Culture e Studi del Sociale-CuSSoc
ISSN:2531-3975
Editors-in-Chief
Felice Addeo, Giuseppe Masullo, Giovanna Truda

The Italian perspective on the use of Big Data in Sociological Field: Implications, Empirical Findings and an Impact Analysis on the Discipline

MICHELA CAVAGNUOLO

Come citare / How to cite

CAVAGNUOLO, M. (2022). The Italian perspective on the use of Big Data in Sociological Field: Implications, Empirical Findings and an Impact Analysis on the Discipline. *Culture e Studi del Sociale*, vol. 7(1), 47-65.

Disponibile / Retrieved http://www.cussoc.it/index.php/journal/issue/archive

1. Affiliazione Autore / Authors' information

Department of Communications and social research, Sapienza University of Rome, Italy

2. Contatti / Authors' contact

michela.cavagnuolo@uniroma1.it

Articolo pubblicato online / Article first published online: June 2022

Informazioni aggiuntive / Additional information

Culture e Studi del Sociale

The Italian perspective on the use of Big Data in Sociological Field: Implications, Empirical Findings and an Impact Analysis on the Discipline

Michela Cavagnuolo

Department of Communications and social research, Sapienza University of Rome, Italy

Email: michela.cavagnuolo@uniroma1.it

Abstract

Digital development has invested predominantly in all fields of society: the field of sociology and in general the field of social sciences are not exempt. Starting from these considerations, the social scientist is called to integrate everything within his studies with technological innovations and new tools that modifies and innovates the typical toolbox of the study. This change creates a strong interdisciplinary context. An analysis of the changes and impact of the use of Big Data within the social sciences and how they modify the work of the social scientist therefore becomes necessary.

Keywords: Big Data, Symphonic Social Science, Digital Sociology.

1. Introduction

In recent years, there has been an exponential growth in data production, accompanied by a significant reduction in costs for their management and storage. Due to the perennial collection of data, modern technologies have become an integrated part of all aspects of our daily lives, creating our digital double; which is not an avatar that resembles us in physical appearance, but a dataset with our tastes, preferences, and actions that we do. All this is made possible thanks to computer-mediated communication (Addeo, 2016): «people use blogs, social networks, chats and other online channels on a regular basis in order to express identity, share information, ideas, values, knowledge, and build relationships». Our social life seems to be literally invaded by the Internet, making it difficult to distinguish between «online life» and «offline life» (Pink, 2008).

Owing to these changes, social scientists have to include this new type of knowledge generated online within their studies. In this context, the Internet is not only considered as an «object», but also as a «tool», which sociologists must exploit in order to study these new sources of information available on the Web. Therefore, a context of strong interdiscipline made of sociology, computer science, statistics, but also economics, demography and communication is created; in short, nowadays all those disciplines that somehow wonder about the human behavior have something in common: the digital world (Rogers, 2013; 2019). The main innovation for all social sciences is digitization; the chance to extract and analyze an increasing amount

of data, whether these are *Open or Big Data*[1], allows the scholar to interact with the entire world through a screen.

Since 2004, the widespread of social media has made more evident the inadequacy of an obsolete conception of the Web, considered as a mere theatre of fake identities and deviant behavior, but for the first time we find ourselves having as many available data as are the actions that users can do online (Jurgenson, 2012; Fuchs, 2021). Digital Life, Algorithmic Identity, Digital Footprint, Digital Shadow are just some of the definitions that describes the human activity of the 21st century (McQuade, 2021). Sonia Stefanizzi (2016, basing on a work by Kitchin (2014), underlines a new scientific paradigm and well describes the power of algorithms and computers, defining the Web as the context of the new paradigm of social research.

In the last few years (2014-2020), publications in the field of Big Social Data increased by about 40% (Scopus, 2019) and the interest in this new interdisciplinary field has led to the birth of new Degree Courses, Master's Degrees and Phd Programs.

For many social scientists, the introduction of the analysis of Big Data, changes the way of understanding and «doing» sociology. Due to this fact, it is important to understand what are the motivations that push the different authors to modify or implement the classic strategies of analysis. In the light of these changes, the essay has the following goals: 1. reconstruct the development of the Big Data phenomenon within the Social Sciences – especially Sociology; 2. understand how the introduction of Big Data analysis, impacts, modifies and redefines the boundaries of the scientific community; 3. reconstruct a mapping under a point of view of authors, publications, research centres, networks of relations and conferences.

In order to reconstruct the mapping, this study makes use of the metadata analysis of the main scientific publications in terms of Big Data and, more specifically, the analysis of keywords that through the use of a statistical approach allow extrapolating different application areas, relevant issues and strategies adopted by comparing the use of sociology and other disciplines of social sciences.

2. The Big Data phenomenon in the field of Social Sciences

3.

«Big Data» - also known as BD – is the expression used to describe a data collection so extensive in terms of volume, speed and variety that specific technologies and analytical methods for value extraction are required. The progressive increase in the size of datasets is linked to the need for analysis on a single set of data, with the aim of extracting additional information to those that could be obtained by analyzing small series, with the same total amount of data (Mayer-Schongerger & Cukier, 2017, p. 15-24). 90% of the existing data have been generated in the last 2 years. The rate this information is produced is so high that every two, a volume of data equal to the amount of information generated by the entire humanity until 2003 is created. The birth of the acronym BD is justified by the presence of a new type of data; that is unstructured data requiring a different treatment and name than the previous ones. Since it is difficult to give an intelligible definition, they are often explained using "7 V's"[2], which correspond to seven

[1] Set of data in digital format that are collected, stored and managed through large datasets, not treatable through software and hardware systems traditionally used in the field of social research (Lombi 2015, p. 215).

[2]At the beginning there were just 3 V's: volume, velocity, variety. Then, over time, they have been integrated with four more V's: veracity, variability, visualization, and value (Haslinda Mohd Din, 2017).

ISSN: 2531-3975

questions we can ask about them, with particular reference to the most important first "three V's" (Haslinda Mohd Din, 2017). Mayer-Schonberger and Cukier (2013) have clarified that the acronym BD refers to two main interpretations: the first one (used by the media) simply refers to the crossing of large amounts of data; whereas the second one is that used by computer scientists, mathematicians, sociologists, and researchers who can use them either as a support or as a new tool to explain social phenomena. The latter talk about mega dataset rather than BD.

As it has been explained above, the need to introduce the analysis of BD within social research arises because our actions are mediated by the digital and these are transformed into assemblies of metrical data without users of technologies even realizing it (Burrows, 2012). These digital assemblies generate a kind of Digital Doppelgänger[3].

