

The Complexity of Datafication: Putting Digital Traces in Context

Andreas Breiter and Andreas Hepp

16.1 INTRODUCTION

A prominent characteristic of deep mediatization is an ongoing datafication. This means that in a moment when more and more media become digital they are not only means of communication but increasingly also of generating data. These data can be used for very different purposes. The basis of this datafication are ‘digital traces’. Whatever users do, as soon as they live in this highly mediatized social world they leave ‘foot-prints’ of their digital media use that build ‘digital traces’. Partly, users do this consciously, for example by uploading photographs or writing comments on the ‘time lines’ of digital platforms. But often users are not aware of it and it ‘happens’ as an (unintended) side effect of our media-related

The original version of the book was revised: Incorrect reference and corresponding reference citation have been corrected. The erratum to the book is available at https://doi.org/10.1007/978-3-319-65584-0_19

A. Breiter (✉) · A. Hepp
ZeMKI, Centre for Media, Communication and Information
Research, University of Bremen, Bremen, Germany
e-mail: abreiter@uni-bremen.de

A. Hepp
e-mail: ahepp@uni-bremen.de

© The Author(s) 2018
A. Hepp et al. (eds.), *Communicative Figurations*,
Transforming Communications – Studies in Cross-Media Research,
https://doi.org/10.1007/978-3-319-65584-0_16

activities. This is, for example, the case when using a search engine or when reading newspapers online, where only a limited group of users are aware of the scope of related traces and their further use, for example in the advertising industry (Turow 2011). But digital traces go even further: they are not just made by the users themselves but also by others when they interact online with reference to them, for example by synchronizing their address books with our digital addresses, by tagging pictures, texts or other digital artefacts with the names of other users. Digital traces nowadays even begin before the date of birth and beyond death. One example for this is the ‘mediatization of parenthood’ (Damkjær 2015), which results in processes of constructing ‘parenthood’ before birth, as pregnancy is accompanied with an ongoing flow of communication via apps and platforms that produces digital traces of a ‘forthcoming child’. Then the question ‘*who* is allowed to leave these traces of an even unborn?’ becomes an issue in a kind of family communication policy. In such a sense, as individuals, collectivities or organizations ‘we cannot not leave digital traces’ (Merzeau 2009: 4) in times of deep mediatization. Therefore, datafication reflects an increasing complexity of the social world by adding a new level of social construction that is delegated to algorithms and software.

Methodologically speaking, the emergence of such kinds of digital traces is a problem for empirical media and communication research. Existing research on datafication shows that one problem is the access to such kinds of data. In many cases, the application program interfaces (APIs) which open access to this kind of data are controlled by companies in outstanding power positions, such as Apple, Twitter, Facebook or Google. However, even if such an access is given, yet another problem arises. How can this data be put into context in a way that one is able to analyze it in a socially meaningful way?

In this chapter, we want to deepen the discussion of this second problem of contextualizing digital traces. First, we will reflect on digital traces as a phenomenon of complexity more generally. Then we will take the example of data from learning management systems to discuss possible strategies of how to put such automatically generated data into context by the use of qualitative methods that become triangulated. On such a basis, we finally want to draw some conclusions about the future challenges of this kind of research. Overall, this chapter can only argue in an exemplary way, taking a specific and thus limited case of analysis. But we hope that our more detailed discussion makes it possible to outline different options for future methodological developments in media and communication research.

16.2 DIGITAL TRACES AS A PHENOMENON OF COMPLEXITY

Understanding digital traces as the sequence of ‘digital footprints’ which are left by the use of digital media and services represents quite a new area of media and communication research. At the same time, we can refer this back to more prolonged discussions about whether ‘new’ media also require ‘new’ methods of research (see for example Golding and Splichal 2013; Hutchinson 2016), and have to contextualize it in the much more far-reaching discussion surrounding ‘digital humanities’ and its methods (Baum and Stäcker 2015; Gardiner and Musto 2015). As a phenomenon, digital traces have evoked a sophisticated but also controversial methodological discussion (Kitchin 2014). In this respect, we can notice a multiple complexity of the phenomenon.

