

Blurry Boundaries: Internet, Big-New Data, and Mixed-Method Approach

Enrica Amaturò and Gabriella Punziano

Abstract Internet as place of everyday life, as well as a vast repository of information, becomes an integral part of the society essential for understanding complex phenomena and social issues. Social researchers cannot remain anchored to traditional research practices and conceptual categories, but instead must cope with the exponential growth of new data, much of which is user-generated and freely available online. In this manner, Web-mediated research is already transforming the way in which researchers practice traditional research methods transposed on the Web. But the steady growth of interest in these new analytical frontiers has not been properly accompanied by a solid reflection on the implications that the new data available and the analysis of such data can generate. It is necessary to undertake a critical discussion of the quality of this new data and the processes involved: data collection, organization, analysis, and treatment of ethical issues. Over the next years, this domain will raise the need of to much efforts in social research including the use of new methodological approaches. Along this track, the current chapter aims to problematize future challenges for social research, with a particular emphasis on the advent of big-new data.

Keywords Big-new data • Mixed-methods approach • Internet • Online textual data • Privacy and other constraints • Data quality • Data availability • New technical frontiers

E. Amaturò (✉)

Methodology of Social Sciences, Department of Social Sciences,
University of Naples Federico II, Vico Monte della Pietà, 1, 80136 Naples, Italy
e-mail: amaturò@unina.it

G. Punziano

Urban Studies and Social Sciences, Gran Sasso Science Institute – GSSI,
Viale Francesco Crispi, 7, 67100 L'Aquila, Italy
e-mail: gabriella.punziano@gssi.it

© Springer International Publishing AG 2017

N. Carlo Lauro et al. (eds.), *Data Science and Social Research*,
Studies in Classification, Data Analysis, and Knowledge Organization,
DOI 10.1007/978-3-319-55477-8_5

1 Introduction

Nowadays, the Internet, which more and more holds a place in people's daily lives, is configured as a consideration that is essential for understanding many different phenomena and complex social issues. If research fails to take into account this vast repository for information, its potential results are of limited value. Any research that aims at making a general investigation must examine the research environment as a whole, meaning that the Internet must be an integral part. Social researchers cannot remain anchored to traditional research practices and conceptual categories, but instead must cope with the exponential growth of new data, much of which are *user-generated* and freely available online. In this manner, Web-mediated research is already transforming the way in which researchers practice traditional research methods transposed on the Web; for example, it will be considered as online ethnographic research (i.e. *netnography*) or *social network analysis* that actually can leverage a greater potential compared to relational data relying on information provided by social networks. The steady growth of interest in these new analytical frontiers has not been properly accompanied by a solid reflection on the implications that the new data available and the analysis of such data can generate. More concretely, this means that it is necessary to undertake a critical discussion of the quality of this new data and the processes involved, which include data collection, organization, analysis, and treatment of ethical issues.

First of all, there are issues of data access because of the constraints of *privacy*. Materials on the Internet, which assume the status of *personally published documents* (Amaturò and Punziano 2013), are produced and used for different purposes than analytical use (as is generally true for all secondary data). This leads to ethical and value questions over characterization of the data. Indeed, the data can be considered either as simple information that is freely available or as the result of a contextual construction that should not be investigated outside the scope in which it was produced. The ethical boundaries over use of such data are becoming increasingly blurred, as the line between public and private material also becomes blurred.

Secondly, questions arise as to the nature of the data and the ability to analyse them. The development of the Internet and smart mobile devices, with their social applications, is making obsolete the usual ways of dealing with data. This erodes the boundaries of time (synchronous/asynchronous) and space intended as boundaries between real and virtual. Consequently, new types of data have emerged, which are classified as big data (user-generated content, streaming digital data, log data, and information generated from any communicative and Internet processes).

The problems, along with that of ethics, consist in archiving, processing and analysing these new data. Over the next years, this domain will raise the need of to much efforts in social research including the use of new methodological approaches such as the *mixed-methods* approach (cf. Tashakkori and Teddlie 2010 and

Amaturo and Punziano 2016, among others) that seems to increasingly blur the boundaries between qualitative and quantitative research.

These questions represent one of the current methodological frontiers, and along this track, the current chapter aims to problematize future challenges for the approach, with a particular emphasis on the advent of big-new data. In fact, this is an approach that has stopped fighting for emergence and is now fighting for consolidation and specialization, in particular by identifying those holes where thematic areas and issues are yet missing a consolidated discussion in the current literature on mixed methods (as well as on research in general) and the challenges coinciding with the advent of big data.

2 New Contexts and New Data

This paper discusses how academic research can address real-world problems and contribute to relevant and forward-looking solutions when the context involved—and in which the data and the knowledge are generated—is the transposition on the Web of relational phenomena.

An important issue for social research concerns the diffusion of the Internet and the so-called Web 2.0 (Boccia Artieri 2015; Hesse-Biber and Griffin 2013). This is because this mechanism has changed, and at the same time innovated, the nature of the data, along with its collection methods and analytical procedures. The social research field has seen the opening of a world of opportunities and challenges with respect to the data produced and accessible from the Internet: ‘A real *revolution* destined to change the ways of doing social research’ (Amaturo and Aragona 2016, 26).

The fast growth of data is not a novelty. But how can we make sense of such a data explosion? This is the question with which many researchers have to fight. Web traffic and social network flows, as well as the software, sensors, and tools for monitoring these flows, have become instruments in the search for a guide to rapid and effective decision-making and learning about social processes. In contemporary society, data are constantly produced as the direct and indirect effect of bureaucratic, legislative, planning, and other activities, but also as the spontaneous accumulation of information resulting from the use of social networks as means of exchange, social relationships, and knowledge (*ibidem*).

Understanding how ideas, opinions, and behaviours are constructed, produced, and spread in large communities is a process that cannot be achieved by ignoring the Web. It is a sort of *revolution* in which ‘we are really just getting under way. But the march of quantification, made possible by enormous new sources of data, will sweep through academia, business, and government. There is no area that is going to be untouched’ (Lohr 2012). Nowadays, data can be considered a new class of economic asset, such as currency or gold (Chen et al. 2012), because of the knowledge power embedded therein. Data can help to combat poverty, crime, and pollution, and it can aid in understanding world changes and trajectories or assist in

decision-making processes. Today, the extent of data is not only a lot greater than in the past, it also includes entirely new types of data. Data seems to be more communicative because they are generated from any digital process as well as through sensors in a completely new way. In accordance with the foregoing discussion, the *data revolution* can be defined as the process that has radically changed the routines for building, organizing, and analysing the data consolidated in scientific disciplines. This combined with developments in information technology, governance, and research techniques developments has, in turn, changed the way the data are used to produce knowledge about social phenomena (Amaturio and Aragona 2016).

