

# Chapter 20

## What People Leave Behind Online: Digital Traces and Web-Mediated Documents for Social Research



Laura Arosio

### What People Leave Behind (Online)

Individuals and groups leave evidence of their lives when they are engaged in their activities. They move through time and space and modify the environments in which they live, leaving behind signs of their passage. These signs include a variety of materials that differ in content and form, such as written texts, images, material objects, audio tracks, links, maps, metadata and hypertexts. They may relate to personal interests, have to do with larger organizations or be cultural products. This evidence is not created for research purposes, but it can provide a great deal of insight into individual and group behaviours, attitudes and values (Webb et al., 1966).

Leaving behind a sign of one's activities has been part of the human condition since the first appearance of humankind, and it is seen as a necessity for maintaining memory and ensuring the existence of the world in which people live (Gleick, 2011). This process has accelerated in certain historical contexts and because of breakthroughs in communications (such as the introduction of writing, printing and mass media) or even technical progress in the field of manufacturing and transportation.

The creation of a vast amount of documentation, generally in written form, has become a constituent of contemporary societies. We can consider Max Weber's idea of bureaucracy (Weber, 1922) and Jacques Derrida's reflection on writing to be characteristic elements of our age (Derrida, 1967). These materials are thought of as a source of legitimation of the existence of groups and organizations and of the activity of individuals. In this respect, a practice such as photography becomes a tool for people to make everyday experiences "real" (Sontag, 1977). In addition, in contemporary societies, there has been a continuing trend towards the creation and

---

L. Arosio (✉)

Department of Sociology and Social Research, University of Milano-Bicocca, Milan, Italy  
e-mail: [laura.arosio@unimib.it](mailto:laura.arosio@unimib.it)

spread of material objects, whether mass-produced or personalized, that circulate on a planetary level and are charged with meaning for those who produce and consume them (Kopytoff, 1986; Appadurai, 1986).

Today, the advent of new information technologies and the growth of the World Wide Web have encouraged the creation, dissemination and preservation of many different types of materials that people leave behind online. The Web is a place of interaction in which a very large number of people move about, spend time and practice a variety of activities, leaving signs of their passage. In surfing the Net, people produce a large amount of material on different topics. These materials are stored and recorded in different places, such as on personal and institutional devices, in forums and on blogs and social network pages.

With the advent of the Internet, many aspects of social life have been coded and quantified, and these data have been stored and potentially made available to third parties (see the concept of datafication by Mayer, Schoenberger and Cuki (Mayer-Schoenberger and Cukier 2013) and its critics (see van Dijck, 2014)). The use of new technologies has introduced a revolution that is not only technological but also social and cultural (among others, see the idea of *documediality* (Ferraris & Martino, 2018) and *self-tracking culture* (Lupton, 2019)). These changes bring a new kind of reflexivity and can be read as opening a new era of social sciences (Boullier, 2015).

The materials that individuals and groups leave behind while performing their online activities can provide much information about their behaviours, values and ways of thinking. Messages, posts, photos, videos, audio files, searches and online activities become persistent data that account for a wide range of experiences. As some scholars have argued, digital data can be seen as a kind of individual and social memory (Hand, 2016) or identity (Reigeluth, 2014; Kneidinger-Muller, 2018). If properly used, these data provide information about experiences, beliefs and values. Even though they were not originally intended as research materials, researchers can use them to study contemporary societies.

Before they are used in research practices, digital data left behind need to be placed in a rigorous theoretical and methodological framework. They cannot be used indiscriminately and without preliminary investigation. The aim of this contribution is to debate what people leave behind (WPLB) online from a methodological point of view, recognizing elements of both continuity and novelty in comparison to other types of data sources. We first connect WPLB online data to the unobtrusive measures framework and point out the major strengths and weaknesses of these data. Then, we propose dividing the broad family of unobtrusive measures collected online into three different categories. This categorization has important implications for research. We then discuss the characteristics that allow us to distinguish different online materials. As a result, the importance of contextualizing (digital) data is emphasized.