First of all, it is important to be aware that they are more than just (big) data. As ‘big data’ is used as ‘a catch-all, amorphous phrase’ (Kitchin and McArdle 2016), it provokes substantial discussions about its capacity. Heavily criticized by one group of scholars (boyd and Crawford 2012; Andrejevic 2014), it is regarded as the future of empirical research by others (Mayer-Schönberger and Cukier 2013; Townsend 2013). Hence, we follow a different direction while discussing some questions of big data later more in detail. Digital traces are a kind of digital data which become meaningful because this sequence of ‘digital footprints’ is in a technical procedure of construction related to a certain actor or action, typically an individual but in principle also a collectivity or an organization. By such procedures of connecting data with entities of the social world they become meaningful information, and this is the reason why companies and other organizations of data processing are highly interested in this kind of data aggregation in relation to ‘real’ people. For the purpose of empirical research, a good starting point is to define digital traces as numerically produced correlations of disparate kinds of data that are generated by practices of individual, collective and corporate actors in a digitalized media environment.¹ The complexity of digital traces is reasoned by the variety of their production, but also the variety of possible correlations.

Recently, digital traces and related possibilities of data generation became an issue of fundamental critique of social science methods; one that we do not share in detail but have to be aware of. The argument at this point is that with increasing datafication, methods of social sciences increasingly entered a ‘crisis’ as digital traces seem to be a much more

proper data source than the kinds of data typically used in social sciences (Savage and Burrows 2007). While the sample survey and the in-depth interview once represented innovative contributions to a methodologically informed description and understanding of the social world, nowadays because of datafication—and hence accessible data sources—they would produce a much more limited access to the procedures of how society is constructed. Its main governing organizations—companies, administrations, educational and government institutions—get much of their information via an ongoing observation and analysis of the various digital traces left by the people. Against such sources, any proposition academic research can produce based on surveys and interviews seems to be flawed. Many established methods would come under pressure with recent datafication as they cannot deliver proper answers to the problems under question, something that is described as the ‘social life of methods’ (Savage 2013: 5). Therefore, we would need to ‘reassemble social science methods’ (Ruppert et al. 2013: 22). A widely discussed conclusion from this is to think about new forms of data collection and analysis that are based on ‘digital methods’ (Rogers 2013: 1, 13). Methods such as crawling, scraping or data mining take digital traces as sources for empirical research. They do not use special procedures for data collection to produce data that is then analyzed; but rather they are methods of using digital traces as a source for analysis.

Some proponents even go one step further, arguing that digital traces would allow for the first time a direct access to ongoing processes of social construction. Maybe the most prominent example is Bruno Latour’s integration of digital traces investigation into his overall approach to social analysis (see Latour 2007). A ‘digital traceability’ (Venturini and Latour 2010: 6) then becomes a possibility for analyzing processes of social construction in situ: ‘Being interested in the construction of social phenomena implies tracking each of the actors involved and each of the interactions between them’ (Venturini and Latour 2010: 5). With digital traces, so the argument, we might have such a direct access, as they would allow us to witness processes of assembling in the moment they take place (see Latour et al. 2012; Venturini 2012).

From our point of view, this move largely misunderstands the main points of digital traces and the complexity of their analysis. First of all, there remains the fundamental problem of misinterpreting the social world as ‘flat’ and therefore as reconstructable solely by an analysis of correlated ‘footprints’ in digital media. This is one point of access which is

non-responsive, but one that reduces the present complexity of the datafied social world to the ontology of a flat society.² Second, and even more fundamentally, such an approach misunderstands digital traces as something ‘neutral’, offering us a ‘direct access’ to society. However, digital traces are not ‘neutral phenomena’; rather, they rely on the technical procedures of governing institutions: the companies, administrations and agencies that produce this kind of data. With governing we mean that these institutions are organizations that are in a powerful position to define the character and structure of data and metadata as well as its possible purposes of use. Actors can access this purposefully constructed and not objective data as individuals (independent workers, civic hackers), collectivities or organizations only in a controlled way. Therefore, as in any established method of social science, digital traces as indicators of social reality have to be critically reflected with regard to their particular perspective and the underlying biases in which they are produced.