According to Amaturio and Aragona (2016), in the word of new and big data, many distinctions can be made. Data can be distinguished according to the way in which they are generated: direct data (products from direct measurements; for example, recordings of surveillance cameras, aerial photographs for the monitoring of territories, etc.) are data that require human intervention to be collected and analysed; automated data, that which can be collected and produced by an automatic device (for example, data detected by a sensor related to smog, temperature, or movements in an enclosed space; or metadata about a telephone conversation, such as the recorded date, time, and duration); and volunteer data, data produced by individuals who upload or transmit data to a particular system (as in the case of images, texts, audio posted on social networks, blogs, sites, etc.) (Kitchin 2014). It is also possible to distinguish these kinds of data by the area from which the data are produced, e.g. public sector (which is generally accessible to researchers), private sector (generally used for internal investigation purposes), and users of digital infrastructures and platforms (generated by social media and thus difficult to access, organize, and analyse because they are produced for purposes other than research) (Elias 2012; Van Dijck 2009). However, one main distinction is certainly the flexibility of the data collected, for which it is possible to recognize the following categories: structured data (text and numbers arranged in relational databases or matrices, where the number of rows and columns is fixed); semi-structured data (which do not have a fixed pattern, but are organized in a flexible manner based on the available content); and unstructured data (which do not share the same format, such as data from social networks, web-sites, etc.). In this last group, it is necessary to mention the so-called *Big Corpora*, which consist of very large collections of textual data that require treatments of information extraction, cleaning and creation of the data, and—only at the end of this process—suitable organization in order that the data may be processed and analysed.

More generally, big-new data can be defined as referring to data that are *too big-large* (petabyte-scale collections of data that come from click streams, transaction histories, sensors, and elsewhere, thus entailing massive quantities of data), *too fast* (big amounts of data that also need to be processed quickly), *too complex* (to sense from social media data), and *too hard* (data that do not fit neatly into existing processing tools or that need some kind of preparation or analysis that existing tools cannot readily provide) to process for existing tools (Madden 2012). ‘The big data of this revolution is far more powerful than the analytics that were used in the past. We can measure and therefore manage more precisely than ever before. We can

make better predictions and smarter decisions. We can target more-effective interventions, and can do so in areas that so far have been dominated by gut and intuition rather than by data and rigor' (McAfee et al. 2012, 62). The big data movement, like analytics before it, seeks to glean intelligence from data and translate data so as to advance knowledge in every field that can be impacted by this revolution. However, three key differences could be highlighted: the volume of the data flow, the velocity in the speed of data creation and its real-time consistency, and the variety of production sources (messages, updates, and images from social networks, readings from sensors, GPS signals from cell phones, and others) that provide enormous streams of data tied to people, activities, and locations:

The development of the Internet in the 1970s and the subsequent large-scale adoption of the World Wide Web since the 1990s have increased business data generation and collection speeds exponentially. Recently, the Big Data era has quietly descended on many communities, from governments and e-commerce to health organizations. With an overwhelming amount of web-based, mobile, and sensor generated data arriving at a terabyte and even exabyte scale, new science, discovery, and insights can be obtained from the highly detailed, contextualized, and rich contents of relevance to any business or organization (Chen et al. 2012, 1168).

More and more high-impact application areas are involved in such radical processes, such as e-commerce and marketing intelligence, e-government and politics 2.0, science and technology, smart health and well-being, and security and public safety. Furthermore, any of these areas adopt or develop appropriate analytics and techniques to drive the intended effect of extracting knowledge from available big-new data.

Big-new data has been accelerated by advances in computing that allow us to measure phenomena in fine detail and as it happens in real time. In fact, the big-new data trend has also been fuelled thanks to the improved access to information and to the development of e-databases that archive old paper-based systems of knowledge. In this sense, data are not only becoming more available but also more understandable to computers. Words, images, video on the Web, and streams of sensor data generate a set of unstructured data that are not directly organizable in traditional databases. Consequently, computer tools for generating knowledge and insights from the flow of these unstructured data have grown fast because of rapidly advancing techniques in artificial intelligence, such as natural-language processing, pattern recognition, and machine learning. The application of these techniques has involved many fields, though it is true that some big challenges for big-new data are still far from a stable solution—such as, for example, the possibility of parsing vast quantities of data and making decisions instantaneously. Experience and intuition are progressively being replaced by decisions based on data, analysis, and empirical evidence. In this shift towards data-driven decision-making and evidence-based problem solving, the demand for greater predictive power from data has arisen in fields ranging from sporting bets to public health systems and economic development and economic forecasting.

Therefore, we are swimming in an expanding sea of data that is either too voluminous or too unstructured to be managed and analysed through traditional

means. New sources, among others, are clickstream data from the Web, social media content (tweets, blogs, Facebook wall postings, etc.), and video data from retail and other settings or from video entertainment. Big-new data also encompasses everything from voice data to genomic and proteomic data from biological research and medicine. And very little of this information is formatted in the traditional rows and columns of conventional databases.

As Davenport and Bean (2012) highlight, big data offers the ability to take advantage of real-time information from sensors, radio frequency identification, and other identifying devices to understand environments at a more granular level (especially from a social sciences perspective), to create new products and services (for market competition), to treat and discover cures for threatening diseases (in life sciences), and, more generally, to respond to changes in usage patterns as they occur. Big data has become the horizon from which recovery of information is what is needed to stand apart from traditional data analysis environments. In particular, the need has now emerged in the marketplace to pay attention to data flows as opposed to stocks and to rely on data scientists and product and process developers rather than data analysts. The authors also note that big data can be useful in supporting processes (such as decision-making, prevention, and security) using real-time processed information; in involving continuous process monitoring to detect changes, trends, and trajectories; and in exploring network relationships, like suggesting friends on LinkedIn and Facebook by matching characteristics and known information. In all these applications, the data do not serve as the 'stock' contained in a data warehouse but rather maintain a continuous flow. A substantial change has thus occurred from the past, when data analysts performed multiple analyses to find meaning in a fixed supply of data. Given the volume and velocity of big-new data, conventional approaches to decision-making are often not appropriate in such settings. Therefore, in big-new data environments it is important to analyse, make decisions, and act both quickly and often. The new data scientist needs to understand not only analytics, but also computer science, computational physics, and biology- or network-oriented social sciences. He or she requires skills that range from data management to programming, mathematical and statistical skills, business acumen, and the ability to communicate effectively with decision-makers. All this goes well beyond what was necessary for data analysts in the past. A key issue for big data is the fact that the world and the data that describe it are constantly changing, and organizations that can recognize the changes and react quickly and intelligently to those changes will have the upper hand.

