

---

JOURNAL OF COMPARATIVE RESEARCH IN  
ANTHROPOLOGY AND SOCIOLOGY

Copyright © The Author, 2021  
Volume 12, Number 1, Autumn 2021  
ISSN 2068 – 0317  
<http://compaso.eu>



## Between digital traces and documentary analysis: The methodological challenge of self-tracking data

Laura Arosio<sup>1</sup>

### **Abstract**

*Trace analysis is a particular form of non-participant observation. It studies the physical and digital "footprints" left by individuals in the environments where they live. In this paper I use the classic and more restrictive definition of trace analysis and focus on intentionality as a feature that marks a boundary between trace analysis and other research methods. By doing this, I stress out the situationality and contextuality of (digital) data's meaning. I use an example involving digital data collected by individuals through mobile and wearable devices. The focus is on running routes mapped out using activity tracking apps. What happens when "traces" are intentionally created by users and shared with other people through social networks? What is the methodological challenge proposed by self-tracking measures when people attribute them a communicative intent? What is the boundary between trace analysis and documentary analysis?*

### **Keywords**

*Trace analysis; Documentary analysis; Intentionality; Digital data; Self-tracking data; GPS art; Big data;*

---

<sup>1</sup> PhD Associate Professor in Sociology, University of Milano Bicocca (Italy) [laura.arosio@unimib.it](mailto:laura.arosio@unimib.it).

## What is (digital) trace analysis

Trace analysis is a particular form of non-participant observation. It studies how environments are transformed by individuals passing through and leaving traces (Webb *et al.*, 1966; Kellehear, 1993; Lee, 2000). Traces may be produced through “accretion” or “erosion”. Accretion traces are left when subjects add something to the environment they visit; erosion traces are left when subjects remove something from the environment (Webb *et al.*, 1966). Trace analysis is designed to study the ways in which environments are experienced and to reconstruct formal and informal rules that guide social activities, along with norms and values that inspire them.

Trace study is an unobtrusive research technique. The research team do not interact with the subjects being observed (Lee, 2000). Data production is not connected to research purposes because data is not solicited from researchers. For these reasons, the study of traces is considered a non-reactive research practice, such as covert observation and documentary analysis. Precisely because there is no direct contact between the researchers and the observed ones, subjects do not change their behaviors as a result of being part of the research (the so-called Hawthorne effect, referring to Mayo, 1949). In contrast, in reactive methods such as interviewing and non-hidden observation, subjects may adapt their behaviors as they know they are involved in a study (Given, 2008).

The observation of traces can be done in places where interaction occurs in a face-to-face mode (such as in a park, a museum or a street), or where interaction is mediated by new technologies (above all, the Internet). In the first case we are dealing with physical traces, in the second case with digital traces (Lee, 2000). From a conceptual point of view, there is great continuity between physical traces and digital traces. In the classic study by Webb *et al.* (1981), it is suggested that noseprints on the glass of museum displays can give insight into visitors' interests. Similarly, electronic footprints left by Web users as they move through the contents of an online exhibit can be used to study what interests people on the Internet (Hine, 2011). Likewise, when users update their social profiles, look up a term on a search engine or navigate from one page to another on the Net, they leave digital traces. These traces give information about behaviors, tastes and interests.

Nowadays digital traces are increasingly widespread because of the growing use of digital devices and applications (Estrin, 2014; Li *et al.*, 2011; Ugoretz, 2017). They have made available a large amount of behavioral data that were previously very difficult and expensive for social scientists to collect (Veltri, 2019). The study of this kind of data can give a lot of information about contemporary societies because, as some scholars argue, the collection of digital traces can be seen as a kind of individual and social memory (Hand, 2016) or identity (Reigeluth, 2014).

Given the high interest surrounding digital traces as research data, it is important to pay great attention to them from a theoretical and methodological point of view. This article points out that digital data and digital traces cannot be treated as something that is abstract and has a unique value. Instead, their meaning depends on the circumstances of their production, creation and dissemination.

The issue of contextuality of data – even more in the case of digital data - is central to the understanding of their meaning (Boyd & Crawford, 2012; Leonelli, 2016). Considering the contingencies of data production and use it allows researchers to access the very nature of the data. For example, it makes it possible to distinguish between digital traces and online documents, as in the case of running routes mapped by activity tracking apps. This is precisely the case being discussed and exemplified in this article.