## WPLB Online as Unobtrusive Measures

From a methodological point of view, WPLB online has a first distinctive feature: it was not produced for the purpose of scientific research. The materials have been created spontaneously by individuals and groups while performing their activities and have not been solicited in a research context. For this reason, WPLB online can be considered to be part of the so-called unobtrusive measures. This term was coined by Webb and colleagues in 1966 (Webb et al., 1966) and refers to data collected through methods that do not require direct elicitation by researchers (Webb et al., 1966, 1981; Kellehear, 1993; Lee, 2000). In unobtrusive data collection, the research team does not interact with the subjects to be studied and does not require active cooperation from them. Therefore, unobtrusively collected data are considered nonreactive: because there is no direct contact between researchers and those observed, the subjects do not alter their behaviours because they do not know that they are being studied (Given, 2008).<sup>1</sup> Unobtrusive measures provide complementary—not alternative—information to be used in conjunction with data collected through direct elicitation methods (Sechrest, 1971).

WPLB online data share the wide potential of traditional unobtrusive measures. They are not reactive and allow hidden populations or practices to be studied (Hine, 2011). They can be used together with data gathered by intrusive methods such as interviews, questionnaires and participant observation. They can encourage integration among methods (Tashakkori & Teddlie, 1998; Creswell & Plano, 2011) and support methodological triangulation (Denzin, 1978; Morse, 1991). Furthermore, the use of unobtrusive data collection online enhances the study of materials such as written texts, images, audio tracks and other data that risk being marginalized by mainstream research. The use of WPLB online can foster imagination and creativity, against the risk of relying excessively on self-reported measures and on the technical requirements of the research process (Mills, 1959). The study of different materials promotes contamination and boundary crossing between academic studies. By emphasizing that every human activity is cultural and full of meaning, WPLB online data promote the Internet as a place for research and the study of online environments as a source of information. WPLB online data welcome the challenge of a “punk” sociology, which is able to consider new methods, new knowledge and new representations of social life (Beer, 2014).

WPLB online data magnify some of the benefits of traditional unobtrusive measures (Hine, 2008; Janetzko, 2017). Worldwide, an increasing number of people with diverse characteristics are currently online, performing a wide variety of activities: they participate in discussion forums, leave opinions and reviews, upload photographs and videos, find a partner, learn and offer skills, make purchases, spend

---

<sup>1</sup>Data collection through direct elicitation methods such as interviews and participant observation is considered reactive methods: researchers can systematically distort their measurements by their presence because the subjects can modify their behaviour in order to provide a desirable picture of themselves—the so-called Hawthorn effect (Mayo, 1949).

their free time and send and receive messages. Unobtrusive data collection online allows us to research a large number of people and facilitate the gathering of a large volume of data, breaking down geographical distances and increasing the speed of communication. Unobtrusive measures conducted online are often cumulative and allow the gathering of longitudinal data. In this respect, WPLB online data can also support comparative studies (Smelser, 2013) of social phenomena. Online data are already registered and stored; they are readily available, inexpensive and easy to access.

WPLB online also present the limitations of unobtrusive measures. Some of these limitations are at risk of being reinforced by the online collection method. Although everything is currently increasingly connected, some activities are not performed on the Internet and therefore do not leave digital evidence. Some data are kept private and are difficult to access. Some are selected for preservation, while others are not. Digital data can be incomplete, inaccurate and dispersed (Pink et al., 2018). As a consequence, some behaviours and opinions can be collected and recorded by unobtrusive methods online, while others cannot (Janetzko, 2017). There are also differences among those who use digital devices in terms of age, gender, socio-economic condition and geographical area: while some groups are totally confident in utilizing digital tools and create a variety of contents, others are excluded (the so-called digital divide; see, among others, Norris, 2001). While the main disparities between those who have access to digital devices have diminished, there are still deep differences between those who produce content and those who do not (second-level digital divide; see Hargittai, 2002). In many cases, the identity of authors is not known, as they are anonymous or use pseudonyms. People use strategies of identity management online (Janetzko, 2017) and tend to provide wrong or misleading information. Communications can be self-censored because of privacy concerns (Eynon et al., 2008; Joinson et al., 2010). For these reasons, unobtrusive data collected online have limitations in relating content to authors' characteristics. The purposes and recipients of the data may also be unclear. Additionally, there are ethical questions to be considered in using unobtrusive methods online: to what extent can WPLB online be used for research purposes? Should the authors be informed about the use of their data? Should consent be required? Do information and sensitive issues exist that should be protected? (For a wider discussion, see, among others, Johns et al. (2004) and McKee and Porter (2009)).