Our Digital Doppelgänger are born from the intersection of four types of digital archives: data of our economic transactions, our use of social media, our specific online conversations, and the use we make of crowdsourcing platforms such as Wikipedia or Kickstarter. We can talk about digital doppelgänger as informational entities that transcend us and end up overpowering us (Haggerty & Ericson, 2000) or complex algorithmic identities (Cheney & Lippold, 2011; Coulthard, Mallett & Taylor, 2020) that show how machines, in their different components are part of our incarnate processes and our way of being in the world (Haraway, 2015). Today, algorithms work in networks of associations that modulate part of the society; numerical values, subjective ratings, ranking and ratings that are transformed into what has been defined «self-quantification» (Aragona, 2015).

All data which have been produced by third parties should not be underestimated in social research; for example, those created by institutions and that are available to the population: the Open Data. They are a subset of BD but the most important differentiation lies in the purposes and utilization that characterize these two large categorizations of data:

- BD are also collected in order to profile the tastes and trends of citizens without letting the interested party know and are often used for private purposes and market analysis;
- Open Data are public data collected as part of the action of Public Administrations; they must be available, reusable, and are available to the community to encourage participation in the management of public affairs (ADI[4]).

Starting from these assumptions, to understand the importance of these data, it is noteworthy the statement made by Bossewitch and Sinnreich (2016, p. 226): «Nowadays digital technologies act as cognitive prosthesis, because their traces extend, increase and even replace memories». Therefore, sociologists cannot escape this new challenge to introduce the use of BD in social research with all its criticisms and limitations, which have been widely debated in recent year.

3.1 The birth of Big Social Data Analysis

Within the sociological field – basing on Putnam (2000), Wilkinson, Pickett (2009) and Piketty (2014) argumentative modes - Halford and Savage (2017) labelled the integration of the use of BD as "symphonic social science". Symphonic

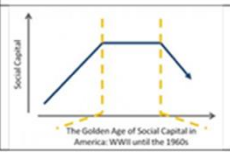
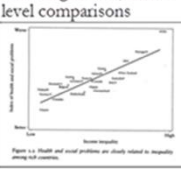
[3] Digital Double.

[4] (https://www.agendadigitale.eu/dati/open-data-e-big-data/) Website visited on 28/05/19

social science classifies the «BD phenomenon» both as an empirical phenomenon and an emerging field of practice where affirmations of knowledge of the social world are made (Halford *et al.*, 2014; Tinati *et al.*, 2014; Halford 2015; Savage & Burrows 2007; 2009; Savage 2013; 2014). BD provide information about the daily life of millions of people and in real-time. Interest in these new data can be found in studies such as *Bowling Alone* (2000) by Robert Putnam, *The Spirit Level* (2009) by Richard Wilkinson and Kate Pickett, *Capital* (2014) by Thomas Piketty. Although none of these authors uses BD terminology, their innovative approach is a form of data collection and analysis. Halford and Savage, 2017 suggest that the books mentioned above "establish a new mode of reasoning that reconfigures the relationship between data, method and theory so that it brings both surprising similarities and key differences compared to BD analysis groups". These important pioneering studies contain different topics and have different disciplinary backgrounds: none of them come from sociology but each one shaped many sociological debates. However, our interest focuses on the fact that they embrace different arguments, different analytical strategies, and different style of argumentation so that they could be a way of thinking to put at the base of the use of BD in social research strategies.

Figure 1. (Halford & Savage, 2017) shows the significant points of the books taken into consideration focusing on datasets used, different techniques, and models of visualization. This work is an empirical evidence that actually sociologists have been using BD for a long time before realizing it; the only things that have changed today are the awareness and the speed with which they are used.

FIGURE 1. *Symphonic social science and the future for Big Data research*

	Putnam (2000) *Bowling Alone*	Wilkinson & Pickett (2009) *The Spirit Level*	Piketty (2014) *Capital*
Data	US Census, surveys of social and political trends, membership data, Gallup polls, etc.	National survey data, registration data, ethnographic data.	Multiple and diverse taxation records from 1700-2010, registration data.
Methods	Descriptive statistics, bi-variate frequencies	Descriptive statistics, linear regression, macro level comparisons	Descriptive statistics, frequencies over time.
Visualisation			

Source: Halford e Savage, 2017

Each databook mentioned above distributes heterogeneous numerical assemblages on a large scale reintroducing the results of numerous sources, rather than dedicated sources such as a national representative sample or a case study. Among different sources, *Bowling Alone* uses the US census, surveys on social and political trends, data relating to the membership of 40 organizations, Social Survey and Gallup[5] surveys. Similarly, Wilkinson and Pickett in *The Spirit Level* compare

[5] US Public Opinion Poll Agency (American Institute of Public Opinion, also known as Gallup Poll), based in Princeton, New Jersey, founded in 1935 by George H. Gallup (Jefferson 1901-Tschingel, Switzerland, 1984), statistician and professor of journalism. Mostly known for its surveys of presidential campaigns, it also conducts numerous economic, sociological and psychological investigations. Then provides information throughout various publications (books, reports) (www.sapere.it/enciclopedia/Gallup).

ISSN: 2531-3975

different types of national data sources, including not only surveys, but also ethnographic data and patent registrations. On the other hand, in *Capital* Piketty criticizes sample surveys and instead uses extensive data from the *World Incomes Database*[6], which meticulously collects tax data from numerous nations to show long-term trends in income inequality and wealth emphasizing in particular that the last decades have seen a shift towards a concentration of income and wealth. In addition to this study, Halford and Savage (2017) claim that "social scientists have been the first people who started to show how new data assemblies can be deployed to formulate powerful arguments on social life and social changes that shape academic and public debates".

Savage (2017) points out that each book taken into consideration is based on "repeated refrains", just like classical music symphonies introduce and return to recurring themes with slight modifications. Hence, while conventional social sciences focus on formal models, often trying to predict the results of specific dependent variables, *symphonic social science* draws on a more aesthetic repertoire. In this case what truly matters is "prolixity" with intelligent and subtle repetition of examples belonging to the same kind of relationship, rather than "thrift" promoted by traditional social sciences.

Symphonic social science is based on three assumptions:

- re-proposition of multiple and various data sources;
- emphasis on correlation;
- use of visualization.

These assumptions not only break with conventional social science approaches but also are the hallmarks of BD analysis (Mayer-Schönberger & Cukier, 2013; Kitchin, 2014; O'Neill & Schutt, 2014). Symphonic social science could be what the social sciences were waiting for to make the best use of BD analysis. *Symphonic authors* (Savage, 2014) took years to analyze all projects. This type of approach involves the pursuit of more complex research questions and an analysis of more data flows, easier longitudinal researches to implement and the use of large and different datasets. Using a symphonic approach means having a new view of data, method and theory. Now more than ever, we must not forget that data need to be observed critically, that combinations of different strategies involves the use of a pluralist methodology (Savage, 2017), and that a symphonic approach to BD requires abdicative reasoning which focuses on the interaction of co-producing data, method, and theory (Fotopoulou, 2021).

