from predictive and "a-theoretical" "infatuation", and the tendency to use algorithms in an affirmative and a-critical way, to reflection on ethical issues, exploratory research and reflective and critical methodologies that seek to bring to light the hidden side of data. The new "worlds of data", outlined above, question the dominant canon and any reductionist perspective, and sociology has the duty to bring to the fore the social implications that hide behind each number.

Today's challenge asks sociology to rethink itself as a *holistic science* through an opening/blending with the increasingly established field of data *science*, to favour the fusion of perspectives and methods that can help this branch of knowledge come out from the margins of the debate on these issues, to let it glimpse new and interesting possibilities that question the scientific community on an epistemological and methodological level.

Acknowledgements The Author thanks the referees for their valuable suggestions.

Author contributions The author is a single person, there are no other contributions to declare.

Declarations

Conflict of interest and funding The Author have no financial interests to disclose. The author declare that no funds, grants, or other support were received during the preparation of this manuscript.

References

AA.VV.: Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. International Journal of Information Management, 57 (2021) 101994. https://doi.org/10.1016/j.ijinfomgt.2019.08.002

Abbington, B.: Handbook of Computational Sociology. Clanrye International: (2019)

Adorno, T.: Lezioni di sociologia. Einaudi, Torino (1956)

AGID: Agenzia per l'Italia digitale. Programma nazionale per la cultura, la formazione e le competenze digitali, LINEE GUIDA Indicazioni strategiche e operative: (2014)

Alvarez, R.M.: Computational Social Science: Discovery and Prediction. Cambridge University Press, United States of America (2020)

Anderson, C.: The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. Wired, 16(7) (2008): http://archive.wired.com/science/discoveries/magazine/16-07/pb_theory. Accessed 9 November 2021

Boudon, R.: Effects pervers et ordre social. Press Universitaire de France, Paris (1977)

Bourdieu, P., Passeron, J.: Reproduction in Education, Society and Culture. Sage, London (1977)

Bourdieu, P.: Sul potere simbolico. In A. Boschetti (2003), La rivoluzione simbolica di Pierre Bourdieu, Marsilio, Venezia, pp. 119-129 [Sur le pouvoir symbolique, Annales, n. 3, maggio-giugno 1977: 405-411]: (2003)

Bowker, G.C., Star, S.L.: Sorting things out. Classification and its consequences. MIT Press, Cambridge (2000)

Boyd, D., Crawford, K.: Critical questions for big data. Inform. Communication Soc. **15**(5), 662–679 (2012). https://doi.org/10.1177/2053951716631130

Boyd, D., Crawford, K.: Six provocations for big data. Presented at A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society, The Oxford Internet Institute, 21 September: (2011)

Bruschi, A.: Metodologia delle scienze sociali. Bruno Mondadori, Milano (1999)

Capogna, S.: Sociology Between Research and the Society of Data, Sociology Study, Nov.-Dec. 2021, Vol. 11, No. 6, 276-289. doi: https://doi.org/10.17265/2159-5526/2021.06.005

Chen, S.H., Yu, T.: Big Data in Computational Social Sciences and Humanities: An Introduction. Computational Social Sciences: (2018). DOI:https://doi.org/10.1007/978-3-319-95465-3_1

Cheung, M., Jak, S.: Analyzing Big Data in Psychology: A Split/Analyze/Meta-Analyze Approach, Published 23 May 2016. Medicine Frontiers in Psychology. doi:https://doi.org/10.3389/fpsyg.2016.00738

Chin, W.: Technology, war and the state: past, present and future. Int. Affair. **95**(4), 765–783 (2019). doi:https://doi.org/10.1093/ia/iiz106

Cooper, D.: Bigger Theoretical Picture around the Idea of a University Third Mission. J. Appl. Sci. August (2017). https://doi.org/10.1177/1936724417722580

Creswell, J.W., Plano Clark, W.L.: Designing and Conducting Mixed Methods Research. Sage publications, Los Angeles (2017)

Dalton, C.M., Taylor, L., Thatcher, J.: Critical data studies: A dialog on data and space. Big Data & Society. **3**(1), 1–9 (2016). https://doi.org/10.1177/2053951716648346

Davidson, E., Edwards, R., Jamieson, J., Weller, S.: Big data, qualitative style: a breadth–and–depth method for working with large amounts of secondary qualitative data. Qual. Quant. **53**, 363–376 (2019). https://doi.org/10.1007/s11135-018-0757-y