These first remarks show how WPLB online can be reconnected to the classical methodological framework. It shares the advantages and disadvantages of unobtrusive methods and at the same time poses unprecedented challenges. Moreover, it is clear that WPLB online data cannot be used without first being analysed, examined and interpreted. Their meaning depends on the circumstances of their production, creation and dissemination (Boyd & Crawford, 2012; Leonelli, 2016).

### Three Categories of WPLB (Online) Data

Data collected online by unobtrusive methods are often viewed as being undifferentiated and referred to as “traces”, which is used as a synonym for “evidence”. We believe that this term is inaccurate and does not reflect the complexity and rich variety of digital data. On the one hand, online materials can have different characteristics; therefore, we propose dividing them into three categories of unobtrusive data collected online. On the other hand, the term “traces” has already been used in the methodological literature in a restrictive sense; therefore, we propose a different vocabulary. It is not only a terminological issue but also a conceptual and methodological one. Distinguishing between different categories, as well as indicating them in clear and unambiguous terms, allows us to understand the real nature of the data and their specific contribution to knowledge. It also makes it possible to match the different categories of digital data to the traditional methodological approaches to which each of them belongs. Furthermore, drawing boundaries between different categories of unobtrusive materials online stresses the need to analyse these data before they are used for research purposes.

In their seminal work on nonreactive measures in social research, Webb and colleagues distinguish three types of unobtrusive data: *found data*, *retrieved data* and *captured data* (Webb et al., 1966).

By the term “found data”, the authors refer to material inadvertently left behind by subjects and groups as they go about their lives. Found data are defined as the remnants of their passage (pressed grass, discarded items, removed flyers, worn tiles, etc.). They give this type of material the name “traces” (Webb et al., 1966). Traces can be left by erosion or accretion. In the first case, something is removed from the environment (floor wear in the halls of a museum); in the second case, something is added (garbage thrown in the baskets of the halls of the same museum). In these two examples, traces are remnants of visits to the museum that can be used to understand the behaviours, habits and preferences of the visitors.

“Retrieved data” are defined by Webb and colleagues as materials intentionally created by individuals and groups while pursuing their aims. They can be public (laws, regulations, newspaper articles, billboards, songs) or private (family photographs, letters to friends, personal notes). Webb et al. (1966) distinguish “running records”, which are archival materials that have a continuous form and cover long periods (data gathered for administrative purposes, actuarial records, sales data, media materials that appear in regular form), from “episodic records”, which are discontinuous (a sentence, some letters, a few novels). Retrieved data can take different forms and use different languages. They reveal tastes, attitudes, choices and behaviours and show how events and meanings are socially constructed. It is worth noting that retrieved data correspond to the definition of documents in documentary analysis (McCulloch, 2004; Prior, 2003; Scott, 1990; Scott, 2006). Webb et al. use the term “documents”, particularly in relation to personal and episodic records, and the term “archives” for public and running records.

“Captured data” are defined as behaviours and non-verbal cues such as movements, postures, gestures and even conversations “in situ” captured using nonparticipant observation methods, such as simple observation, meaning unobserved, passive, unobtrusive observation (Webb et al., 1966). These are not persistent but ephemeral data that arise during social interactions and vanish in the moment they are realized, so they need to be captured by researchers. Examples are analysing non-verbal behaviours, such as looking, touching and verbal latency, to understand the social dynamics of a group, listening to market conversations between sellers and customers to understand how a product’s identity is constructed and studying eye movements to reveal interest or other attitudes (Lee, 2000).