Concluding from this, our approach to digital traces refers back to a critique of any naïve understanding of ‘big data’ (cf. Puschmann and Burgess 2014). Especially beyond academic research, there is high hope of the promise of new forms of analysis with reference to a so-called ‘revolution of big data’. The core argument of this hope is that huge amounts of data-based information can be related and analyzed with automated procedures without predefining theoretical assumptions, and at the same time can lay the ground to predict future developments. This would make a new, purely data-oriented knowledge production possible that is partly positioned against theoretically informed forms of academic research. As prominent representatives of big data analysis put it, ‘no longer do we necessarily require a valid substantive hypothesis about a phenomenon to begin to understand our world’ (Mayer-Schönberger and Cukier 2013: 55). Or, as formulated in the sub-title of a best-selling practical guide (Marr 2015), it is about ‘using smart big data, analytics and metrics to make better decisions and improve performance’. In education, ‘learning analytics’ (Ferguson 2012; Papamitsiou and Economides 2014) based on big data become the new vision to control and manage individual learning processes purely by algorithms. Similarly, student assessment data based on psychometric tests are used by administrators to rank schools, incentivize teachers and to create their accountability systems (Anagnostopoulos et al. 2013). But as Perrotta and Williamson (2016) clearly point out, the production of the underlying data structures and algorithms and their construction power in social life are often neglected.

Such an approach reduces the complexity of the phenomenon of digital traces to a ‘big data paradigm’ that is about ‘managing data and transforming it into usable and sellable knowledge’ (Elmer et al. 2015: 3). From the point of view of empirical research methods in social sciences, such hopes are partly based on what we can call a ‘mythology of big data’, that is ‘large data sets offer a higher form of intelligence and knowledge that can generate insights that were previously impossible, with the aura of truth, objectivity, and accuracy’ (boyd and Crawford 2012: 2). This kind of ‘social analytics’ (Couldry et al. 2015) refers back to the ‘gradual normalisation of datafication’ (van Dijck 2014: 198) as a new paradigm in science and society. This is exactly the point where we have to be careful: researchers of big data ‘tend to echo these claims concerning the nature of social media data as natural traces and of platforms as neutral facilitators’ (van Dijck 2014: 199). The idea is that once the easy work of gathering data is completed, the ‘data will speak for itself’ (Mosco 2014: 180). The hope becoming articulated in such a discourse is that big data would offer a possibility to reduce the complexity of analyzing the social. Or put another way: big data is constructed as an easy way to handle the complexity of our datafied social world by datafication.

As we know in the meantime, (meta)data cannot be considered as ‘raw resources’ that offer any direct access to a complex datafied social world (Gitelman and Jackson 2013, 7; Bowker 2014: 1797; van Dijck 2014: 201; Borgman 2015). In contrast, the main methodological task for empirical research on digital traces is to make them meaningful in a social sense, that is to explain the causalities and relations that go beyond pure aggregations and correlations as they are put up by automated collections of data. As a consequence, the methodological challenge for researching transforming communications is less than just an automated analysis of big data, as often postulated: rather, the methodological challenge lies in how to relate digital traces to further sources of data by means of which such traces become validated as well as interpretable and can subsequently be referred to in more sophisticated explanations and procedures of theory building (see Crampton et al. 2013; Lohmeier 2014). We must be very careful to avoid possible misunderstandings at this point. We share the position that competences in new forms of ‘digital methods’ (Rogers 2013) and ‘automatized analysis’ (Neuendorf 2017) are a necessity for media and communication research that endeavours to be up to date, and we subscribe to this discussion about datafication (Hepp 2016: 234–237). This said, we are critical of any approaches that

understand data purely as a direct source for describing the society. We need the combination with further information about the figuration under investigation. Following the semiotic theory, information is data in context referring to its semantics (O'Connor et al. 2001).