Desrosières, A.: The quantification of the social sciences. An historical comparison. In The social sciences of quantification, ed. Isabelle Bruno, Florence Jany-Catrice & Béatrice Touchelay, 183-204. Cham: Springer: (2016). doi:https://doi.org/10.1007/978-3-319-44000-2_15

Desrosières, A.: Retroaction. How indicators feed back onto quantified actors. In: Rottenburg, R., Merry, S.E., Park, S.-J., Mugler, J. (eds.) The world of indicators, pp. 329–353. Cambridge University Press, Cambridge (2015). doi:https://doi.org/10.1017/CBO9781316091265.013

Diaz-Bone, R., Didier, E.: The sociology of quantification—Perspectives on an emerging field in the social sciences. Hist. Social Re- search. **41**(2), 7–26 (2016/a). doi:https://doi.org/10.12759/hsr.41.2016.2.7-26

Diaz-Bone, R., Didier, E.: Conventions and quantification—Transdisciplinary perspectives on statistics and classifications (special is- sue). Historical Social Research 41 (2) (2016/B): https://www.gesis.org/hsr/archiv/2016/412-conventions-and-quantification Accessed 9 November 2021

Diaz-Bone, R., Horvath, K., Cappel Source, V.: Social Research in Times of Big Data. The Challenges of New Data Worlds and the Need for a Sociology of Social Research. Historical Social Research, Vol. 45, No. 3, Special Issue: 2020: 314-341 Published by: GESIS—Leibniz Institute for the Social Science: (2020). https://doi.org/10.12759/hsr.45.2020.3.314-341

DiMaggio, P.: Adapting computational text analysis to social science (and vice versa). Big Data & Society, July-December: 1-5: (2015). https://doi.org/10.1177/2053951715602908

Dodge, M., Kitchin, R.: Codes of life: Identification codes and the machine-readable world. Environ. Plann. D Soc. Space. **23**(6), 851–881 (2005). https://doi.org/10.1068/d378t

Durkheim, E.: Le regole del metodo sociologico, Torino, Einaudi (1897) (Edizione 2008)

Edwards, P.N., Mayernik, M.S., Btcheller, A.L., Bowker, G.D., Borgman, C.L.: Science Friction: Data, Metadata and Collaboration. Soc. Stud. Sci. **41**(5), 667–690 (2011): https://doi.org/10.1177/0306312711413314

Frade, C.: Social Theory and the politics of big data and method. Sociology. **50**(5), 863–877 (2016). https://doi.org/10.1177/0038038515614186

Franzosi, R.: What's in a text? Bridging the gap between quality and quantity in the digital era. Qual. Quant. **55**, 1513–1540 (2021). https://doi.org/10.1007/s11135-020-01067-6

Fronzetti Colladon, A.: The Semantic Brand Score. J. Bus. Res. **88**, 150–160 (2018). https://doi.org/10.1016/j.jbusres.2018.03.026

Garcia, D., Gluesing, J.C.: Qualitative research methods in international organizational change research. J. Organizational Change Manage. **26 No**(2), 423–444 (2013). https://doi.org/10.1108/09534811311328416

Gehl, R.: Critical reverse engineering: The case of Twitter and Talk Open. In: Langlois, G., Redden, J., Elmer, G. (eds.) Compromised Data: From Social-Media to Big Data. Bloomsbury, London (2015)

Gloor, P.A.: Coolhunting for Trends on the Web, Conference Paper, June: 2007: DOI:https://doi.org/10.1109/CTS.2007.4621731

Gluesing, J.C.: Building connections and balancing power in global teams: toward a reconceptualization of culture as composite. Anthropol. Work Rev. **18**, N° 2–3 (1988) pp. 18–30

Gluesing, J.C., Riopelle, K.R., Danowski, J.A.: Mixing Ethnography and Information Technology Data Mining to Visualize Innovation Networks in Global Networked Organizations. In: Dominguez, S., Hollstein, B. (eds.) Mixed Methods Social Networks Research, pp. 203–234. Cambridge University Press (2014). https://doi.org/10.1017/CBO9781139227193.011

Goldberg, A.: In defence of forensic social science. Big Data & Society, July-December: 1-3: (2015)

Goldthorpe, J.: Sociology as a Population Science. Cambridge University Press, Cambridge (2016)