Webb and his colleagues (1966) talk about physical materials and social interactions that occur in face-to-face environments. We believe that this distinction can be adapted for WPLB online. There is a wide family of data collected online by unobtrusive methods. Within this family, we can distinguish three categories: *online found data* (unintentional digital traces/traces in the restrictive definition), *online retrieved data* (web-mediated documents with communicative ends) and *online captured data* (ephemeral behaviours that occur online).

### ***Online Found Data***

Online found data are remnants of other online activities produced inadvertently by users while navigating the Internet. Online found data include log file data (i.e. reports of technical operations carried out online generated automatically by computer applications—they can be access log files, request log files or email log files), mouse clicks, search requests, links, cookies and time measurements. Log file data can be used to generate statistics on the number of pages requested, time spent on a particular site and web browsing patterns. Email log files report information on senders, receivers, times and data of messages, disclosing networks of relationships and their characteristics. Cookies can gather information on visits to websites (date and time, action performed). Time measurements capture durations and latencies (for an in-depth presentation, see Janetzko, 2017). Online found data are a residue left unintentionally. According to the more restrictive methodological definition, online found data are digital traces (Lee, 2000). The methodological roots of this approach can be found in classical trace analysis (Webb et al., 1966; Kellehear, 1993).

### ***Online Retrieved Data***

Online retrieved data are materials that Internet users intentionally create and upload to the Web. They can be texts, videos, images, audio tracks, and hypertexts. Online retrieved data are created not for research purposes but to achieve the authors’ goals (private purposes, administrative aims, communication, artistic expression).

Examples are messages and photographs uploaded to social networks, administrative acts published online, news broadcast on the Internet, movies and songs on personal, institutional and cultural pages. Texts and photographs published on blogs can be used to explore the authors' representations of their lives (Snee, 2013). Online retrieved data have communicative purposes and express the point of view of individuals and groups. Owing to their characteristics, they are web-mediated documents (Arosio, 2010). The methodological roots of this approach can be found in classical documentary analysis (McCulloch, 2004; Prior, 2003; Scott, 1990; Scott, 2006).

### ***Online Captured Data***

Online captured data are behaviours, conversations, gestures, non-verbal cues and expressive movements captured simultaneously by observers while people are interacting online. Examples are synchronous interactions taking place in digital contexts such as online conferences, lectures, chat rooms and virtual worlds when actors are simultaneously connected. Online captured data come from nonpersistent social interactions of which no record would remain. They are ephemeral data captured through simple observation (hence the need to record digital field notes, as suggested by Boellstorff et al., 2012). The researcher is either invisible or hidden behind a false identity in order to be unobtrusive.<sup>2</sup> Online captured data refer to the so-called netnography (Kozintetz, 2010; Costello et al., 2017) as far as simple observation is concerned (digital ethnographic research often uses unobtrusive methods in conjunction with direct elicitation methods both because of the centrality of the dialogue with the subjects to be studied and because of ethical issues; see Ugoretz, 2017). The methodological roots of this approach can be found in classical simple observation (Lee, 2000).

Our proposal is summarized in Table 20.1.

### **How to Operate a Distinction: The Issue of Intentionality**

We proposed dividing online WPLB data into three different categories (Table 20.1). On the one hand, online captured data are easy to identify by their very nature: they are social interactions that occur online that researchers can observe as they are in progress. Online found data and online retrieved data are both persistent data stored on the Net that researchers encounter online at a later stage. Online found data and

---

<sup>2</sup>Covert and passive observation in online environments is also known as "lurking" (Ugoretz, 2017).

**Table 20.1** What people leave behind online. *Three categories of unobtrusive digital data: A proposal*

|                                | What people leave behind online                                 |  |  |
|--------------------------------|---|--|--|
| Category                       | Online found data   | Online retrieved data  | Online captured data   |
| Research object                | Digital traces  | Web-mediated documents   | Online behaviours  |
| Main examples                  | Log files, links, mouse clicks, feeds, search requests, cookies | Images, texts, videos, audio files, hypertexts published online by users | Gestures, behaviours, non-verbal cues, in situ conversations |
| Nature                         | Remnants of other activities                                    | Communicative acts   | Social interactions  |
| Quality                        | Inadvertent data  | Intentional data   | Ephemeral data   |
| Research approach              | Online trace analysis (Lee, 2000)                               | Web-mediated documentary analysis (Arosio, 2010)                         | Netnography (Kozintetz, 2010; Costello et al., 2017)         |
| Roots in classical methodology | Trace analysis (Webb et al., 1966; Kellehear, 1993)             | Documentary analysis (Prior, 2003; Scott, 2006)                          | Simple observation (Webb et al., 1966)                       |

online retrieved data can easily be mistaken for each other and require an element to operate a distinction. We focus on them.