The next paragraph will be about the birth of BD and their worldwide interest, focusing on the existing nomenclatures that seem to be connected to them.

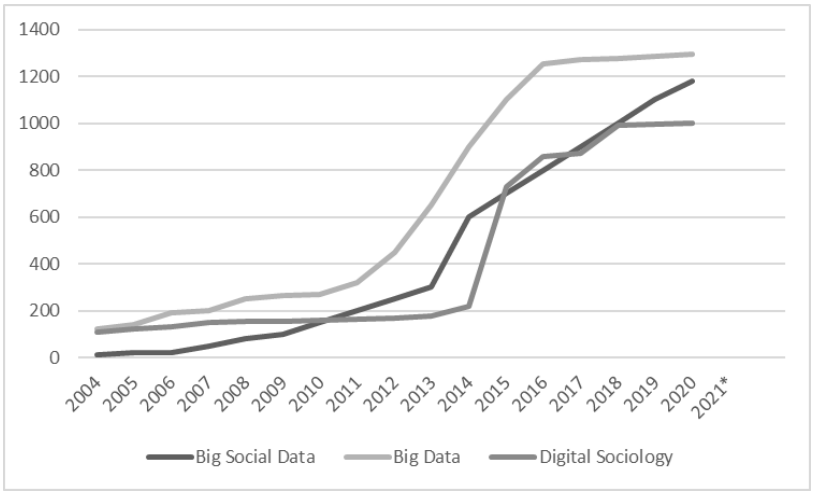
3.2 A Scraping on the development of international interest

Having outlined how BD are used within the discipline, the next step is to understand how and when the interest in Big Data, Big Social Data and Digital Sociology develops from an international point of view. Graph 1. is the result of a

[6] Open and accessible database on the historical evolution of the world distribution of income and wealth (https://wid.world/)

revision of re-elaborated raw data extracted thanks to Google Trends[7]. The data on the Big Data trends, Social BD and Digital Sociology have been extrapolated and reworked separately and then a comparison was made. Observing the curves, it is clear that the interest in Big Data increases between 2011 and 2012; whereas the interest in Big Social Data (term chosen to differentiate and specify the use of the latter within the social research) and Digital Sociology increases between 2013 and 2014. The comparison of the curves made possible the space-time identification of the following analysis: (2014-2020).

GRAPH 1. Interest in Big Data, Big Social Data and Digital Sociology



Source: Google Trends 18/05/2021

3.3 Methodological note

After having reconstructed through the reference literature the creation and development of the use of Big Data in the field of social sciences and having observed through Google Trends how the "world's interest" has shifted through the years, the next objective is to reconstruct a mapping under the point of view of authors, publications, research centers, networks of relations and conferences through an analysis of the metadata in the most relevant scientific publications.

For the analysis of metadata, which follows a standard/quantitative approach, it has been decided to work on an integrated database built with the papers extracted from the databases: Scopus, Web of Science, and Journals that mostly dealt with the "Big Data" phenomenon: Journal of Sociology (AIS), Journal of Big Data, Rassegna Italiana di Valutazione (RIV), Journal "Sociologia e Ricerca Sociale". The decisions of adding some journals to the database extracted from Scopus and Web of Science derives from personal experience and from the main readings that have compelled the writers to create such mapping. The database has been constructed in various steps: I: The first step included the observation of all the filters that the platforms Scopus and Web of Science use to reduce the number of articles useful in achieving the objective of the researcher. II: After having identified all the filters useful to the current work, it was decided to use the Scopus platform to extract all articles that met the following criteria: the term "Big Data" in the keywords; networks "Italian"

[7] Google Trends is a search trends feature that shows how frequently a given search term is entered into Google's search engine. This tool is for web scraping and enables the collection of data available online about the words in analysis.

(which enabled the extraction of not only Italian authors but also articles in which at least there is an Italian co-author; field "Social Science"; language "English and Italian"; Range "2014 -2020". III: scraping of articles taking into account the same filters on Web of Science database (use of the same filters plus filter "sociology", see § 3.5), n.70 extracted articles. IV: addition of extrapolated articles from the identified Journals. VI: final database construction (n. 260 of which 70 linked to the field of sociology).

As will be shown in the coming paragraphs, all the analyses that follow are the result of personal processing on the database.

4. A mapping of empirical evidence: from the social science scenario to the focus on sociological discipline

4.1 Publications

Looking at publication dates of all the articles, on the database it appears that 65,8 % of them were published in the range indicated by Google Trends analysis (2014-2020). From an international point of view, Computer Science and Engineering are the fields where BD are mostly used; the field of study of social sciences is placed in the third place for use with a total of 54.594 articles out of 321.905. Then, the fields of medicine, mathematics, agriculture, psychology and chemistry.

On the other hand, the analyses that follow are the result of integrated databases built on the extraction criterion mentioned that takes into consideration mainly Italy and collaborations between italian authors and international authors[8] (n. 190 for the field of social sciences and n.70 for the specific field of sociology).

FIGURE 2. Keywords Theme Cloud of the analysed in the field of social sciences

Source: Our database elaboration; Excel

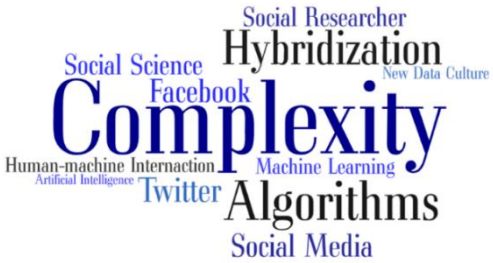


TABLE 1. Frequency of keywords articles

Social Science Keywords	Freq.	%
Complexity	142	0,15
Algorithms	124	0,13
Machine Learning	114	0,12
AI	105	0,11
Hybridization	95	0,10
Human-machine Interaction	76	0,08
Facebook	47	0,05
Twitter	38	0,04
New Data Culture	29	0,03
Social Science	28	0,03
Social Researcher	19	0,02
Other	133	0,14
Total	950	1

Before introducing the authors, it was decided to do an analysis of the keywords that the authors themselves have inserted to describe their work when the contributions were published. Following a lexical approach, it is possible to understand what the topics are. Figure 2. shows that, in addition to BD keywords, the field of social sciences mainly refers to the concept of complexity, algorithms, the world of digital with social media, human-machine interaction (Roth & Luczak-Roesch, 2020) but also to more specific concepts such as "hybrid", a new culture and artificial intelligence theme that is increasingly linked to the use of BD in different disciplines. Therefore, in the field of social sciences, the main reference is to the sociotechnical apparatus and the technical analysis of the use of BD rather than empirical fields of use and social repercussions.