'Big Data technologies [...are defined...] as a new generation of technologies and architectures, designed to economically extract value from very large volumes of a wide variety of data by enabling high-velocity capture, discovery, and/or analysis. There are three main characteristics of Big Data: the data itself, the analytics of the data, and the presentation of the results of the analytics. Then there are the products and services that can be wrapped around one or all of these Big Data elements' (Gantz and Reinsel 2012, 9). Clearly, one of the most relevant problems that occurs in this context is the need to bridge the gap between large-scale data processing platforms and analysis packages, but this issue involves informatics and

engineering infrastructure and thus lies outside many other relevant problems. This also suggests another important aspect in processing data: the computer and mathematical models used to tame and understand data: ‘These models, like metaphors in literature, are explanatory simplifications. They are useful for understanding, but they have their limits. A model might spot a correlation and draw a statistical inference that is unfair or discriminatory, based on online searches, affecting the products, bank loans and health insurance a person is offered, privacy advocates warn [...] there seems to be no turning back. Data is in the driver’s seat. It’s there, it’s useful and it’s valuable, even hip’ (Lohr 2012). The development of statistics and machine-learning algorithms also requires a parallel development of a consistent data management ecosystem around these algorithms, so that user can manage and evolve their data; enforce consistency properties over them; and browse, visualize, and understand their algorithms’ results. Hidden behind these necessities is the fact that every system is actually collecting more data than it knows what to do with. Transforming this information into relevant and useful knowledge is a very difficult challenge.

Furthermore, ‘Since the early 2000s, the Internet began to offer unique data collection and analytical research and development opportunities. [...] Web intelligence, Web analytics, and the user-generated content collected through Web 2.0-based social and crowd-sourcing systems (Doan et al. 2011; O’Reilly 2005) have ushered in a new and exciting era of BIandA 2.0 research in the 2000s, centred on text and Web analytics for unstructured Web contents’ (Chen et al. 2012, 1167). This means considering the data flow as a conversation that requires the integration of mature and scalable techniques in text mining (e.g. information extraction, topic identification, opinion mining, question answering), Web mining, social network analysis, and spatial-temporal analysis. The developments and advancements in these areas have had an impact on both industry and the academic sector, leading to a progressive convergence between the two.

In this revolution, another element has become noteworthy: ‘At the same time, the steadily declining costs of all the elements of computing—storage, memory, processing, bandwidth, and so on—mean that previously expensive data-intensive approaches are quickly becoming economical’ (McAfee et al. 2012, 63). A new frontier has been opened; large amounts of information exist on virtually any topic of interest, and each of us is now a walking data generator. This statement illustrates the power of big data to inform more accurate predictions, better decisions, and more precise interventions, and to enable these things on a seemingly limitless scale (ibidem). Big-new data’s power does not erase the need for vision or human insight nor does it negate the role of intuition and experience in the decision-making process.

There is, also, excitement about big-new data technologies, automatic tagging algorithms, real-time analytics, social media data mining, and the myriad of new storage technologies. In addition, big-new data are already transforming the study of how social networks function and what kinds of information they can generate accidentally. The development of sentiment analysis, for example, has greatly enriched the fields of advertising and marketing, politics, and public opinion

management, not only in matching people to interest or people to people, but also in creating huge online data sets of collective behaviours and opinions, from which it is possible to retrieve the necessary information to develop knowledge systems on the changing society.

Big-new data environments must make sense of new data. As big-new data evolves, the architecture will develop into an information ecosystem, a network of internal and external services continuously sharing information, optimizing decisions, communicating results, and generating new insights. Furthermore, we are used to thinking of big-new data as always telling us the truth, but this is actually far from reality.

Some problems occur in data acquisition (determining whether all the data are important, or how to filter out the right information) and in extraction of information (pulling out the required information from its underlying sources and expressing it in a structured form suitable for analysis), and there also remains an important question of a suitable database design that permits differences in data structure and semantics to be expressed. As was mentioned earlier, one problem with current big-new data analysis is the lack of coordination between database systems and analytics packages that perform various forms of data processing, such as data mining and statistical analyses. Moreover, the ability to analyse big-new data is of limited value if users cannot understand the analysis: ‘There is a multistep pipeline required to extract value from data. Heterogeneity, incompleteness, scale, timeliness, privacy, and process complexity give rise to challenges at all phases of the pipeline. Furthermore, this pipeline is not a simple linear flow—rather there are frequent loops back as downstream steps suggest changes to upstream steps. There is more than enough here that we in the database research community can work on’ (Labrinidis and Jagadish 2012).

Inside this greater context of concerns, one relevant question is linked to privacy. Amidst the enormous increase in digital data and transactional information, concerns about information privacy have emerged globally: ‘Privacy issues are further exacerbated now that the World Wide Web makes it easy for the new data to be automatically collected and added to databases’ (Agrawal and Srikant 2000, 439). These concerns come from the previous Internet era, in which it was an established fact that people disguised some basic characters behind false identities because of misunderstandings or accentuated concerns about privacy. The fact that such people had then, in this context, acted upon information related to their true selves showed that they were not equally protective of every field in their data records. Such disguises have become less important for people, organizations, and everything else occurring on the Internet today. The big-new data stream and flows are further and further from concerns arising from the possibility of veiled identities, actions, and patterns of usage.

What we can say is that a new context for information, the so-called digital universe, has generated a new tangible and digital geography (Gantz and Reinsel 2012), which is conceivable as a vast, barely charted place full of promise and danger. A majority of information in the digital universe is created or consumed by consumers, but the amount of information individuals create themselves (writing documents, taking pictures, uploading) is far less than the amount of information

being created about them in the digital universe (ibidem). Because of that, issues strongly emerge concerning copyright, privacy, and compliance with regulations even when the data zipping through networks and server farms are created and consumed by consumers. In a report titled ‘The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in Far East’, much attention is paid to the data producers and, as many producers are essentially consumers on the Internet, much of the data in the digital universe is unprotected. The question arises, then, whether some type of protection is required for such data (to protect privacy, adhere to regulations, or prevent digital snooping or theft): ‘Therefore, like our own physical universe, the digital universe is rapidly expanding and incredibly diverse, with vast regions that are unexplored and some that are, frankly, scary’ (Gantz and Reinsel 2012, 4). The need for information security is growing fast because of the rise in mobility and participation in social networks, the increasing willingness to share more and more data, and the new technologies that capture more data about data (the so-called metadata). For data that may have some level of sensitivity, there are several levels of security definable as privacy only (such as an email address or a stream of downloads), compliance-driven (such as emails that are contentious or subject to storage rules), custodial (such as account information potentially subject to identity theft), confidential (information protected from the origin such as personal memos, transactions, and so on), and lockdown (information requiring the highest security, such as financial transactions, personnel files and documents, medical records, etc.). In this sense, the issue may be more sociological or organizational than technological (ibidem). Also, if our digital universe is bigger and will be bigger, more valuable, and more volatile than ever, one solution contemplates the harmonization of law and regulations governing information security around the globe. As digital data transcend conventional boundaries, a global knowledge of information security may be the difference between approval and denial of a data request.