16.3 SCHOOL LEARNING MANAGEMENT SYSTEMS AS AN EXAMPLE: ANALYZING DIGITAL TRACES AS PUTTING THEM INTO CONTEXT

If we follow the line of argument up to this point, the main challenge is how we can analyze digital traces in a way that we can contextualize them within the figurations of humans that produce these sequences of 'digital footprints' but also *use* them as a means for social construction. From such a point of view, we have to think about how to relate the 'information' of digital traces to specific actor constellations, frames of relevance, and practices of communication in and by which they are produced. The main examples on which we want to discuss this challenge are data systems as they are nowadays widespread in schools, originally especially in the USA and the UK, but increasingly also in Germany.

School learning management systems as a software define the 'space' in which data are produced as 'digital traces' which, however, are also used by others to subsequently construct social reality. Or put differently: the school information systems are not only the means to 'collect' data; they are also means for powerful processes of construction, typically on the part of their providers who do 'data analysis'. The way in which data are embedded into communicative practices in schools plays an important role: for example, the use of grades for decision-making, the use of upload and download traces to define student involvement or teacher or parent engagement.

Learning management systems (Ifenthaler 2012) in schools and higher education institutions are supposed to support the learning process of students and the management processes of teachers. Most studies reflect the forms of instructional use, teachers' and learners' attitudes, and the impact on learning (e.g. De Smet et al. 2012). But the organizational processes of schools, that is interactions between students, teachers and parents and within their groups, between school management and staff, school district and school board, are often neglected (see Breiter 2014). In an empirical study of German secondary schools headed by one of the authors,³ the goal was to reconstruct the school

as a social organization by analyzing communicative practices of key stakeholders. Hence, online, face-to-face as well as paper-based forms were studied. A subset of our research addressed the interdependence of communication networks between teachers in the world of the school building and in the world of the learning management system. The underlying hypothesis assumed a very similar activity structure inside and outside the technical system: those who interact regularly and intensively will do so online. For this purpose, we collected digital traces that teachers left in the learning management system. As in most server-based systems, the paths of users can be traced back by using log-files. Log-files provide information, problems or errors pertaining to the system and its applications (Markov and Larose 2007; Suneetha and Krishnamoorthi 2009; Liu 2011; Oliner et al. 2012), often in the Extended Common Log-file Format:

Looking at these log-files from a webserver as in Fig. 16.1—here in an anonymized and therefore fictive form—it is possible to identify the user by her internet protocol (IP) (1.2.3.4) and additionally the Browser Operating System combination if multiple users use the same internet connection. Once a user is identified, one can track the movement within the site because the second last entry contains the page the user came from, the so-called referrer. In the example given here, the user enters the site at `index.php`, stays on the site for 14 seconds and moves on to `page2.php` by using a hyperlink. These ‘clicks’ are called actions. Using this information, we can track all movements from all users separately. There are mainly five ways to conduct a log-file analysis: (1) display which pages of a website are accessed more than others and how many users selected a specific function; (2) show paths from visitors through the site; (3) cluster visitors into groups, the clusters being based on movements or paths through the system; (4) social network analysis to identify connections between users and/or websites based on the ‘clickstream’ data; and (5) other statistical methods and algorithms (e.g. multi-level analysis).