3.2 Authors

The analysis shows that there are 563 authors of which 62% are only italians, while the remaining 38% have worked with international scholars.

Among the most significant italian authors stand out Davide Bennato, Gabriella Punziano, Paolo Parra Saiani, Biagio Aragona, Felice Addeo, and Fabrizio Martire. In particular, Davide Bennato (2016) studied BD in relation to the birth and development of computational science, up to the interest in Digital Sociology. Gabriella Punziano not only focused on the analysis of online texts (2017) but also on the birth and development of Data Science as a new branch of sociology. Over the years, Paolo Parra Saiani has worked in order to understand the advantages and disadvantages of using BD in social sciences (2016). Felice Addeo (2019) focused on Bds, but also on Digital Sociology with reference to *Online Social Research*. On the other hand, Biagio Aragona (2018) studied the integration of institutional databases with online ones, describing this union as BD. He also dealt with new strategies to insert BD into the evaluation research. Finally, Fabrizio Martire (2018) focused on *Food Print* analyses, emphasizing that BD analysis is not intrusive. He later pointed out the possibility – thanks to BD - of integrating social sciences and biological sciences into *biosocial surveys*.[9].

Among the most important international collaborations, some working groups can be seen: the group formed by Antoniou C., Barceló J., Breen M., Bullejos M., Casas J., Cipriani E., Ciuffo B., Djukic T., Hoogendoorn S., Marzano V., Montero L., Nigro M., Perarnau J., Punzo V., Toledo T., van Lint H, which in a volume entitled "*A Utility-based Dynamic Estimation Model that Explicitly Accounts for Activity Scheduling and Duration"* published in 2017, studied the behavior of users through an estimation model based on the actions of users themselves, all based on a simulation model; the group formed by Caputo F., Walletzky L., Štepánek P, which dealt with the concepts of *smart community* (2016) and sociology of health (2015); and the group formed by Ayma V.A., Ferreira R.S., Happ P., Oliveira D., Feitosa R.,

9: For further details, see the full contributions of authors: (Addeo, F., (2019), Doing Social Research on Online Communities: The Benefits of Netnography, in Athens journal of social sciences. vol. 7. p. 9-38; Aragona, B, (2017), Types of Big Data and designs of evaluation research, in Rassegna italiana di valutazione, 68(2):48-62.; -,(2016), Big Data or data that are getting bigger?, in Sociologia e ricerca sociale, 109(3): 42-53; Bennato, D. (2008), Le metafore del computer. La costruzione sociale dell'informatica, Roma, Mantelmi; -, (2015), Il computer come macroscopio, Big Data e approccio computazionale per comprendere i cambiamenti sociali, Roma, Franco Angeli; Martire, F & Pitrone, M.C., (2016), Lo studio dell'opinione pubblica al tempo dei big data.Una sfida per la ricerca sociale, in Sociologia e ricerca sociale, 109(3). 102-115; Mazzaro, F & Punziano, G., Online Textual Data and Political communication analysis, in AIS, Journal of Sociology: Aprile 2017 (11): 143-158; Parra Sainai, P. (2016), Le risorse e il controllo. I big data oltre il mito, in Sociologia e ricerca sociale, 109(3): 28- 41).

Costa G., Plaza A., Gamba P, which analysed both Artificial Intelligence (AI) and Data Science trying to create an interdisciplinary mash-up between computer science, social sciences and mathematical methods.

The authors mentioned above make a step forward from a methodological point of view: due to their studies there is a shift of interest from mixed methods approaches to multidimensional studies, where each discipline can help to the knowledge formation of the phenomenon taken in analysis. All this is possible thanks to the collaboration between authors, universities, and institutes (see also §3.3, 3.4) we are perhaps witnessing the era of a second "new data culture" Aragona (2008). The author, reworking on the thought of Sgritta (1988) describes the social, technological, and institutional[10] path that statistical system has crossed until its current institutionalization. Today, it looks like that the same path is being crossed by new types of data available. As for the national statistical system - wanted and organized by various disciplines such as economy, jurisprudence, and administrative sector - the meeting between different disciplines and the application of BD analysis to the study of various sectors of society can be a strong incentive to create what we could call as the *second new data culture*; justified both by interdisciplinary collaborations of the authors mentioned above, and specifically by the networks of relations that have been established thanks to this phenomenon. The collaboration between disciplines, institutions and authors -both nationally and internationally- can only strengthen the already predominant BD phenomenon.

Parra Saiani (2016) point of view is also interesting: he focused on how cuts to research funding (specifically on cuts to social sciences fundings in favor of hard sciences), can see in BD a new opportunity for social research, which goes beyond the use of them to study the new social complexity in the era of Internet and social networks, but as a mean of responding to the shortfall in financial resources that cuts produce. Again, collaboration seems to be the answer: an extracted database through an online scraping technique can not only be used -as the traditional statistical databases- by several scholars but can also contribute to the integrated study of a hypothetical analysed phenomenon. It is the case of Fabrizio Martire (2016) study, which focuses on public opinion: the same database could be used to make specific lifestyle analyses or could be analyzed to design a new campaign from a political marketing point of view, but also for an initial market analysis.

Although it is difficult to separate authors, networks, and research centres for more detailed information, the main research centres and groups working on BD are as follows.

3.3 Research centres

In order to create a mapping of the research centres, it has been necessary to start from the universities that are working to integrate the study and use of BD into their curricula. Figure 3. shows that the Italian regions that host universities dealing with BD are mainly Lombardy and Lazio. In fact, these regions are also connected to each

[10] See: Legislative Decree no. 322 of 6 September 1989, enacted in implementation of the directive contained in Article 24, Law 400/88, with which Sistema statistico nazionale (Sistan) – National Statistical System was established; reform introduced by Law 81/93, on the company's administrative management and the formal recognition of some profiles belonging to the Third Sector (Pavolini, 2003); Decree no. 322/89 which, together with welfare state reform, strengthened the link between production of statistical data and their practical use in the definition of policies and the study of social phenomena in the territories (Aragona, 2008, p. 161).

other thanks to conferences and publications. Regarding Lombardy, Bicocca and Cattolica universities are the most involved ones: in recent years, they have been working with the Department of Communication and Social Research of Sapienza University of Rome. Regarding Lazio, beside Sapienza, recently Rome Tre University has carried out an election campaign monitoring for the European elections of 26 May 2019, building an integrated online database thanks to the project "European election campaign 2019" where it was possible to monitor and archive the election campaigns of the leaders of all 27 countries; this is an example that the use of these large databases is applied in a wide range of fields research. However, it must be taken into account the statistical over-representation of Lombardy and Lazio, as they host a larger number of universities. In fact, while observing in detail Figure 3. it can be seen that almost all Italian universities are actually involved in these types of studies. Considering other regions, the University of Catania with a research group, led by Professor Davide Bennato, is one of the excellences in this field because for years it has been working with computational social sciences, web analysis and BD.