3 The Study of the Online Textual Data¹

3.1 *What Is New in Approaches and Analytics?*

Given the huge amounts of data available and the different nature that different sources of data can assume, in the following section, we want to develop a more narrow discourse concerning big-new data produced by the interaction and Internet transposition of relational phenomena, although it is true that the impact of data abundance coming from the Web extends well beyond social processes and also involves the economic, business, political, and many other fields. This combination of methods, approaches, and scientific disciplines has to be seen as an opportunity,

¹The reflections in this section recall the articulate discussion in Punziano (2016, 149–157).

because the new demand for knowledge from the Web, and from society itself, is becoming increasingly data-intensive. In this scenario, the real concern is not only the possibility to apply a computer-automated analysis, but the ability to govern it.

Moreover, in the transition from analogue to digital data, the study of content carried on the digital network in addition to the abundance of available data also established a very happy marriage with the development of automated analysis in one field of study, that of *content analysis*, which had often been accused of being strongly subjective and particularly expensive in terms of time and resources. Much progress has been made in terms of computer programming and hardware improvement, but the most essential improvements have been in textual statistics, machine learning, and on advanced and intuitive methods of visualization that allow for analysis without having to resort to a priori assumptions that are thus very iterative. This makes such systems both fundamental and extremely valid from the point of view of exploratory analysis, while presenting non-trivial difficulties from a computational point of view. This moves the subjective limitations of content analysis towards objectification, even if significant barriers are still found. However, the extensive development of available software has prompted an enormous acceleration in the use of documents as sources for social research, thus triggering the perverse mechanism of an exponential growth of additional software and applications. This direction of development has meant that the trend towards automation of analytical systems has led to a continuous approaching of quantification targets and the extension of qualitative studies, as well as to a progressive increase in-depth pretention in quantitative studies. This ensures that both approaches have seen advantages in terms of the ability to handle rapidly and easily a large amount of data, by giving to the entire analytical process a greater scientific rigour, database available for inspection, and reproducibility of the analytical procedures. However, both approaches have equal value in the cognitive scope with respect to a given phenomenon, bringing the results closer from different angles. As a result of the complexity of the phenomena under investigation in social research, and particularly in communication and politics studies, it is necessary to combine these two approaches so that the weakness of one can be overcome by the strengths of the other (Denzin 1978). This is still a difficult task, as the results obtained, in order for them to be read together, require the definition of a common vocabulary and an approach of two totally different ways of thinking, or rather must be thought of as opposites.

Therefore, if quantitative methods are used to obtain reliable models, qualitative methods are used to capture the essence of a phenomenon. The *blurring of boundaries* and the integration into routine analytical procedures, however, leads to the emergence of a third analytical approach, which aspires to establish itself as an autonomous and independent approach in social research: the *mixed-methods* approach. This approach integrates the qualitative and quantitative visions in three dimensions—epistemological, ontological, and methodological—leading to an integrated but not necessarily convergent approach. It does not use the different methods in the form of triangulating the results in order to assess their consistency, but instead uses the methods as part of an investigation toolbox that can broaden the

vision of certain phenomena and may allow for the emergence of significant and independent results (Crassewel 2003; Taschakkori and Teddly 2010). Basically, the mixed-method approach is an interesting approach devoted to integrating methods, data, and researchers. It takes the form of an approach to theoretical and practical knowledge that attempts to consider multiple points of views and perspectives, including both qualitative and the quantitative ones (Johnson et al. 2007). It is not, however, merely a summation of approaches and analytical methods, but rather an integrated and integral way of looking at the investigated reality (Bryman 2004) which allows researchers to overcome the limitations of individual approaches in order to achieve a complex model, along with a container and amplifier, of the methods individually used (Denzin 1978). Hesse-Biber and Johnson (2013), describe the mixed strategy as a means of 'coming at things differently' (Hesse-Biber and Johnson 2013, p. 103). The traditional forms of data collection, using only one method at a time, may not be adequate to address complex questions that sometimes require a variety of qualitative and quantitative lenses to achieve the desired optimal results from the study. The purpose of this approach is to improve the width, depth, and complexity of the produced knowledge (Daigneaut and Jacob 2014). In addition, it shows itself as dynamic, flexible, and inclusive, allowing in this way the deepening of complex and strongly changeable phenomena as those that take shape on the Internet.

The study of social and communication phenomena through *online textual data* is a new frontier and a new challenge in understanding society in its transposition on the Internet. But in order to make this study possible, it is necessary to consciously consider and identify the limits and new prospects for the use of online data in social research.

In particular, the use of online social text data becomes ever more essential for understanding and analysing new forms of participation in the flow of communication on the Internet. However, it is also important for the researcher to investigate the quality of online data, the ways in which these data are produced, and the possibilities that they offer to answer different research questions. This requires an acute reflection on the role of social research techniques 'on' and 'with' the Internet, that is, the so-called *Web Content Analysis* (Lilleker and Jackson 2011).

WCA, as an analytical perspective, has shown potential for development at a time of crisis between the underlying approaches to content analysis. Jointly to the emergence of the Web, which has dramatically expanded the availability of data, serious problems have arisen in relation to cognitive questions and this abundance of content and its conveyed meanings, as well as the selection, veracity, and traceability of information sources, and the use and consumption of these generated materials. The approach to statistical techniques or other quantitative analyses in this area has made it possible to dominate and systematize the availability of analytical objects by creating specific characterizations, compared to the analysis of the content published on the Web, or the WCA. That prospect has particular features that show points of similarity with the general technique from which it results, content analysis, but it reflects the advantages and disadvantages of the Web as an object and as a place of analysis. After all, the growing attention to the Internet and

its content lies in the development from a one-to-many transmission model to many-to-many—in this way, changing the rules of the game in the classic communication phenomena. In its evolutions, WCA uncovers this ambiguity and leads to a distinction involving two particular families of techniques: the traditional application of the techniques of content analysis to the Web interpreted as *strictly* in this way (McMillan 2001), and the analysis of Web content through specific techniques (Mitra and Cohen 1999; Wakeford 2000) that allow for systematic identification and study of linkage models, of message characteristics with interactive content, etc. It was specifically for the analysis of messages produced by interaction that an intense strand of computerized analysis studies of speech was developed, as described by Herring (2004), while for understanding the hypertext nature of websites was born the interconnections models (formed by hyperlinks), which dealt with the classical methods of social network analysis (SNA).