Following the definition used in section “Three Categories of WPLB (Online) Data”, the distinctive element between found data and retrieved data is intentionality. While the former are inadvertent data, the latter are intentional data. Digital traces are left as a residue of other activities; they do not have a purpose of their own. Web-mediated documents are created to achieve a purpose (communication). The element of distinction between traces and documents is intentionality, understood as the will of the subjects to carry out that action, giving it a purpose.

Intentionality is not understood as “awareness”, and leaving traces does not necessarily imply lack of awareness. Individuals may be more or less conscious of leaving online traces. Increasing attention to privacy issues and online security discourses has increased people’s awareness that online data can be recorded and stored on the Net. Scandals that in recent years have received widespread media coverage have highlighted that major technology companies have access to personal information and may use it for commercial and political purposes, and people have no control over these actions.

We can illustrate these concepts using the example of email, a widespread form of online communication that can be analysed by researchers to gather information about contemporary societies (for a review, see Perer et al., 2006). Whenever email messages are sent, evidence of the activities carried out by individuals and groups is left behind. On the one hand, there are the texts of the emails. These are materials that subjects and groups intentionally create and disseminate for communication purposes. The message contained in the email is the main purpose of the action. They are “web-mediated documents”: personal web-mediated documents if sent for private purposes and institutional web-mediated documents if sent for organizational purposes (Arosio, 2010). On the other hand, whenever emails are sent and received,



**Table 20.2** Email communications between web-mediated documents and digital traces

| Content of email messages | Communicative act   | Intentional | Main purpose of the action | Web-mediated documents | Messages and points of view                   |
|---------------------------|---------------------|-------------|----------------------------|------------------------|---|
| Email log files           | Involuntary remnant | Inadvertent | Leftover of the action     | Digital traces         | Frequency and directionality of communication |

automatic systems generate and store log files that contain information about the sender and the recipient and the date and time of the messages, depending on the protocol used (Janetzko, 2017). Email log files are remnants of another activity because they are not the primary purpose of the action (the subjects sit at a computer to send a message, not to generate log files). Email log files are therefore digital traces (Lee, 2000).

As illustrated in Table 20.2, the aim of email communications is to send and receive the content of the message, and the creation of email log files does not have intentionality, as it is not the main purpose of the action but rather a leftover. Subjects leave traces without attributing a meaning to them. Digital traces and documents are both useful in research; they just have different characteristics and provide different information. Email log files capture the frequency and directionality of communications and allow networks of relationships to be reconstructed. Email texts capture messages and points of view.

Even though intentionality is a central feature in understanding the very nature of data, it cannot always be easily identified. Arosio (2021) discusses the example of self-tracking data, where the level of intentionality is not self-evident. Self-tracking data may be the result of an unintentional action, as in the case of subjects wearing a mobile device with tracking functions enabled by default. In other cases, self-tracking data can have a communicative intent, for example, when data are collected to be shared by subjects on their social networks or even used as a form of artistic expression. On the surface, the data appear to be the same, but in fact, they are very different. In the first case, researchers are dealing with digital traces, and in the second, they are dealing with web-mediated documents. A major work of reconstruction of the context of production and use of (digital) data is therefore required (Arosio, 2021).

## Why Operate a Distinction?

Every day, people pass through social environments in living their lives: they work, study, go on vacation, compose songs, take a walk in the park, cook or visit a museum. The same happens for groups and organizations. Each of these activities leaves something behind: an attendance sheet, a legal act, a textbook, a souvenir photo, crushed grass, discarded objects or worn tiles. These are all sources of

information that can be used by researchers to understand behaviours, attitudes and values.