On the other hand, regarding the international scene, American universities turn out to be the most involved and for a longer period of time. This study starts from a theoretical consideration on symphonic sociology by Halford and Savage (Stanford). Nevertheless, also other universities stand out on this topic lately, such as King's College London that is dealing with BD on several fronts: from Marketing to privacy.

Concerning research centres, one of the most interesting is Foundation Bruno Kessler (based in Trento) that in 2016 started to collaborate on BD studies with MIT (*Massachusetts Institute of Technology*). The Foundation is also popular for its involvement in healthcare and for the formation of the *Data Scientist*, a new professional. In recent years, also Gran Sasso Institute played its part in the field of social sciences.

Relationship networks established during the range of the analysis (2014-2020) will be shown below.

3.4 Research Groups and Relationship Networks established

It was decided to start from the Miur[11] mapping study published in June 2016 so as to analyze the research groups and relationship networks that have established over time.

Then, Miur identified three worldwide research groups that also constitute an important network (one is italian): *EuroTech Universities:* Technical University of Denmark (DTU), École Polytechnique Fédérale de Lausanne (EPFL), École Polytechnique (L'X), Il Technion, Eindhoven University of Technology (TU/e), Technical University of Munich (TUM).

[11]http://www.istruzione.it/allegati/2016/bigdata.pdf). Aim of the study: mapping reconstruction of the universities and institutes that deal with Big Data to analyze the progress of the phenomenon.

TABLE 2. *Mapping of Italian Universities dealing with Big Data*

Italian Universities using Big Data*	**Region**
Universities of Milan; Pavia, Bergamo; Brescia and Bicocca University	Lombardy
Sapienza, Roma Tre, LUISS, LUMSA and LINK Universities	Lazio
Federico II and Salerno Universities.	Campania
Universities of Florence and Pisa; Normale University	Tuscany
Alma Mater Studiorum University and Modena e Reggio Emilia Universities.	Emilia-Romagna
Padova and Verona Universities	Veneto
Turin University.	Piedmont
Salento University.	Puglia
University of Catania and Enna Kore University	Sicily
University of Calabria	Calabria
Basilicata University	Basilicata
Genoa University	Liguria
Gabriele d'Annunzio University	Abruzzo
Carlo Bo University	Marche
Perugia University	Umbria
Sassari University	Sardinia
Total: 29 Universities	

FIGURE 3. Map of universities by region

Source: Our database elaboration; Excel

This network currently involves several universities that aim to implement their analysis on BD and face the challenges of the digitalization of society. The network mainly deals with information technology techniques and engineering, but also healthcare, smart city, innovation, and artificial intelligence.

European Institute of Innovation & Technology (EIT): Polytechnic University of Milan, Technische Universiteit Eindhoven, Universidad Politecnica de Madrid, Université Nice Sophia Antipolis, Technische Universität Berlin e KTH Royal Institute of Technology.

The EIT is an EU body created by the European Union in 2008 to strengthen Europe's ability to innovate. It is an integral part of Horizon 2020, the EU's Framework Programme for Research and Innovation and uses BD-based research techniques in many of his studies.

Michigan Data Science: Institute which mainly deals with man-machine relationship. It includes several universities of the United States. According to the MIUR report (2016), the United States ranks first in the world for interdisciplinary investments.

The relations emerged from metadata analysis are as follows.

3.4.1 Networks emerged from the database

The analyses revealed 37 relationship networks based on 56 universities and 24 research institutes involved: 17 networks among universities alone, 18 networks between universities and institutes, and just 2 networks among institutes alone.

One of the most important networks is formed by the Department of Sociology and Social Research of the University of Trento and Scuola Normale Superiore di Pisa (Mattoni & Pavan, 2018). They are dealing with political activism through BD and creating a clear net between political studies and social media. On the other hand, the network formed by the University of Cassino and the Link Campus University of Rome is studying digitalization of society so much that new single courses in digital sociology have been added to study programs. For the academic year 2019/2020, the Department of Social Sciences of University of Naples Federico II finally introduced a Master's Degree called Digital Sociology and Web Analysis. The most interesting indicator in this scenario is represented by the evolving interest towards higher education courses. In fact, among the networks which emerged in the analysis it is interesting to note that, out of four PhDs [12] in the field of BD, in addition to the Gran Sasso Science Institute[13] course, a new PhD course was inaugurated in 2016 thanks to the collaboration between the IMT School for Adanced Studies Lucca, the Scuola Normale Superiore, the Sant'Anna School of Advanced Studies, the University of Pisa, and the National Research Council- CNR.

Finally, regarding the relationship between universities and research institutes, it is very interesting to note how contributions from European centres can be the source of research funding. The universities that focused on the use of BD in social research are mainly Federico II of Naples, Sapienza University, and the Universities of Florence and Catania, which cooperate with several research centres such as ISTAT, CNR, ANVUR, CENSIS.

Also in this case, as for the authors, the collaboration between universities and institutes is a warning from the points of view of interdisciplinary, finance, and methodology. For example, consider bio-surveys and new ideas that might arise thanks to interdisciplinary. Starting from these considerations, it was decided to focus on sociology in order to analyse the impact of these new data on discipline.

3.5 Specifications on the sociological discipline

As mentioned in the previous paragraphs, it was decided to focus on the sociological discipline. After having analysed 70 articles, comparing their topics, research questions and conclusions, it can be said that real conclusions do not exist neither at a national nor international level. Despite working on the "BD phenomenon" for a decade, in Italy there are still some scholars who do not agree on their use within the sociological field, whereas abroad, other scholars are open minded on their use. However, it is interesting to note that in conferences reports there is a high number of studies that have used BD, making redundant the discussion about their use. They have been used for a wide range of studies: from politics to the quality of life, but also social nets, emergency communications[14], specific social

[12] The other PhD interdisciplinary courses that are not focused on social research are those of the Department of Mathematics in the Universities of Ferrara and Insubria, and the Department of ICT in the University of Milan.

[13] An international PhD school and a center for advanced studies in physics, mathematics, computer science and social sciences.

[14] The emergency communication system and the analysis of interdisciplinary phenomena made with Big Data - is a very important contribution in the period in which work takes shape. In fact, a group of

networks. Moreover, on interdisciplinary studies, a widely discussed topic both in the last international conference of 2019 "Data Science and Social Research", and in the XII National Congress AIS "*Sociologia in Dialogo: algoritmo, cervello, valutazione"* (Sociology in Dialogue: algorithm, brain, evaluation) held last January in Naples attended by scholars from different disciplines.