In addition, a macro-division can be carried out between: (i) *ethnography of the Web*, on the qualitative side, focusing the analysis on Web content essentially as sociocultural texts and describing and exploring them from a qualitative point of view to discover the basic meaning; and (ii) *analysis of content and functionality of the Web*, which focuses on the web page by viewing sites such as units of analysis, and developing a coding system that is applied to them in order to provide quantitative measurements of the content features, functionality, usability, and design. There is also a mixed-methods form of content analysis on the Web, whereby advocates Gibson (2010) and Lilleker and Jackson (2011) have developed mixed integrated qualitative and quantitative methods for the study of electoral campaigns on the Web, showing that, with the development of Web 2.0, it is impossible to think of static approaches in extension (quantitative) or in-depth (qualitative), but instead there is a need to look at nonparametric dynamic metrics that model the analysis at the speed of change inherent in the Web itself. Also important is the development of interactive methods based on control of a vast database and large samples, finalized by comparative analysis and with the emergence of standardized measures and automated features for the acquisition and analysis of data from social media.

3.2 *Concerns and Boundaries*

The concerns raised by the innovative reach of Web content and the methods to analyse such content are a necessary part of the process in the evolution of a research paradigm and in the need for *scientific rigour* of new approaches and new techniques associated with these developments.

Concerning the nature of the data on the Internet, the data collected from social networks are often ambiguous, problematic, and of uncertain authors, with insufficient analysis of the content carried on the Internet. Classical analysis techniques are not very suitable to enter a scene with this level of complexity, because these

techniques lead the researcher to focus on the data itself rather than on the potential of the container and the opportunities that it offers in terms of resonance. The limits of the classical analysis techniques reside precisely in their inability to reflect the ambiguities and potential scenarios opened by the Internet. The contribution of classical analysis techniques remains valid so long as the data to be analysed are highly repetitive and in a 'manageable' dataset. Such techniques are therefore not able to explain the potential of interactivity inherent to the social Web. These are techniques that generally do not work on relational matrices or interconnections between the data produced, but they allows for analysing the overall data extrapolated from the container (the Internet). This leads to the need to fragment a production scenario, which in itself presents a strong complexity and is difficult to read whether decomposed and simplified through the use of a few variables in a descriptive sense. The scenario or container, then, and the different production practices—use and generation of information for social users, as the producers, consumers, and reusers of the information transmitted through the investigated channel—are elements that the classic techniques continue to sacrifice because they are unfit to grasp the particularity of these items. New horizons have unfolded, however, thanks to the introduction of technical specifications for analysis on the Internet born from the union of text mining developments, applications of neuronal networks, big data processing through machine-learning systems, and in general from the emergence of data science as a possible applicative approach for statistical analysis in social research. Fruitful applications that demonstrate the potential to develop specific techniques for the Internet have been produced in the field of content analysis. Consider, for example, applications of *social network text analysis* (Ehrlich et al. 2007) which, unlike the analysis of classically understood content, is directed to the reconstruction of networks of relationships in which only certain concepts or areas of meaning are discussed and disclosed. Another analytic boundary is the analysis of short texts on social channels, which could lead to overcoming the lack of depth of content analysed by classifying them with respect to the polarity of associated opinions and expressing this as *sentiment analysis* (Pang and Lee 2008). This again responds to a new cognitive objective that goes beyond the mere content, or the structure of relationships that drive the content, beyond the scenario, and seeks to retrieve user practices and the interactive practice inherent in the social network. A further reflection can be highlighted. What is happening in today's society is that often specific functions end up being attributed to each platform that is used; this goes from identity spheres to showcases, from accurate reproductions of social worlds within the social network to completely virtual worlds with ever more realistic games. It is through the strong connotative use of Web channels that messages and, more generally, content are conveyed, and thus in the kind of social online data that derive from these. This is the whole crux of the matter regarding the methodological implications in the use of these specific data and in the necessary associated analysis techniques. The limits and potential of the proposed analyses and of the emerging analytical boundaries are still open for debate. These new techniques offer the possibility of jointly analysing both the

content—as well as the way in which it is used and reused—and the container in which it all takes place.

Additionally, with the advent of the Internet, additional ethical issues are generated that can also become quite thorny, for *mixed-methods* as well as for social research in general. Such is the case with the use for research purposes of *personal documents published on the Web* (Amaturò and Punziano 2013, p. 88) without those who produce them being aware of the fact that such documents can become the object of a potential study. And again, the development of Web 2.0 and smart mobile devices, with their increasing number of social applications, may soon make obsolete methods to process data that confuse the *boundaries of time* (synchronous/asynchronous) and *of space* intended as *boundaries between real and virtual*. *Ethical boundaries* become increasingly murky as the *boundary between public and private* blurs. In addition to these ethical concerns, the greatest emerging problems, given the huge amount of product data in the network, are the storage, processing, and analysis of these big-new data, and these problems in the coming years are likely to drive mixed-methods researchers to rethink, innovate, and produce new paradigmatic perspectives, as well as new structures and research designs:

The exponential growth of “big data”, arising from newly emergent user-generated and streaming digital data from networking sites such as Twitter and Facebook, will place pressures on MM researchers to transform traditional modes of collecting and analyzing data generated from these sites. Big data “is [sic] too big, moves too fast, or doesn’t fit the strictures of your database architectures. To gain value from this [sic] data, you must choose an alternative way to process it [sic]” (Dumbill 2013, p. 1). How will MM researchers incorporate big data and “big analytics” (i.e., large-scale algorithms developed to understand the meanings contained within individual social networking outputs)? In the coming years, big data methods and analytics may also drive and challenge MM researchers to rethink and innovate and produce new paradigmatic perspectives and research designs and structures. In turn, MM perspectives and praxis can provide models for interpreting and deriving critical insights that that may give a more complex understanding of big data that can bring a set of new questions and understanding to the trending data currently extracted from user-generated social networking sites (Hesse-Biber and Johnson 2013, p. 107).

4 Open Questions²

Regarding what has been discussed in the previous sections, and taking up the arguments of Hesse-Biber and Griffin (2013), one issue still under exploration for mixed-methods appears to be the use of technology mediated by the Internet for data collection. Advantages, disadvantages, and ethical issues are all factors involved in this process, and it is impossible to disregard these in discussing how

²The reflections in this section recall the complex debate exposed in Punziano (2016, 157–162) and its developments in Amaturò and Aragona (2016, 25–50).

the Web has changed the possibilities of movement for the researcher and opened scenarios of knowledge that were roughly unthinkable until now.