Today, due to the widespread use of new communication technologies, many activities are performed in digital environments, and much evidence is left behind online. WPLB online data can offer a relevant contribution to the study of social reality in contemporary societies (among others, see Back & Puwal, 2012; Ruppert et al., 2013; Lupton, 2015). On the one hand, the use of the Internet as a source of information represents a challenge that social researchers need to address to capture the dynamics of social change. On the other hand, there is a need to develop a solid methodological framework before using digital data for social research. The innovative features of these data need to be known, as do the characteristics that connect them to the classical methodological repertoire (see among the others, Amaturio & Aragona, 2021).

This work offers a contribution in this direction. First, it reconnects WPLB online data to traditional unobtrusive measures. Second, it stresses the complexity and varied features of online unobtrusive measures. We focus on the difference among three different categories of online data: online behaviours, digital traces and web-mediated documents. In doing so, investigating the context of production, dissemination and use proved to be a key point. Contextualizing data is not easy, especially unobtrusive measurements and digital data, because unobtrusive measures and online environments reduce the degree to which the researcher has control over the type of data collected (Trochim, 2006). This difficulty does not detract from the need for the nature and limitations of data to be understood before they are used.

Distinguishing different types of online WPLB data is not just a purely terminological issue. Rather, it has important theoretical and methodological implications. Online behaviours, digital traces and web-mediated documents can all provide deep insights into individuals, groups and society. However, they offer different perspectives and different pieces of information. For example, whereas traces are automatically generated by a computer system, online documents are the result of a communicative will. They contain an interpretation, a subjective point of view on reality, which is affected by the author, recipient, purpose and circumstances of sending the message. In this sense, documents should be critically investigated before being analysed (Scott, 1990; Cohen et al., 2000).

The emergence of various categories of digital data calls for reflection on other circumstances. Unobtrusive online collection methods are generally thought of as a source of big data. However, even in this case, digital data can be differentiated. Digital data can be thought of as *big data* (large amounts of data that need to be processed by some kind of data mining/scraping procedures to reveal patterns and trends) or as *small data* (data that are smaller in volume and format and allow for in-depth study).<sup>3</sup> Digital traces, automatically generated by computer systems in

---

<sup>3</sup>The big data/small data distinction echoes the methodological debate on quantity/quality or correlation/causation. In this context, we cannot address the issue extensively. For a first discussion, see Cows and Schroeder (2015).

huge quantities over an extended period, can be interpreted as big data. The simple observation of social behaviours on the Net or the analysis of documents such as blogs and social network pages comes closer to the definition of small data (Lindstrom, 2016).

Ethical issues, which are central to the use of unobtrusive online data, as well as key issues such as privacy, informed consent and the use of data for research purposes should also be explored in more detail when considering the three categories of WPLB online data.

Properly identifying the different natures of digital data has many implications. We have mentioned some of them, and many others need to be explored further. Our remarks are intended as a starting point to develop a deeper understanding of WPLB online from a methodological point of view.