The theme cloud n.2 (Figure 4) shows the most relevant topics discussed in conferences held in Italy regarding the general field of social sciences. As it can be seen, the areas of application are very different: it is interesting to note how BD combine to public administration, healthcare, tourism and Smart Cities. The greatest innovation is to see words such as responsibilities, electoral forecasts, dialogue between university and industry because it suggests a departure from the statistical determinism with which BD are used in favour of sociological analysis. The authors reasoned on the advantages and disadvantages of using these data within the sociological discipline by testing them, using BD and evaluating their performance. All the extracted articles, especially review conference ones, are full of empirical evidence about their use.

Then, going into detail of the Cloud, it has been decided to collect the main aspects of the analyzed articles and review conferences, which, once again allowed to focus on the study and the use of BD in terms of interdiscipline and multi-method. Making a quick comparison between Key Words used in social science authors' contributions and those of the specific field of sociology, it is interesting to note that in the sociological field the areas of application and the future of BD are in detail. Some examples also present in international studies: Big Data (Sandvik, 2020); Responsibility (Lane, 2020; Liu & Chen, 2021) Public Administration; Smart City (Meerstra-de Haan, Haartsen, Meier & Strijker, 2020; Willig, 2022); Tourism; Health; Facebook (Fussey & Roth, 2020); Twitter; Open Data (Westermeier, 2020); Electoral Predictions (Chen, Wu, Hu, He & Ju, 2021).

One immediately realizes that in the field of social sciences there is more reference to the sociotechnical apparatus of the BD use rather than to its empirical repercussions. One immediately realizes that in the field of social sciences there is a reference to the sociotechnical apparatus of BD use rather than its empirical repercussions. Hence, under this point of view, sociology is – as symphonic authors say – one step beyond compared to other disciplines and context of studies in which the phenomenon in analysis has been studied in several ways.

researchers at Johns Hopkins University created a georeferenced map of the spread of the Chinese virus COVID-19 also called Coronavirus, which shows the confirmed cases in real time and country by country, extracted from reliable sources such as WHO. In addition, some researchers at the Johns Hopkins Whiting School of Engineering have created an online dashboard that shows the daily progress of coronavirus infections globally. Website visited on 23/02/2020) https://gisanddata.maps.arcgis.com/apps/opsdashboard/index.h. A careful monitoring together with good communications can somehow help to keep under control the spread of the virus and keep citizens around the world constantly updated. Then, these data could be cross-referenced with other data of different studies. For example, they could be cross-referenced with social network data to study the spread of the virus and the reactions of users or with data that allow to study how media have addressed the problem to understand the level of alarmism transmitted by the extent of the phenomenon.

FIGURE 4. *Keywords Theme Cloud of sociological field*

Source: Our database elaboration; Excel

TABLE 3. Frequency of keywords *the articles in*

Sociology Keywords	Freq	%
Big Data	42	0,12
Responsibility	35	0,10
Public administration	29	0,08
Smart City	28	0,08
Tourism	29	0,08
Health	15	0,04
Facebook	18	0,05
Twitter	17	0,05
Open Data	14	0,04
Electoral Predictions	11	0,03
University & I D	7	0,02
Missing	105	0,30
Total	350	1

For these reasons, conclusions are considered as such just on formal grounds: actually, they represent the current level of existing art and scientific literature and an attempt to highlight the most interesting changes identified in the analysis of the works.

5. Conclusions

At the end of this study it is not difficult to recognize that the rapid growth and availability of digital data have led to a new development of social sciences and, more specifically, Sociology. The phenomenon of dating or quantification (Espeland & Stevens, 2008) allows to use data directly produced by social actors - like those extrapolated from social media – with a timing that never before would have been imagined. In this context, we are assisting to a new era of sociology (Brensinger & Eyal, 2021).

After analyzing the genesis of the phenomenon, the pioneering studies, the international interest and reconstructing a mapping of the use of BD in social sciences and sociology, the final aim of this study is to understand how the BD phenomenon had an impact on the scientific community. In order to do that, some key points taken from the contributions that constitute the impact of the utilization of BD in the sociological field will be used:

Birth of new branches of sociology

Digital Sociology. *Digital Sociology* is a "sub-discipline of sociology that focuses on understanding the use of digital media as part of everyday life, and how these various technologies contribute to patterns of human behavior, social relationships, and concepts of the self". The first scholarly article is "Digital Sociology" by Jonathan R. Wynn (2009). The author reflects on the ways in which digital technologies may influence not only daily lives but also sociological research, teaching, and sociologists. In 2012, the British Sociological Association approved a new study group in digital sociology; Goldsmiths, University of London, offers the first master's degree in digital sociology; in 2013, following the conference of

the Australian Sociological Association the first book about Digital Sociology was published (Orton-Johnson & Prior, 2013); in 2015, there has been the first academic conference dedicated to this area of studies at New York University. Over time - underlines Deborah Lupton (2015, p. 43), sociologists have labelled media study in different ways: from *Cybersociology* to *Internet Sociology*, *Sociology of New Media*, or *Sociology*, *Sociology of digital media*. The peculiarity of Digital Sociology can be found in the greatest range of interests: it deals with cyberculture, the role of the media, the use of digital devices and the techno-social world. Therefore, it is a way of re-practising sociology; it is a kind of sociology that studies the challenge of digitalization, which recognizes the epochal turning point of the way sociology can be done. One of the easiest ways to explain what this is about is to think about the fact that it is no longer a study of *the* digital but a study *through* the digital itself, in all its forms. In short, putting in place the famous sociological imagination at the dawn of the third millennium (Santoro, 2018).

Toolbox Implementations [some examples]

Data mining. The term *data mining* refers to the procedure of extracting a part of the data (that have been chosen to analyze) from large datasets. In this case, it is important to emphasize that this procedure is useful within social research to build, for example, classification trees, study neural networks, use a cluster analysis technique on large datasets, for different types of regression, for association techniques, but also analyses in main components, factorial analyses, within SPSS, a software used in our field, or in the emerging field of social simulation. The innovation lies in the fact that this procedure manages to make researchers work on much larger numbers of cases even keeping on using classic techniques of social research (Zani & Cerioli, 2015; Kotras, 2020).

GeoMapping. One of the most interesting contributions on this technique is by Elena Battaglini (2015), who, in the fields of evaluation research and territorial development, focuses on the use of this technique for the analysis of the territorial system, the descriptive techniques of organization of the territory, but also for the analysis of territorial assets and ecological data. In this context, the maps show different types of information; but according to the author, the power of images is very important because it can help researchers to associate map's evidence and other types of information.