### **The issue of intentionality: traces vs documents**

In this article, I refer to the seminal and more restrictive definition of trace analysis first introduced by Webb and colleagues in 1966 (Webb *et al.*, 1966) and then used in subsequent methodological works (Webb *et al.*, 1981; Kellehear, 1993; Lee, 2000). Trace analysis is presented as one of a variety of unobtrusive measures, including non-participant observation, content analysis, archival work and use of documents<sup>2</sup>.

According to this definition (Webb *et al.*, 1966), traces are referred to as the residue of other activities (“*remnant*”), what is left behind while people are busy doing something else. Subjects leave traces without attributing meaning to them, and often without being aware of it. Thus, traces lack direct or indirect communicative purposes. Subjects do not use traces to send a message, to communicate their ideas or to convey information. For example, subjects intend to spend a day outdoors: the footprints left on the lawn at the end of the excursion are a residual outcome, which is without communicative intent and left unintentionally. Similarly, subjects wish to make a purchase and complete an electronic payment: the digital traces are a residue of this operation, which is left without intentionality and without full awareness<sup>3</sup>.

Because traces are unintentional, their study is different from other unobtrusive research practices. In particular a distinction can be made with documentary analysis. Documents are spontaneous products that are created by individuals and groups without research purposes (Scott, 2006; McCulloch, 2004). There are “traditional” and “web mediated” documents (Arosio, 2010).

Unlike traces, documents are intentional products. Individuals and groups produce documents for a purpose (and purpose distinguishes documents into: personal, institutional, media, and cultural documents). The ultimate purpose of documents is communication. Documents contain ideas, points of view, interpretations to be transmitted (Prior, 2003; Scott, 2006; Arosio, 2010). Referring to the previous examples, graffiti left on the picnic table during the trip out are documents, as are comments on a product left by users on an e-commerce website. Both are created intentionally and have a communicative purpose.

---

<sup>2</sup> Nowadays, the term “traces” tends to be used extensively and to some extent has become synonymous with “unobtrusive measures”, including non participant observation, content analysis, archival work and use of documents. In my opinion it is worth keeping the more restrictive definition because it has important theoretical and methodological implications, to be discussed in the next pages.

<sup>3</sup> People are increasingly aware that their use of digital technologies can leave traces. This supports the idea that there are different degrees of intentionality in the creation of digital data, which need to be studied.

This difference between traces and documents has important methodological consequences. The message of documents is shaped by the context of production and contains a point of view about reality (Prior, 2003). Through the production and dissemination of documents, individuals and groups want to convey an image of reality. Therefore, the content of documents is affected by the subjectivity of those who produce or commission them. The document's audiences and purposes also influence their content. Therefore, documents need to be contextualized in order to be understood. Specifically, the author, recipient, commissioner, language, and purpose of documents must be reconstructed (Scott, 1990). Through this critical analysis, the message of the document can be received (Cohen *et al.*, 2000).

In contrast, traces have no intentionality and therefore contain no interpretation. They are not designed to convey specific content. Traces do not require the critical questioning required by the documents to interpret the point of view and message they contain (Prior, 2003). On the other hand, traces need to be placed in their context of production to understand the information they may provide (Kneidinger-Muller, 2018). The contingencies surrounding the creation and use of digital data shape their meaning and therefore need to be known (Pink *et al.*, 2018).

### **Wearable devices and self-tracking data: traces or documents?**

In this paper I want to explore the existing boundary between (online) trace analysis and (web mediated) documentary analysis relying on the characteristic of intentionality. This line is not always easy to draw. By doing this, the importance of considering (digital) data in their context of production and use will be highlighted.

A research example concerning data collected through mobile and wearable devices is proposed. These are watches, bracelets, glasses, and other high-tech miniaturized electronic objects connected to the Internet and operating through dedicated applications. Constant contact with the body makes wearable devices particularly suitable for medical and sports purposes. These devices and their increasing popularity make possible the collection, storage and analysis of a large amount of data about the subjects who wear them.

Wearable devices include activity trackers, which monitor and record body metrics such as heartbeats, calories burned, body temperature, sleep quality. Some devices gather data about environmental conditions such as air temperature, humidity, air quality. Data can be stored on websites where they can be shared and compared with other users (Lupton, 2014).

The concept of self-tracking has recently emerged in sociological literature. Self-tracking has been defined as the practice of gathering data about oneself on a regular basis and then recording and analyzing the data to produce statistics relating to habits, behaviors and feelings (Lupton, 2014).