## References

- Amaturo, E., & Aragona, B. (2021). Digital methods and the evolution of the epistemology of social sciences. In P. Mariani & M. Zenga (Eds.), *Data science and social research II. Methods, technologies and applications* (pp. 1–8). Springer.
- Appadurai, A. (1986). *The social life of things: Commodities in cultural perspective*. Cambridge University Press.
- Arosio, L. (2010). Personal documents on the internet: what's new and what's old. *Journal of Comparative Research in Anthropology and Sociology*, 1(2), 1–16.
- Arosio, L. (2021). Between digital traces and documentary analysis: The methodological challenge of self-tracking data. *Journal of Comparative Research in Anthropology and Sociology*, 12(1), 1–10.
- Back, L., & Puwal, N. (2012). A manifesto for live methods: Provocations and capacities. *The Sociological Review*, 60(S1), 6–17.
- Beer, D. (2014). *Punk sociology*. Palgrave Macmillan.
- Boellstorff, T., Nardi, B., Pearce, C., & Taylor, T. L. (2012). *Ethnography and virtual worlds: A handbook of method*. Princeton University Press.
- Boullier, D. (2015). Les sciences sociales face aux traces du Big Data. Société, opinion ou vibrations? *Revue Française de Science Politique*, 65(5–6), 805–828.
- Boyd, D., & Crawford, K. (2012). Critical questions for big data. *Information, Communication & Society*, 15(5), 662–679.
- Cohen, L., Manion, L., & Morrison, K. (Eds.). (2000). *Research methods in education*. Routledge.
- Costello, L., McDermott, M. L., & Wallace, R. (2017). Netnography: Range of practices, misperceptions, and missed opportunities. *International Journal of Qualitative Methods*, 16(1), 1–12.
- Cowls, J., & Schroeder, R. (2015). Causation, correlation, and big data in social science research. *Policy & Internet*, 7(4), 447–472.
- Creswell, J. W., & Plano, C. V. L. (2011). *Designing and conducting mixed methods research*. Sage.
- Denzin, N. K. (1978). *The research act: A theoretical introduction to sociological methods*. Praeger.
- Derrida, J. (1967). *De la grammatologie*. Ed. de Minuit.
- Eynon, R., Fry, J., & Schroeder, R. (2008). The ethics of internet research. In N. G. Fielding, R. M. Lee, & G. Blank (Eds.), *The Sage handbook of online research methods* (pp. 22–41). Sage Publications.

- Ferraris, M., & Martino, V. D. (2018). What is Documediality and why traces, documents and archives are normative. *Law Text Culture*, 22(1), 21–30.
- Given, L. M. (Ed.). (2008). *Unobtrusive research. The SAGE Encyclopedia of qualitative research methods*. Sage Publications.
- Gleick, J. (2011). *The information: A history, a theory, a flood*. Pantheon Books.
- Hand, M. (2016). Persistent traces, potential memories: Smartphone and the negotiation of visual, locative, and textual data in personal life. *Convergence*, 22(3), 1–18.
- Hargittai, E. (2002). Second-level digital divide: Differences in people's online skills. *First Monday*, 7(4), 1–16. <https://doi.org/10.5210/fm.v7i4.942>
- Hine, C. (2008). The internet and research methods. In N. Gilbert (Ed.), *Researching social life* (3rd ed., pp. 304–320). Sage Publications.
- Hine, C. (2011). Internet research and unobtrusive methods. *Social Research Update*, 61, 1–3.
- Janetzko, D. (2017). Nonreactive data collection online. In N. G. Fielding, R. M. Lee, & G. Blank (Eds.), *The Sage handbook of online research methods* (pp. 161–174). Sage Publications.
- Johns, M. D., Chen, S. L., & Hall, G. J. (2004). *Online social research: Methods, issues & ethics*. Peter Lang.
- Joinson, A., Reips, U., Buchanan, T., & Paine Schofield, C. (2010). Privacy, trust, and self-disclosure online. *Human & Computer Interaction*, 25, 1–24.
- Kellehear, A. (1993). *The unobtrusive researcher: A guide to methods*. Allen & Unwin.
- Kneidinger-Muller, B. (2018). Self-tracking data as digital traces of identity: A theoretical analysis of contextual factors of self-observation practices. *International Journal of Communication*, 12, 629–646.
- Kopytoff, I. (1986). The cultural biography of things: Commoditization as process. In A. Appadurai (Ed.), *The social life of things: Commodities in cultural perspective* (pp. 64–92). Cambridge University Press.
- Kozintetz, R. V. (2010). *Netnography: Doing ethnographic research online*. Sage.
- Lee, R. M. (2000). *Unobtrusive methods in social research*. Open University Press.
- Leonelli, S. (2016). *Data-centric biology: A philosophical study*. Chicago University Press.
- Lindstrom, M. (2016). *Small data: The tiny clues that uncover huge trends*. Hodder & Stoughton.
- Lupton, D. (2015). *Digital sociology*. Routledge.
- Lupton, D. (2019). *Data selves: More-than-human perspectives*. Polity Press.
- Mayer-Schonberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work and think*. John Murray.
- Mayo, E. (1949). *Hawthorne and the Western electric company. The social problems of an industrial civilisation*. Routledge and Kegan Paul.
- McCulloch, G. (2004). *Documentary research in education, history and the social sciences*. Routledge.
- McKee, H. A., & Porter, J. E. (2009). *The ethics of internet research: A rhetorical case-based process*. McLaughlin.
- Mills, C. W. (1959). *The sociological imagination*. Oxford University Press.
- Morse, J. (1991). Approaches to qualitative-quantitative methodological triangulation. *Nursing Research*, 40, 120–123.
- Norris, P. (2001). *Digital divide: Civic engagement, information poverty and the internet world-wide*. Cambridge University Press.
- Perer, A., Shneiderman, B., & Oard, D. W. (2006). Using rhythms of relationships to understand E-mail archives. *Journal of the American Society for Information Science and Technology*, 57(14), 1936–1948.
- Pink, S., Ruckenstein, M., Willim, R., & Duque, M. (2018). Broken data: Conceptualising data in an emerging world. *Big Data & Society*, 5(1), 1–13.
- Prior, L. (2003). *Using documents in social research*. Sage Publications.
- Reigeluth, T. (2014). Why data is not enough: Digital traces as control of self and self-control. *Surveillance & Society*, 12(2), 243–254.