Scraping. Web Scraping is used for extracting data and information from websites thanks to automated processes. Unlike data mining, the main goal of this technique is to extract all available information, both textual and numerical. It is possible to use this technique within social research in different ways; for example, for studying all the online contents of an unknown topic when the research style is still uncertain, or for using other techniques in sequence, such as online *Sentiment Analysis* on the phenomenon (Marres, 2013).

Hashtags Analysis. Hashtags analysis on social media may have different functions. First and foremost, this kind of analysis can be both quantitative and qualitative. Statistic for recurring words can be analyzed quantitatively as well as some interpretations can be extracted from these words to understand the contexts meant by users who published the hashtag.

Netnography/cyber-etnography. In recent years, thanks to netnography, anthropology has experienced a kind of internal revolution; in fact, it is defined as the discipline that studies relational dynamics, opinions, habits, digital communities. The term was coined by sociologist Robert Kozinets in 2010 to denote the use of naturalistic observation techniques on the web. This word is a neologism born by combining the words ethnography and internet and has different fields of application: from Tribal Marketing to digital ethnography studies, as in this case (Addeo, 2019).

Web Content Analysis. Web Content Analysis, as an analytical perspective, showed the potential for developing a new type of content analysis. From the development of the Internet, statistical techniques and other types of quantitative analyses, have dominated research for years. Thanks to the development of this new type of content analysis, qualitative techniques are regaining power. The subject to analysis are both different socio-cultural contents and simple online texts.

New professionals

Data Scientist. The Data Scientist is a new professional role. An expert in statistics, computer science and mathematics, who thanks to its technical skills helps to solve problems that underlie disciplines such as sociology, marketing, biology. Also, in this case, it is important the interdisciplinarity concept, which is about the transversal skills that anyone should have once joined such a complex team. The feature of this professional is to have a strong technical knowledge and adequate theoretical formation regarding to the field in which he/she operates. The research groups must produce a continuous dialogue between technical and sociological theory in order to avoid a statistical/computerised micro-reductionism.

A new kind of relationship with theory

One of the first topics to consider when discussing the theory/method debate is the fact that, with BD, this becomes a hybrid relationship. The datasets used in the research, are not free from theoretical references because they are the result of a construction process; most of all, we must keep in mind that use and production are co-produced. On the one hand, the production process affects utilization but on the other hand, the need for utilization affects production mode. Moreover, researchers seek ways to use them in social One of the first topics to consider when discussing the theory/method debate is the fact that, with BD, this becomes a hybrid relationship. The datasets used in the research, are not free from theoretical references because they are the result of a construction process; most of all, we must keep in mind that use and production are co-produced. On the one hand, the production process affects utilization but on the other hand, the need for utilization affects production mode. Moreover, researchers seek ways to use them in social research to produce knowledge. This relies on the fact that someone produces them. According to Neresini (2017), data become the process themselves.

According to the scientific community and sociology, the most important change is to consider "data" as the tools that allow us to get different points of view, and not to be the points of view that allow us to choose how to generate data. Once again, it should be emphasized that -as it is typical of postmodern society- also in this field is a beta phase. Scholars should keep on learning from the interdisciplinary method and work with different scholars who can help to maximize the benefits of these studies.

Only by mixing social theory, calculation, data and models in an innovative way, researchers can give a social processes clearer view, integrating new and traditional approaches into social research (Amaturo *et al.*, 2017; Cavagnuolo, Capozza & Matrella, 2021).

Another important topic to consider is the method of social research as social object. The idea is to theorize methods, aiming for a perspective in which society is produced by research methods and at the same time produces them. The methods in this co-production relationship are both material and social (Law & Ruppert, 2013), the so-called *methodological devices*. […] Methodological devices, like any other device, "do things". The methods of social research have a double social life: they are shaped by the social and at the same time, as social vehicles, make the social (Law & Ruppert 2013, p. 233). Savage (2017) states that, in this perspective, a methodological device is an assembly of artefacts, users, practices and ideas in constant change. For this reason, they themselves are subjects of analysis and have their own social life. Consequently, this last statement makes us think that method and theory can coexist; according to this point of view, theory is closely linked to method. Theory and method are in continuous co-production.

In conclusion, authors agree on the fact that sociologists, when doing research, become part of an assembly made of human components, data and methodological devices, which modifies as the research progresses. These different assemblies co-produce in turn creating a hybrid system of knowledge, interdiscipline and devices in constant change. In this perspective, the assembly become one of the first aims of the sociological analysis, creating as previously stated, a deadline of the socio-epistemological theory/method debate (Mauceri, 2016). Dissociating from the idea that BD would sanction the "end of theory" by Anderson (2008), sociology is, perhaps, one of the main disciplines that is able to provide a theoretical-interpretative contribution to data. In this context, sociology represents a strong interdisciplinary link that too often prefers statistical significance rather than interpretation; by working in this direction, hopefully in a few years, BD will simply be absorbed within the discipline and used -as we know- like other tools of social research that each time are chosen according to the object of investigation: Big Data should not simply "work" (Sabetta, 2017; Burrell & Fourcade, 2021), but they must generate value, a value considered in all its possible forms.

References

Addeo, F., Delli Paoli, A., Esposito, M. & Bolcato, M.Y. (2019). Doing Social Research on Online Communities: The Benefits of Netnography, *Athens journal of social sciences*, (7): 9-38.

Agnoli, M.S. & Parra Saiani, P. (2016). Sulle tracce dei Big Data. Questioni di metodo e percorsi di ricerca, *numero monografico di sociologia e ricerca sociale*, XXXVII, n. 109.

Amaturo, E. & Punziano, G. (2017). Blurry boundaries: Internet, Big-New Data and Mixed methods Approach, *Data science and Social Research, Epistemology, Methods, Thecnology and Application*, New York, Springer international publishing.

Anderson, C. (2008). *The End of Theory: The Data Deluge makes the Scientific Method Obsolete*, Wired, June 23.

Aragona, B. (2017). Types of Big Data and designs of evaluation research, *Rassegna italiana di valutazione*, 68(2):48-62.

Aragona, B. (2016). Big Data or data that are getting bigger?, *Sociologia e Ricerca Sociale*, 109(3): 42-53.

Aragona B. (2008), Una nuova cultura del dato, *Sociologia e Ricerca Sociale*, 87(3): 151-173.

Auriemma, M., Esposito, E., Iadicicco L., Mazzaro, F., Polimene, A, Punziano, G., Sarnelli, C. (2015). Euroscetticismo a 5 stelle. Stili comunicativi e onlinetext data nel caso delle elezioni europee 2014, *Sociologia della comunicazione*, 47, pp. 36-54.

Battaglini, E. (2015). *Sviluppo Territoriale, dal disegno della ricerca alla valutazione dei risultati*, FrancoAngeli, Milano.

Bennato D. (2008). *Le metafore del computer. La costruzione sociale dell'informatica*, Roma, Mantelmi .