It is argued that self-tracking is not an individualistic practice, but a social one. It is promoted by social institutions (cultural, medical, educational, sports, insurance institutions) and makes sense within socially constructed meanings. Moreover, many self-

trackers use social media to compare and share personal data, and feel part of a community (Boesel, 2013; Nafus & Sherman, 2014). Social, cultural and political dimensions of the phenomenon of self-tracking have been studied over the past decade (among the others, Lupton, 2014 & 2018; Maturo, 2015; Nafus & Sherman, 2014; Pantzar & Ruckenstein, 2014; Whitson, 2013).

In social research, self-observation data can be used by researchers to obtain a lot of information about contemporary societies. Self-tracking practices can be considered as a new big data source, in that they gather a large amount of detailed data on individuals, on a broad range of aspects of their lives, rapidly and for a long period (Almalki *et al.*, 2013). At the same time, the epistemological and material uncertainty of these data must be considered. The incompleteness, inaccuracy and dispersed nature of personal self-tracking data has been noted, by using the concept-metaphor of “*broken data*” (Pink *et al.*, 2018). This issue supports the need for a more thorough analysis of the data before researchers can use them for scientific purposes.

From a methodological point of view, self-tracking data tends to be considered as digital traces (Kneidinger-Muller, 2018; Lupton, 2014), although it is recognized that they are the result of an intentional and active process of production (Lupton, 2014 & 2018; Li *et al.*, 2010)<sup>4</sup>.

### **Different degrees of intentionality in self tracking data**

So, should self-tracking data be considered traces? By emphasizing the characteristic of intentionality, I point out that under different conditions self-observation data can be considered either traces or documents. By doing this, the difference between traces and documents is stressed.

In order to make this point clear I take a closer look at running routes, a visual type of self-tracking data that comes from running apps. These are tracking apps designed to collect a large number of metrics about running activity. They are used both by beginners and competitive runners to create a detailed analysis of their running experience<sup>5</sup>. The basic functions involve collecting statistics such as time, distance, speed. The apps often offer training programs, compare runners' performances with those of others, and motivate achievements through virtual awards and medals.

One function of running apps is to track the path completed by runners and map it out with the help of a GPS system (Global Positioning System). The result is an image of the running route that can be saved, stored on electronic devices and shared with others.

Digital routes drawn and recorded by running apps can be considered as electronic traces, the evolution of physical footprints left on the path as the athlete runs along it. I think that this is not always the case.

---

<sup>4</sup> Lupton (2014) suggested that there are five modes of self-tracking: private (for one's own purposes); communal (sharing data with others); pushed (encouraged by others); imposed (forced by others); and repurposed (for the use of others).

<sup>5</sup> Among the most popular running apps see Runtastic, Runkeeper, Strava and Nike+ Run Club.

From a closer inspection of the phenomenon, there are many modes to create and use running routes<sup>6</sup>. Not only are the maps displayed by the apps, but as I have already mentioned they can be saved on personal devices or on specialized websites. In addition, screenshots containing performance measurements and the running routes can be shared on social media (for example, on user's Instagram pages). Even more, before starting their workout runners can imagine paths to be covered so that during the run the device draws selected images on the maps.

I have attempted to bring order to these different ways of creating and using running routes and propose four different modes. Figure 1 contains and illustrates my proposal. It is centered on the different levels of intentionality in the data creation and dissemination process. I think this proposal can be useful to distinguish amongst different types of running maps, self-tracking data and digital data. In particular, it allows to separate digital traces and web mediated documents.

No intentionality			Maximum Intentionality
1	2	3	4
Unintentional geolocation	Private use	Shared use	Artistic use

Figure 1 - Different levels of intentionality in mapping out running routes

A subject may be wearing a mobile device, which has GPS functions enabled by default. The device collects and maps out data without the individual being aware of it (**case 1**): running maps have no meaning for the subject and no communicative purpose.

A person may wear a mobile device and intentionally activate the tracking function for personal purposes (**case 2**): running maps are used by runners to monitor their activity and gather information about their own performance.

A subject may wear a mobile device and intentionally activate the tracking function to collect and share data with others (e.g., Instagram pages) (**case 3**). Running maps are used by runners to communicate their passion for the sport, build their identity, and give meaning to their actions.

A subject may wear a mobile device and intentionally activate the tracking function to draw shapes and images on the map (**case 4**). These drawings are then published and shared on social networks: the runner is an artist that follows a pre-planned route to create a large-scale picture (also known as GPS drawing or GPS art<sup>7</sup>).