The Web 2.0 is a multifaceted scenario, but it is essential to understanding many phenomena and very complex social issues. Apart from noting that to ignore this boundless container of information limits the potential value of research, advocates of mixed-methods research have for decades argued that adopting a single approach, either qualitative or quantitative, confines the possibilities of knowledge. Even the reality to be investigated should be examined as a whole, and every day, the Internet becomes more and more the space where this information ends up.

Just to cite a few figures, consider that, according to the Internet World Statistics (2015), about 30% of the world's population are Web users and, in Europe and America, about 80% of households are connected to the Internet, with access rates ranging from weekly to daily. It is not hard to guess how pervasive the growth of user-generated data has become, and it is therefore ever more limiting for the research community to remain anchored to the traditional practices and concepts. Thus, Internet-mediated research is already transforming the way in which researchers practice traditional research methods transposed on the Internet; for example, important illustrations include the previously mentioned ethnographic online research (Andrews et al. 2003; Denissen et al. 2010; Dicks and Mason 2008; James and Busher 2009; Robinson and Schulz 2011), which nowadays has the more specific connotation of *netnography* (Kozinets 2010; Addeo and Esposito 2015), and *social network analysis* (Wasserman and Faust 1994; Scott 2012), which currently can make the most of its potential compared to relational data by relying on information provided by social networks (Steinfeld et al. 2008).

The steady growth of interest in research mediated by the Web, however, has not been adequately complemented by solid reflection and careful consideration of the extent to which this can actually facilitate the practice of mixed methods, as well as of the other individual approaches. Experiments, comparisons, and demonstrations have proliferated, serving to set up a complex mosaic that still needs to find the right linkage. Hewson (2003, 2007, 2008), reflecting on the contribution provided by Web-mediated research using the mixed-method approach, observes that finding a basin of data collection directly from the Internet allows the researcher to have a more dynamic element for validating research results obtained offline, or at least provides complementary elements to fill the gaps from offline collection when the investigated phenomena also have a significant component online (Davis et al. 2004). Online collection may also assist in defining the most representative samples, which can improve the ability of researchers to generalize the overall results of their studies (Hewson 2003).

In contrast to the exciting possibility that this online scene offers to researchers, there are also quite a few brakes. As is known, there is a wide selection of classical literature on the influence of the medium through which a message is sent; consider the contributions of McLuhan (1964) and Meyrowtz (1994) and of the famous statement: 'the medium is the message'. With the advent of the Internet, there are not insubstantial concerns about the influence exerted by this new medium with respect to information and produced communication, as the technology ends up

structuring an environment with its own rules and thus a process takes place that only in appearance looks like what is happening offline. In seeking to enter into these dynamics, know them, and force them into a given system of interpretation, the researcher can no longer ignore the characteristics and rules of this environment.

Although many disciplinary fields have adopted the Internet-mediated research, there remain many obstacles and concerns for its development in the social sciences. First of all, there are the *questions of access to data* for the *constraints of privacy*, because much of the material produced on the Internet is produced and enjoyed for purposes other than analysis. These are *ethical questions* about the value assigned to the data, in particular whether it can be considered as simple information that is freely available or whether it is the result of a contextual construction that should not be investigated outside the scope in which it was produced. In other words, the Internet has obscured the *ethical boundaries* between what is public and what is private. In addition, many prejudices remain about the veracity and plausibility of data circulating on the Internet, the inability to verify the source of the message, and the reliability of the content. Numerous problems also arise regarding the *issues of representativeness of the achieved sample*, given the changing nature of *cyberspace*, where constancy is as unrealistic as the pretention to photograph a moment, if immediately before and immediately after that moment, the image can change with impressive speed. This has strong implications for the possibility of generalization to a population that is also changing, but not as quickly (Andrews et al. 2003). And again, data produced in very different ways requires a researcher to leave his or her disciplinary zone of comfort, consisting of analytical and technical certainties, in order to develop new skills with respect to both the collection and the analysis of new data. And if this particular point scares disciplinary and methodological purists, it will find mixed-methods researchers more ready to incorporate these changes and new challenges, given the flexibility that the approach in question requires.

Despite the limits and issues already presented for online research, there are many strengths that push towards the value of further development with a goal of overcoming these limits. Among the advantages is the *overcoming of spatial barriers* in data collection, which allows researchers to work on large samples with extremely limited costs; the online environment permits immediate accessibility that classical instruments cannot grant to researchers. Nevertheless, researchers should ignore the need for reflection on the possibilities of compliance and the return of the information, on the delimitations of the real space in which the respondents are located, and especially on who are the real users of the Internet, the entire population or only the part that is more educated, younger, and familiarized with new technologies? Another strength is the ability to leverage both synchronic collection methods, i.e. collecting *streaming data* (conversations, comments, posts, audio-visual material, and anything else that is freely available on social platforms, or even retrieving something closer to traditional data, such as articles on platforms of information) and asynchronous collection methods (email inquiries, questions to Delphi, or online surveys). Whatever the method chosen, though, one effect is the progressive distancing of the researcher from the situation of a face-to-face

collection method, which decreases the influence that the researcher's presence may have on the data that he or she intends to collect. However, the absence of non-verbal signals in online interactions can become a new obstacle that counters the implicit advantage just emphasized. The impact of the Internet is, therefore, extremely controversial, and whenever there is an advantage to its use in data collection and research, in general, it tends to immediately arouse new doubts and controversy. It is for this reason that Hesse-Biber and Griffin (2013) have questioned whether researchers are really aware of the potential impacts that the new technological mode of mediation can have on every aspect of the research process and what new questions such awareness can bring. The authors, finally, identify points of strength that are more or less common in studies developed in the field of mixed-methods research making use of Internet mediation, though this does not go so far from the advantages that can be derived from all the social research implemented in the Internet. Summarizing what has already been discussed, it is essential to take into account:

1. the possibility of reaching populations that are otherwise hard to contact through the identification of groups of interest, and the possibility of achieving a wide range of cases that are territorially dispersed;
2. the increase of sensitivity and fidelity of data, especially if what is under investigation involves unknown phenomena, and the ability to identify networks that have influence in the formation and definition of opinions;
3. the gains in terms of time (such as platforms for online surveys with questionnaires that allow responses to be recorded directly in a data matrix, operations that reduce time for insertion, eliminate potential human error in transcription, and thus automatically enhance the validity of the process) and cost deduction (e.g. cost reductions for travel or office materials such as paper, pens); reducing the time factor in mixed-method sequential or concurrent studies helps to reduce collection times for different data and thus may potentially restrict one of the basic problems of this approach, namely the opportunity to collect both sets of results, discuss them jointly, and publish them in studies that are not fragmented by the availability of first one set of data and then the other; and
4. the potential increase in participation rates and responsiveness thanks in part to the simplicity of the built tools (e.g. online questionnaires), but also because it is possible to answer extremely sensitive questions while masked behind a monitor (increases in terms of the perception of anonymity may consequently decrease the social desirability bias of responses), in the safety of the respondent's own home, and with the ability to use as much time as necessary to provide complete answers to complex questions.