Bennato, D. (2015). *Il computer come macroscopio, Big Data e approccio computazionale per comprendere i cambiamenti sociali*, Roma, Franco Angeli.

Brensinger, J & Eyal, G. (2021). The Sociology of Personal Identification, *Sociological Theory*, 39 (4): 265-292; 10.1177/07352751211055771.

Burrell, & Fourcade, M. (2021). The Society of Algorithms, Annual review of sociology, 47, pp- 213-237; 10.1146/annurev-soc-090820-020800.

Burrows, R. & Savage, M. (2014). After the crisis? Big Data and the methodological challenges of empirical sociology, *BD& society*, 1, n. 1.

Cavagnuolo, M, Capozza, V & Matrella, A. (2021). The walkthrough method: state of the art, innovative aspects and application fields, *Handbook of Research on Advanced Research Methodologies for a Digital Society*; doi:10.4018/978-1-7998-8473-6.ch028

Chen, YS., Wu, XG., Hu, AN., He, GY& Ju, GD. (2021). Social prediction: a new research paradigm based on machine learning, *Journal of chinese sociology*, 8(1); 10.1186/s40711-021-00152-z.

Coulthard, B., Mallett, J. & Taylor, B. (2020). Better Decisions for Children with Big Data: Can Algorithms Promote Fairness, Transparency and Parental Engagement?, *Societies*, 10 (4); 10.3390/soc10040097.

Espeland W.N. & Stevens M.L. (2008). *A Sociology of Quantification*, *European Journal of Sociology*, pp.401-436.

Fotopoulou, A (2021). Conceptualising critical data literacies for civil society organisations: agency, care, and social responsibility, *Information Communication & Society*, 24(11):1640-1657; 10.1080/1369118X.2020.1716041.

Fuchs, C. (2021). History and Class Consciousness 2.0: Georg Lukacs in the age of digital capitalism and big data, *Information Communication & Society*, 24 (15): 2258-2276; 10.1080/1369118X.2020.1759670.

Fussey, P & Roth, S. (2020). Digitizing Sociology: Continuity and Change in the Internet Era, *The Journal Of The British Sociological Association*, 54 (4): 659-674; 10.1177/0038038520918562.

Halford, S. &. Savage, S. (2017). *Speaking Sociologically with Big Data: symphonic social science and the future for Big Data research*, Southampton University; doi:10.1177/0038038517698639.

Kotras, B. (2020). Opinions that matter: the hybridization of opinion and reputation measurement in social media listening software, *Media Culture & Society*, 42 (7-8): 1495-1511; 10.1177/0163443720939427.

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

Lane, T (2020). Along which identity lines does 21st-century Britain divide? Evidence from Big Brother . *Rationality and Society*, 32 (2): 197-222;10.1177/1043463120904049.

Lauro, A., Amaturo, E., Grassia, M.S., Marino, M., Aragona, B. (2017). *Data science and Social Research, Epistemology, Methods, Thecnology and Application*, Napoli, Springer.

Liu, JK & Chen, YS. (2021). Attention to social stratification in the public discourse: An empirical study based on big data of books 1949-2008, *Journal Of Chinese Sociology*, 8(1); 10.1186/s40711-021-00155-w.

Lupton L. (2014). *Digital Sociology*, Canberra, Routledge.

Marres, N., (2013). Scraping the social? Issues in live social research, *Journal Of Cultural Economy*, 6(3),313-335.

Martire, F., Pitrone, M.C. (2016). Lo studio dell'opinione pubblica al tempo dei big data. Una sfida per la ricerca sociale, *Sociologia e Ricerca Sociale*, 109(3). 102-115.

Mayer-Schonberger, V., & Cukier, K. (2013). *Big Data: A Revolution That Will Transform How We Live, Work, and Think,* New York, Houghton Mifflin Harcourt.

Mauceri S. (2016). Contro la deriva scissoria, Le funzioni dei big data nelle strategie integrate della ricerca sociale, *Sociologia e Ricerca Sociale*, 109, pp. 70-82.

Mazzaro, F., & Punziano, G. (2017) Online Textual Data and Political communication analysis, AIS *Journal of Sociology, (11): 143-158.*

McQuade, B. (2021). World Histories of Big Data Policing The Imperial Epistemology of the Police-Wars of U.S. Hegemony, *Journal Of World-Systems Research*, 27(1):109-135; 10.5195/JWSR.2021.1033

Meerstra-de Haan, E, Haartsen, T, Meier, S & Strijker, D. (2020). An Initiators' Perspective on the Continuity of Citizens' Initiatives in Rural Areas, *Rural Sociology*, 85(1):213-234; 10.1111/ruso.12288.

Orton-Jonson, K., & Prior N. (2013). *Digital Sociology: Critical Prospectives,* Palgrave Macmillan, London.

Parra Sainai, P. (2016). Le risorse e il controllo. I big data oltre il mito, in *Sociologia e Ricerca Sociale*, 109(3): 28- 41.

Sabella, L. (2017) *Basta che funzioni? Note su Big Data e teoria sociologica, AIS Journal of Sociology*, (11): 187-194

Roger, R. (2013) *Digital Methods, MIT Press, Cambridge, USA.*

Roger, R. (2019) *Doing Digital Methods, SAGE Publications, Thousand Oaks, USA.*

Ruppert, E., Harvey, P., Lury, C., Mackenzie, A., McNally, R., Baker, A., Kallianos, Y., Lewis, C. (2013). *Socialising Big Data: From concept to practice,* Centre for Research on Socio-Cultural Change, Manchester.

Roth, S & Luczak-Roesch, M. (2020). Deconstructing the data life-cycle in digital humanitarianism, *Information Communication & Society*, 23(4):555-571; 10.1080/1369118X.2018.1521457.

Sandvik, KB. (2020). Wearables for something good: aid, dataveillance and the production of children's digital bodies, *Information Communication & Society*, 23(14): 2014-2029; 10.1080/1369118X.2020.1753797.

Savage, M. & Burrows, R. (2009). *Some Further Reflections on the Coming Crisis of Empirical Sociology*, Sociology, XLIII, n. 4.

Westermeier, C. (2020). Money is data - the platformization of financial transactions, *Information Communication & Society*, 23(14): 2047-2063; 10.1080/1369118X.2020.1770833.

Willig, I. (2022). From audiences to data points: The role of media agencies in the platformization of the news media industry, *Media Culture & Society*, 44(1):56; 71.10.1177/01634437211029861.

Wynn, J.R., (2009), *Digital* Sociology: Emergent Technologies in field and the classromm, *Sociological Forum*, n. 24, Issue 2

Zani, S., Creioli,. (2015). *Analisi dei dati e data mining per le decisioni aziendali,* Edizioni nuova cultura, Roma.