In **case 1**, there is no intentionality and there are no communicative purposes, so we are dealing with digital traces. In cases 2-4 there is intentionality (to varying degrees) and a message to be passed, so we are working with documents. Because running maps are mainly intentional, they are often to be considered documents.

<sup>6</sup> This categorization derives from an original study of the Author, supported by an extensive inspection of running maps published online.

<sup>7</sup> See for example <https://www.pinterest.com/scarletfireuk/gps-art/>

This example supports my opening point. Self-tracking data can be traces or even documents. It depends on the ways data is created and disseminated. Levels of intentionality and communicative purposes are very relevant features to make this distinction. The situationality of data affects their nature and meaning.

## Discussion

I focused on the case of self-tracking data because they pose an interesting methodological challenge. They are usually considered "*digital traces*", but they are often intentionally created and collected by individuals. Moreover, they can be shared with other people to communicate a message and also to create an artwork. These practices give rise to a methodological oxymoron, i.e., intentional traces (the very definition of traces is the unintentional residue of other actions, as stated in sections 1-2).

More exactly, tracking practices uses digital traces for communicative purposes, fills them with intentionality, and turns them into web mediated documents. So, the issue of reactivity is introduced and the need to be critically questioned should be considered.

When researchers do encounter self-tracking data and plan to use them to study social reality, they must first determine whether they are dealing with traces or documents. To do this, the degree of intentionality is crucial. More generally, all the steps through which data were produced, gathered and shared need to be known and all the actors and forces that have shaped them need to be studied.

From the empirical research perspective, the task to be faced is not easy. While documents and traces can look similar on the surface level, intent is difficult to be determined and measured. The subjects themselves would probably not be able to measure their level of intentionality. Intentionality may also be concerned with the passage of time. On the one hand, subjects may intentionally turn on devices and over time may "*forget*" that data is being collected. On the other hand, subjects are increasing aware that their digital engagement leave traces somewhere. An in-depth study of the context and contingencies surrounding the production and use of the data could provide important insights. But, even this study is not easy to conduct, sometimes proving impossible.

The case of self-tracking data can be used as an example to be extended to other (digital) data. The data that is stored on the Net is becoming an important source of information to study contemporary societies, as the Internet is a digital environment where individuals are spending an increased amount of their time on and performing an increased amount of activities on.

A preliminary analysis of the context in which digital data are produced is necessary to understand their very nature, for example to understand whether they are traces or documents. Intentionality can be a critical element. Establishing the nature of the information collected allows researchers to understand what it can be used for and to what extent.

Technical implications are also important. In the case of (digital) traces, it is necessary to reconstruct the context to understand what the data may be saying. In the

case of (online) documents, critical questioning is also necessary to bring out the subjective viewpoints contained in the messages.

Not all the questions being raised in this article can be answered, nevertheless it eventually opens a theoretical and methodological space for further investigation. At least, the importance of data contextualization is emphasized by our discussion.

### **Acknowledgements**

The Author would like to thank the two anonymous referees for their valuable comments.

### REFERENCES

- Almalki, M., Martin-Sanchez, F. & Gray, K. 2013. *The Use of Self-tracking Devices: Big Data Prospects and Challenges*, HISA Big Data 2013 Conference; Australia, Melbourne. [Online] <https://bit.ly/3rOKXNL>.
- Arosio, L. 2010. Personal documents on the Internet: what's new and what's old. In *Journal of Comparative Research in Anthropology and Sociology*, 1(2), pp.1-16. [Online] <https://bit.ly/35ow6GW>.
- Boesel, W. E. 2013. *What is the quantified self now?* [Online] <https://bit.ly/3JAKLML>.
- Boyd, D. & Crawford, K. 2012 Critical questions for Big Data. In *Information, Communication & Society*, 15(5), pp.662-679.
- Cohen, L., Manion, L. & Morrison, K. (eds.) 2000. *Research Methods in Education*. London, Routledge.
- Estrin, D. 2014. Small data, where n = me: Seeking personalized data-derived insights from analysis of our digital traces. In *Communications of the ACM*, 57, pp.32-34.
- Given, L. M. (ed.). 2008. Unobtrusive Research. In *The SAGE Encyclopedia of Qualitative Research Methods*. London: SAGE Publications.
- Hand, M. 2016. Persistent traces, potential memories: Smartphone and the negotiation of visual, locative, and textual data in personal life. In *Convergence*, 22(3), pp.1-18.
- Hine, C. 2008. The Internet and research methods. In Gilbert, N. (ed.). *Researching Social Life*, 3rd edition, London: Sage.
- Hine, C. 2011. Internet Research and Unobtrusive Methods. In *Social Research Update*, (61), 1-3. [Online] <https://sru.soc.surrey.ac.uk/SRU61.pdf>.
- Kellehear, A. 1993. *The Unobtrusive Researcher: a guide to methods*. St Leonards, NSW, Allen & Unwin.
- Kneidinger-Muller, B. 2018. Self-Tracking Data as Digital Traces of Identity: A Theoretical Analysis of Contextual Factors of Self-Observation Practices. In *International Journal of Communication*, 12, pp.629-646.
- Lee, R. M. 2000. *Unobtrusive Methods in Social Research*. Buckingham: Open University Press.
- Leonelli, S. 2016. *Data-Centric Biology: A Philosophical Study*. Chicago: Chicago University Press.