- Ruppert, E., Law, J., & Savage, M. (2013). Reassembling social science methods: The challenge of digital devices. *Theory, Culture & Society*, 30(4), 22–46.
- Scott, J. (1990). *A matter of record: Documentary sources in social research*. Polity Press.
- Scott, J. (2006). *Documentary research*. Sage Publications.
- Sechrest, L. (1971). Unobtrusive measures in data collection. *Academy of Management Annual Meeting Proceedings*, 1, 58–66.
- Smelser, N. J. (2013). *Comparative methods in the social sciences*. Quid Pro Books.
- Snee, H. (2013). Making ethical decisions in an online context: Reflections on using blogs to explore narratives of experience. *Methodological Innovations Online*, 8(2), 52–67.
- Sontag, S. (1977). *On photography*. Penguin Books.
- Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology: Combining qualitative and quantitative approaches*. Sage Publications.
- Trochim, W. M. K. (2006). *Unobtrusive measures*. Retrieved March 31, 2022, from <http://www.socialresearchmethods.net/kb/unobtrus.php>
- Ugoretz, K. (2017). *A guide to unobtrusive methods in digital ethnography*. Retrieved March 31, 2022, from [https://www.researchgate.net/publication/331287677\\_A\\_Guide\\_to\\_Unobtrusive\\_Methods\\_in\\_Digital\\_Ethnography](https://www.researchgate.net/publication/331287677_A_Guide_to_Unobtrusive_Methods_in_Digital_Ethnography)
- van Dijck, J. (2014). Datafication, dataism and dataveillance: Big data between scientific paradigm and ideology. *Surveillance & Society*, 12(2), 197–208.
- Webb, E. J., Campbell, D. T., Schwartz, R. D., & Sechrest, L. (1966). *Unobtrusive measures: Nonreactive research in the social sciences*. Rand McNally.
- Webb, E. J., Campbell, D. T., Schwartz, R. D., Sechrest, L., & Grove, J. B. (1981). *Nonreactive measures in the social sciences*. Houghton Mifflin.
- Weber, M. (1922). *Wirtschaft und Gesellschaft. Grundriß der verstehenden Soziologie*. Mohr.

**Laura Arosio**, PhD is an Associate Professor in Sociology in the Department of Sociology and Social Research at the University of Milano Bicocca, Italy. She has a PhD in Sociology. Her main interests concern processes of family formation and transformation in contemporary Italy, with a focus on homogamy and social heterogamy, marriage, marital disruption and its consequences. She has worked on rites of passage and divorce ceremonies from a life course perspective. She is interested in the field of methodology and empirical research (both qualitative and quantitative). Her recent methodological studies focus on documentary analysis, unobtrusive measures, digital data and the implications of using the Internet in social research.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