Gorz, A.: Économie de la connaissance et exploitation des savoirs, e-Cairn.info, 2004/1 no 15: 205-216: (2004): doi:https://doi.org/10.3917/mult.015.0205

Gray, H.: Toward a Sociology of the Trace Edition. NED, University of Minnesota Press (2010)

Gray, J., Gerlitz, C., Bounegru, L.: Data infrastructure literacy. Big Data & Society. **5**(2), 1–13 (2018). https://doi.org/10.1177/2053951718786316

Greco, F., Riopelle, K., Grippa, F., Colladon, F., A., & Gluesing, J.: Linguistic Sleuthing for Innovators. Qual. Quant. **55**(3), 1027–1045 (2020). https://doi.org/10.1007/s11135-020-01038-x

Halford, S., Savage, M.: Speaking Sociologically with Big Data: Symphonic Social Science and the Future for Big Data Res. Sociol., 51(6): 1132–1148 (2017). https://doi.org/10.1177/0038038517698639

Hamberg, K., Johansson, E., Lingren, G., Westman, G.: Scientific Rigour in Qualitative Research: Examples from a Study of Women's Health in Family Practice. Fam. Pract. **11**(2), 176–181 (1994). doi:https://doi.org/10.1093/fampra/11.2.176

Iliadis, A., Russo, F.: Critical data studies: An introduction. Big Data & Society. **3**(2), 1–7 (2016). https://doi.org/10.1177/2053951716674238

Jedlosky, P.: Un giorno dopo l'altro. La vita quotidiana tra esperienza e routine. Milano, Il Mulino (2005)

Kar, A.K., Dwivedi, Y.K.: Theory building with big data-driven research—Moving away from the "What" towards the "Why". Int. J. Inf. Manag. **54**, 102205 (2020). https://doi.org/10.1016/j.ijinfomgt.2020.102205

Kim, Y.: New Opportunities for Sociological Research: A Discussion of the Usefulness of Mixed Methods with Data Science. Journal of Asian Sociology, Vol. 48, No. 3, September: 343-358: (2019). doi https://doi.org/10.21588/jas/2019.48.3.003

King, G.: Preface: Big Data Is Not About The Data! In R. M. Alvarez (Ed.), Computational Social Science: Discovery and Prediction. Cambridge University Press: 1-4): (2016)

Kitchin, R., McArdle, G.: What makes big data big data? Exploring the ontological characteristics of 26 data sets. Big Data & Society January-June: 1–10 (2016). https://doi.org/10.1177/2053951716631130

Kitchin, R.: The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences. Sage, London (2014)