This last point exactly can be a double-edged sword. In fact, the advantages listed are surely matched by the decline in terms of empathy. An online interview does not have the same advantages, in terms of involvement, of an interview conducted in person, and it can create a sense of disinterest in the respondent if he

or she ends up attributing little weight, relevance, and scientific value to an analysis conducted in this way; such a respondent may also fail to pay due attention to the answers given because he or she considers them of little use in achieving their desired purpose. In this sense, the researcher, due to the lack of interpersonal dynamics, will have a more difficult task in establishing an online relationship rather than an in person one. But more than an increase in terms of process validity, validation of results, and verification of authenticity of the obtained results, the use of online research becomes instrumental in improving the overall understanding of the research problem. Before designing an online research effort, a researcher must have very clear objectives for the research, an understanding of the effects of Internet mediation, and knowledge of how all this can improve the overall research process by limiting the *side effects of online transposition*. One of the biggest problems continues to be that of the under-representation on the Internet of some specific population types (e.g. the homeless, the elderly, and the less-few educated people), so to the increase in terms of possible contact in a large space with reduced time and costs, it goes to take into account the need, sometimes essential, to support online procedures with the classic offline procedures, whose combined use becomes very useful at this stage of advancement of the process of computerization of society. The Internet permit to the social scientists to open new scenarios to provide meaning to complex social environments inside and outside the Internet itself, leading to new methodological challenges and new stimuli for the imagination of the researcher through new research directions.

As seen, the online research is not devoid by drawbacks. For example, Hine (2008) questions the real nature of collected data online and shows decided doubts about the possibility of attributing to an Internet study the characteristics of a study in-depth, as the Web offers a dynamic and fleeting environment as its intrinsic connotation. Among the drawbacks, synthesizing, there are:

1. the lack of generalizability, because Internet users tend to be still mainly composed of more educated people, young, and affluent enough to afford access and tools;
2. the digital divide and technological development, for whom the researcher will have to keep in mind the technological developments and the extent to which they reach the population as a whole before being certain of potentially accessible from the online research coverage level, also bearing in mind that these parameters can vary from country to country and that the exclusion of a significant part of the population by social studies in the Internet can become a great threat to the randomness and representativeness of the investigated samples;
3. the self-selection of respondents and often uncontrollable multiple compilations that ultimately affect the generalizability of the results of the study;
4. the ethical issues concerning the establishment of identity on the Internet and the basic sociodemographic characteristics essential to the performance of any social study such as age, sex, race, or other personal demographic characteristics; this information depends entirely on the honesty of the respondents;

5. the lack of non-verbal cues and the difficulty in interpreting the silence and temporal fluctuations in responses that, unlike it happens in face-to-face interaction, the researcher cannot base his knowledge on the non-verbal support and the interpersonal connections that help establish rapport such as tone of voice, body language, gestures. This, in addition to affecting the wealth of qualitative data in the online research, also requires the researcher to learn how to assess other elements that can give colour to the collected data, such as the emoticons, the use of punctuation, the online jargon, that can add emotion to online interview. Interesting is the interpretation of silence, so Madge and O'Connor (2002) provide a number of alternatives to the intended meanings of silence as a delay in the digitization of the thoughts: the interviewee may need time to process; he forgot to press the sending; he has no clear ideas or did not understand the question that is asked to him; or simply, he is just slower and less comfortable to type rather than verbally expressing his thoughts.

All these problematic issues lead to the need that researchers are properly trained with respect to the research mediated by the Internet in order to be able to conduct not only research, but also to know how to interpret the results. Researchers should be aware of the possibilities of research on the Internet, but also of the available options to them as a plausible alternative research paths. Along with the steady progress of opportunities in terms of research offers by the Internet, innovative efforts have to be implemented in research, especially in the mixed one that permit to provide revolutionary answers through the integration of multiple paths. And this without forgetting that many of the traditional ethical concerns remain unchanged both if the research is conducted online than if it takes place offline. Also, the boundary between what is public and what is private becomes increasingly blurred and the possibility of access to research arena without reveal their intentions, even if it becomes a very easy choice for access to Web-community and social network, it should be equally considered and adequately justified, not only on the ethical side, but also on the methodological one.

References

- Addeo, F., & Esposito, M. (2015). *Informal learning and identity formation: A case study of an Italian virtual community*.
- Agrawal, R., & Srikant, R. (2000, May). Privacy-preserving data mining. *ACM Sigmod Record*, 29 (2), 439–450.
- Amaturo, E., & Aragona, B. (2016). *La “rivoluzione” dei nuovi dati: quale metodo per il futuro, quale futuro per il metodo?*. In F. Corbisiero and E. Ruspini (Eds.), *Sociologia del futuro. Studiare la società del ventunesimo secolo* (pp. 25–50). Milano: CEDAM.
- Amaturo, E., & Punziano, G. (2016). *I Mixed Methods nella ricerca sociale*. Roma: Carocci Editore.
- Amaturo, E., & Punziano, G. (2013). *Content Analysis: tra comunicazione e politica*. Milano: Ledizioni.