- Li, I., Dey, A., & Forlizzi, J. 2010. A stage-based model of personal informatics systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Atlanta: ACM, 557-566.
- Li, I., Dey, A., & Forlizzi, J. 2011. Understanding my data, myself: Supporting self-reflection with ubicomp technologies. In *Proceedings of the 13th International Conference on Ubiquitous Computing*, pp.405-414. New York, NY: Association for Computing Machinery.
- Lupton, D. 2014. Self-tracking cultures: Towards a sociology of personal informatics. In *Proceedings of the 26th Australian Computer-Human Interaction Conference on Designing Futures: the Future of Design*, pp. 77-86, DOI:10.1145/2686612.2686623.
- Lupton, D. 2018. Lively data, social fitness and biovalue: the intersections of health self tracking and social media. In Burgess, J., Marwick, A. & Poell, T. (eds.) *The Sage Handbook of Social Media*.
- Maturo, A. 2015. Doing Things with Numbers. The Quantified Self and the Gamification of Health. In *Journal of Medical Humanities & Social Studies of Science and Technology*, 7(1), pp.87-105.
- Mayo, E. 1949. *Hawthorne and the Western Electric Company. The Social Problems of an Industrial Civilisation*. London: Routledge and Kegan Paul.
- Macdonald, K. & Tipton, C. 1993. Using documents. In Gilbert, N. (eds.) *Researching social Life*. London: Sage.
- McCulloch, G. 2004. *Documentary Research in Education, History and the Social Sciences*. London and New York: Routledge.
- Nafus, D. & Sherman, J. 2014. This one does not go up to 11: The Quantified Self movement as an alternative big data practice. In *International Journal of Communication*, 8, pp.1785-1794.
- Pantzar, M. & Ruckenstein, M. 2014. The heart of everyday analytics: emotional, material and practical extensions in self-tracking market. In *Consumption Markets & Culture*.
- Pink, S., Ruckenstein, M., Willim, R. & Duque, M. 2018. Broken data: Conceptualising data in an emerging world. In *Big Data & Society*. 1(5), pp.1-13.
- Prior, L. 2003. *Using Documents in Social Research*. London: Sage.
- Reigeluth, T. 2014. Why data is not enough: Digital traces as control of self and self control. In *Surveillance & Society*, 12(2), pp.243-254.
- Scott, J. 1990. *A Matter of Record: Documentary Sources in Social Research*. Cambridge: Polity Press.
- Scott, J. 2006. *Documentary Research*. London: Sage.
- Ugoretz, K. 2017. *A Guide to Unobtrusive Methods in Digital Ethnography*. [Online] <https://bit.ly/3LAVbct>.
- Veltri, G.A. 2019. *Digital Social Research*. Polity Press
- Webb, E.J. et al. 1966. *Unobtrusive Measures: Nonreactive Research in the Social Sciences*. Chicago: Rand McNally.
- Webb, E.J. et al. 1981. *Nonreactive Measures in the Social Sciences*. Dallas, TX: Houghton Mifflin.
- Whitson, J. 2013. Gaming the quantified self. In *Surveillance & Society*, 11(1/2), pp.163-176.

**Laura Arosio** is 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 (University of Milano). Her research works concern families, partner selection and marital instability in contemporary societies. She is very interested in the field of methodology and empirical research (both qualitative and quantitative). Her recent studies focus on documentary analysis, non-intrusive measures, and the implications of using the Internet in social research.