Lasswell, H.D.: Politics: Who gets what, when, how. Whitlesey House, New York (1936)
Lee, M., Martin, J.: Surfeit and surface. Big Data & Society, July-December: 1-3: (2015)
Lewin, K.: Principi di psicologia topologica; traduzione dall'inglese di Adriano Ossicini, Firenze, Organizzazioni Speciali (trad. it.)(1961). Principles of topological psychology: (1936)
Lukacs, G.: Storia e coscienza di classe, Milano, Sugar: (1967)
Lupton, D.: Digital sociology. Routledge, London (2015)
Lyle, K.: The Future of Sociology. Sociology, D.: 2017, Vol. 51, No. 6 December: 1169-1185 (2017). https://doi.org/10.1177/0038038516653728
Marr, B.: Big Data: Using SMART Big Data, Analytics and Metrics To Make Better Decisions and Improve Performance. Wiley, Hoboken (2014).
Marres, N., Gerlitz, C.: Interface methods: Renegotiating the relations between digital social research, STS and the sociology of innovation. Sociol. Rev. **64**, 21–46 (2016)
Marx, K.: Das Kapital. Progress Publishers, Moscow (1867)
Marz, N., Warren, J.: Big Data: Principles and Best Practices of Scalable Realtime Data Systems, MEAP edition. Westhampton, NJ: Manning: (2012)
McKinsey Global Institute: Technologies, jobs and the future of work, McKinsey & Company (2011). http://www.eyesreg.it/2016/big-data-e-scienze-regionali-una-relazione-inesplorata/. Accessed 9 November 2021
McNulty, E.: Understanding Big Data: The seven V's. 22 May (2014). Available at: http://dataconomy.com/seven-vs-big-data/. Aaccessed 4 September 2015
Mead, G.H.: Mind, Self and Society. Henceforth, MSS: (1934)
Mills, C.W. Immaginazione sociologica, Milano, Il Saggiatore: (1959)
Peirce, C.S., Collected Papers, ed. by C., Hartshorne, P., Weiss (vol. 1-6), A. Burks: (voll. 7-8), Harvard University Press, 1931-1958. [Esempio, 6.183: vol. 6, par. 183]
Piketty, T.: Capital in the Twenty-First Century. Harvard University Press, Cambridge (2014)
Popper, K.R.: The logic of scientific discovery. Routledge, London (2002)
Putnam, R.: Bowling Alone: The Collapse and Revival of American Community. Simon and Schuster, New York (2000)
Rawat, K.S., Sood, S.K.: Emerging trends and global scope of big data analytics: a scientometric analysis. Qual. Quant. **55**, 1371–1396 (2021). https://doi.org/10.1007/s11135-020-01061-y
Ruppert, E.: Doing the transparent state: Open government data as performance indicators. In: Rottenburg, R., Merry, S.E., Park, S.J., et al. (eds.) A World of Indicators: The Making of Governmental knowledge through Quantification, pp. 1–18. Cambridge University Press, Cambridge (2015)
Salais, R.: Quantification and objectivity. From statistical conventions to social conventions. Hist. Social Res. **41**(2), 118–134 (2016). doi:https://doi.org/10.12759/hsr.41.2017.1.118-134
Salganik, M.J.: Bit by bit. Social research in the digital age. Princeton University Press, Princeton (2018). https://doi.org/10.1177/0038038507080443
Savage, M., Burrows, R.: The coming crisis of empirical sociology. Sociology. **41**(5), 885–899 (2007). https://doi.org/10.1177/0038038507080443
Schoen, H., Avello, D.G., Metaxas, P.T., Mustafaraj, E., Strohmaier, M., Gloor, P.: The power of prediction with social media. Internet Res. **23 No**(5), 528–543 (2013). doi https://doi.org/10.1108/IntR-06-2013-0115
Smith, R.B.: Linking quality & quantity. Part II. Surveys as formalizations. Qual. Quant. **22**, 3–30 (1988)
Srnicek, N.: Platform Capitalism. Polity Press, Cambridge (2017)
Stein, E.: Il problema dell'empatia, (tr. it. di Costantini, Costantini, E. E.S.), Roma, Studium: (1958)
Strathern, M.: The Tyranny of Transparency. Br. Edu. Res. J. **26**(3 Jun), 309–321 (2000). https://doi.org/10.1080/713651562
Thévenot, L.: L'économie du codage social. Critique de l'économie politique **23/24**, 188–222 (1983)
Uzzi, B.: Social structure and competition in interfirm networks: The paradox of embeddedness. Administrative Science Quarterly, 42: 35–67 (1997)
Van Dijck, J., Poell, T., de Waal, M.: The Platform Society: Public Values in a Connective World, New York, NY, Oxford university Press: (2018)
Van Dijck, J.: Datafication, dataism and dataveillance. Big data between scientific paradigm an ideology. Surveillance & Society 12 (2): 197–208 (2014). doi:https://doi.org/10.24908/ss.v12i2.4776
Weber, M. In Engisch K. Pfister P. Winckelmann J. (a cura di) Weber M. Gedächtnisschrift der Ludwig-Maximillians-Universität München zum 100. Wiederkehr seines Geburttages 1964 Berlin, Duncker und Humblot: (1964)
Weber, M.: (postuma; trad. it. 1961), Economia e società, 2 vol., Milano, Edizioni di Comunità: (1922)
Williams, M., Burnap, P., Sloan, L.: Crime sensing with big data. The affordances and limitation of using open-source communication to estimate crime patterns. Br. J. Criminol. **57**(2), 320–340 (2017). doi:https://doi.org/10.1093/bjc/azw031

Wynn, J.: Digital sociology: emergent technologies in the field and the classroom. Sociol. Forum. **24**(2), 448–456 (2009). doi:https://doi.org/10.1111/j.1573-7861.2009.01109.x

Zhang, Y., Xu, S., Zhang, L., Yang, M.: Big data and human resource management research: An integrative review and new directions for future research. J. Bus. Res. **133**, 34–50 (September 2021). DOI:https://doi.org/10.1016/J.JBUSRES.2021.04.019

Zuboff, S.: Big other: Surveillance capitalism and the prospects of an information civilization. J. Inform. Technol. **30**, 75–89 (2015). https://doi.org/10.1057/jit.2015.5

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.