- Andrews, D., Nonnecke, B., & Preece, J. (2003). Electronic survey methodology: A case study in reaching hard-to-involve Internet users. *International Journal of Human-Computer Interaction*, 16, 185–210.
- Boccia Artieri, G. (2015). *Gli effetti sociali del web*. Franco Angeli: Forme della comunicazione e metodologie della ricerca on-line. Milano.
- Bryman, A. (2004). *Social research methods* (2nd ed.). Oxford: Oxford University Press.
- Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business Intelligence And Analytics: From big data to big impact. *MIS Quarterly*, 36(4), 1165–1188.
- Cresswell, J. W. (2003). *Research design: Qualitative, quantitative and mixed method approaches*. London: Sage.
- Daigneault, P. M., & Jacob, S. (2014). Unexpected but most welcome mixed methods for the validation and revision of the participatory evaluation measurement instrument. *Journal of Mixed Methods Research*, 8(1), 6–24.
- Davenport, T. H., Barth, P., & Bean, R. (2012). How big data is different. *MIT Sloan Management Review*, 54(1), 43.
- Davis, M., Bolding, G., Hart, G., Sherr, L., & Elford, J. (2004). Reflecting on the experience of interviewing online: Perspectives from the Internet and HIV study in London. *AIDS Care*, 16, 944–952.
- Denissen, J. J. A., Neumann, L., & van Zalk, M. (2010). How the Internet is changing the implementation of traditional research methods, people's daily lives, and the way in which developmental scientists conduct research. *International Journal of Behavioral Development*, 34, 564–575.
- Denzin, N. K. (1978). *The research act*. New York: McGraw-Hill.
- Dicks, B., & Mason, B. (2008). Hypermedia methods for qualitative research. In S. N. Hesse-Biber & P. Leavy (Eds.), *Handbook of emergent methods* (pp. 571–600). New York, NY: Guilford Press.
- Doan, A., Ramakrishnan, R., & Halevy, A. Y. (2011). Crowdsourcing systems on the world-wide web. *Communications of the ACM*, 54(4), 86–96.
- Dumbill, E. (2013). Making sense of big data. *Big Data*, 1(1), 1–2.
- Ehrlich, K., Lin, C. Y., & Griffiths-Fisher, V. (2007). Searching for experts in the enterprise: combining text and social network analysis. In *Proceedings of the 2007 International ACM Conference on Supporting Group Work* (pp. 117–126).
- Elias, P. (2012). *Big data and the social sciences: a perspective from the ESRC*, presentation at the conference Shaping society.
- Gantz, J., & Reinsel, D. (2012). *The digital universe in 2020: big data, bigger digital shadows, and biggest growth in the far east* (pp. 1–16). IDC iView: IDC Analyze the future.
- Gibson, R. (2010). *Web content analysis*. Manchester: University of Manchester Press.
- Herring, S. C. (2004). *Computer-mediated discourse analysis: An approach to researching online behavior*. Cambridge: Cambridge University Press.
- Hesse-Biber, S. N., & Griffin, A. J. (2013). Internet-mediated technologies and mixed methods research problems and prospects. *Journal of Mixed Methods Research*, 7(1), 43–61.
- Hesse-Biber, S. N., & Johnson, R. B. (2013). Coming at things differently future directions of possible engagement with mixed methods research. *Journal of Mixed Methods Research*, 7(2), 103–109.
- Hewson, C. (2003). Conducting research on the Internet. *The Psychologist*, 16, 290–293.
- Hewson, C. (2007). Web-MCQ: A set of methods and freely available open source code for administering online multiple choice question assessments. *Behavior Research Methods*, 39, 471–481.
- Hewson, C. (2008). *Internet-mediated research as an emergent method and its potential role in facilitating mixed methods research*. In S. N. Hesse-Biber and P. Leavy (Eds.), *Handbook of emergent methods* (pp. 543–570). New York, NY: Guilford Press.
- Hine, C. (2008). Internet research as emergent practice. In S. N. Hesse-Biber & P. Leavy (Eds.), *Handbook of emergent methods* (pp. 525–541). New York, NY: Guilford Press.
- Internet World Statistics (2015). *Internet users and population stats for the Americas*.

- James, N., & Busher, H. (2009). *Online interviewing*. Thousand Oaks, CA: Sage.
- Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Towards a definition of mixed methods research. *Journal of Mixed Methods Research, 1*(2), 112–133.
- Kitchin, R. (2014). *The data revolution: Big data, open data, data infrastructures and their consequences*. London: Sage.
- Kozinets, R. V. (2010). *Netnography: Doing ethnographic research online*. London: Sage Publications.
- Labrinidis, A., & Jagadish, H. V. (2012). Challenges and opportunities with big data. *Proceedings of the VLDB Endowment, 5*(12), 2032–2033.
- Lilleker, D., & Jackson, N. (2011). *Campaigning. Elections and the Internet: US, UK, Germany and France*. London: Routledge.
- Lohr, S. (2012). The age of big data. *New York Times, 11*.
- Madden, S. (2012). From databases to big data. *IEEE Internet Computing, 3*, 4–6.
- Madge, C., & O'Connor, H. (2002). Online with e-mums: Exploring the Internet as a medium for research. *Area, 34*, 92–102.
- McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., & Barton, D. (2012). Big data. The management revolution. *Harvard Bus Rev, 90*(10), 61–67.
- McLuhan, M. (1964, tr.it. 1994). *Understanding media: The extensions of man*. Cambridge: MIT Press.
- McMillan, J. H. (2001). *Essential assessment concepts for teachers and administrators*. Thousand Oaks, CA: Corwin Publishing Company.
- Meyrowitz, J. (1994). Medium theory. In D. Crowley & D. Mitchell (Eds.), *Communication theory today* (pp. 50–77). Cambridge: Polity Press.
- Mitra, A., & Cohen, E. (1999). Analyzing the web: Directions and challenges. In S. Jones (a cura di). *Doing internet research*. London: Sage. 179–202–220.
- O'Reilly, T. (2005). *What is Web 2.0? Design patterns and business models for the next generation of software*.
- Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and trends in information retrieval, 2*(1–2), 1–135.
- Punziano, G. (2016). Il futuro dell'approccio: vantaggi, limiti e nuove prospettive. In E. Amaturio & G. Punziano (Eds.), *I Mixed Methods nella ricerca sociale*. Roma: Carocci Editore.
- Robinson, L., & Schulz, J. (2011). New fieldsites, new methods: New ethnographic opportunities. In S. N. Hesse-Biber (Ed.), *The handbook of emergent technologies in social research* (pp. 180–198). New York, NY: Oxford University Press.
- Scott, J. (2012). *Social network analysis*. London: Sage.
- Steinfeld, C., Ellison, N. B., & Lampe, C. (2008). Social capital, self-esteem, and use of online social network sites: A longitudinal analysis. *Journal of Applied Developmental Psychology, 29*(6), 434–445.
- Tashakkori, A., & Teddlie, C. (2010). *Handbook of mixed methods in social and behavioral research*. London: Sage.
- Van Dijk, J. (2009). Users like you? Theorizing agency in user-generated content. *Media, Culture and Society, 31*(1), 41.
- Wakeford, N. (2000). *New media, new methodologies: Studying the web*. In D. Gauntlett (a cura di). *Web.studies: Rewiring media studies for the digital age* (pp. 31–42). London: Arnold.
- Wasserman, S., & Faust, K. (1994). *Social network analysis. Methods and applications*. Cambridge: Cambridge University Press.