Logfile analyses are non-reactive. All information is gathered on the application layer or server layer and not actively put in by the user. Furthermore, data are stored in a machine-readable format and can be used in real time. But there is a main disadvantage of such a strategy for collecting data: the lack of any information about the user’s practices. Furthermore, there is usually no information about socio-demographic data of the user. Additionally, log-files can cause high privacy concerns.

```
1.2.3.4 - - [25/Aug/2011:12:15:33 +0100] "GET /index.php HTTP/1.1" 200 23578 - "Webbrowser (System etc.)"
1.2.3.4 - - [25/Aug/2011:12:15:47 +0100] "GET /page2.php HTTP/1.1" 200 15789 "http://www.domain.com/index.php" "Webbrowser (System etc.)"
```

Fig. 16.1 Example for log-file entries

The users normally have no control over the log-files that are produced by the server or application. As the IP address is stored, the users are easy to identify. Therefore, log-files must be made anonymous by the researchers. But while this is necessary from the point of view of research ethics, it additionally limits the interpretation of such data.

We gathered anonymous data from a learning management system in a larger German secondary school (>100 teachers and >1000 students). The system is mainly used by the staff for coordination and communication. As it is hosted by an external company it can be accessed from inside the school's network and from home. The learning management system offers the following features: announcements, calendar, file exchange and discussion groups.

The log-files investigated by us span over a period of 12 months, including holiday breaks. In the log-files we analyzed, 120,000 hits from 138 users are recorded. After the deletion of all irrelevant data (e.g. by bots) and by using path completion algorithms, the sum of hits is approximately 62,000. The 138 unique users had a total of 4451 visits.⁴ In Fig. 16.2, a network graph of this data is shown.⁵ Such a visualization makes it possible to identify three main groups in the upper part of the graph, which are connected to the categories 'miscellaneous', 'reports' and 'conferences'. All are mainly linked to dates, some to announcements and materials. Announcements and materials are more likely accessed than dates. This is no surprise as dates can be viewed in a calendar-like overview. The items themselves are mainly linked to the category and not linked among themselves.

In the bottom left are many materials closely connected to each other. Above these materials are two subjects—one bigger and one smaller. In contrast to the representation of the former three categories, the nodes are overlapping each other and are not only linked to the subject itself but also to each other. This indicates that the items are closely linked together and due to the force-driven representation. The relative big node size is another indicator for the intensive material exchange within these two subjects.

To deepen such an analysis, we can do a scatter plot of this data. Scatter plots are mathematical diagrams with two coordinates to visualize values of variables. As the points have different sizes, they represent a third variable (in this case uploader). The scatter plot in Fig. 16.3 compares the number of materials per subject and the sum of hits to these materials. The size of each subject shows the number of different contributors. English has the most hits (2300) and the most materials (23). That is no surprise and was already assumed if we refer to the previous data set. Spanish, on the other hand, is more interesting. It has the second most materials (15), but only around 500 hits and only three contributors. Based on the log-files, we can only speculate about the reasons.

As we can see in this example of digital traces in a school information system, the interpretation of so-called big data is only possible with context-specific knowledge. Log-files can give researchers a broad view into an information system and its usage. They do not allow to identify ‘significant behaviour’. Our analyzed data had a time span of about 300 days. There may be the possibility to overlook significant behaviour as the amount of data is large, and significant behaviour must not be the most common behaviour. But statistical methods such as sequential pattern or cluster analysis try to find a common and frequent pattern, not a rare or unique pattern which is potentially more relevant. This may lead to an opposition of available methods and research aims. Additionally, patterns which can be identified statistically need to be embedded in the physical world of classrooms, different staff rooms, subject- and/or grade-related rooms and ‘water coolers’ (Earl 2001).⁶ To understand schools as communicative figurations, we need to identify the actor constellation which will only partly be mapped in the log-files—non-active members of staff and their communicative practices are neglected, even if they might have a media ensemble which allows data exchange.

In our case, we accompanied our quantitative analysis with in-depth qualitative studies based on participant observations and interviews (Welling et al. 2015). Over a period of one school year, we observed teachers in their staff room as well as in subject-specific rooms. Based on an observation protocol, the use of the information system as well as situations and locations for exchange about administrative and organizational issues were recorded and later analyzed with an open coding scheme. The interviews with different groups of

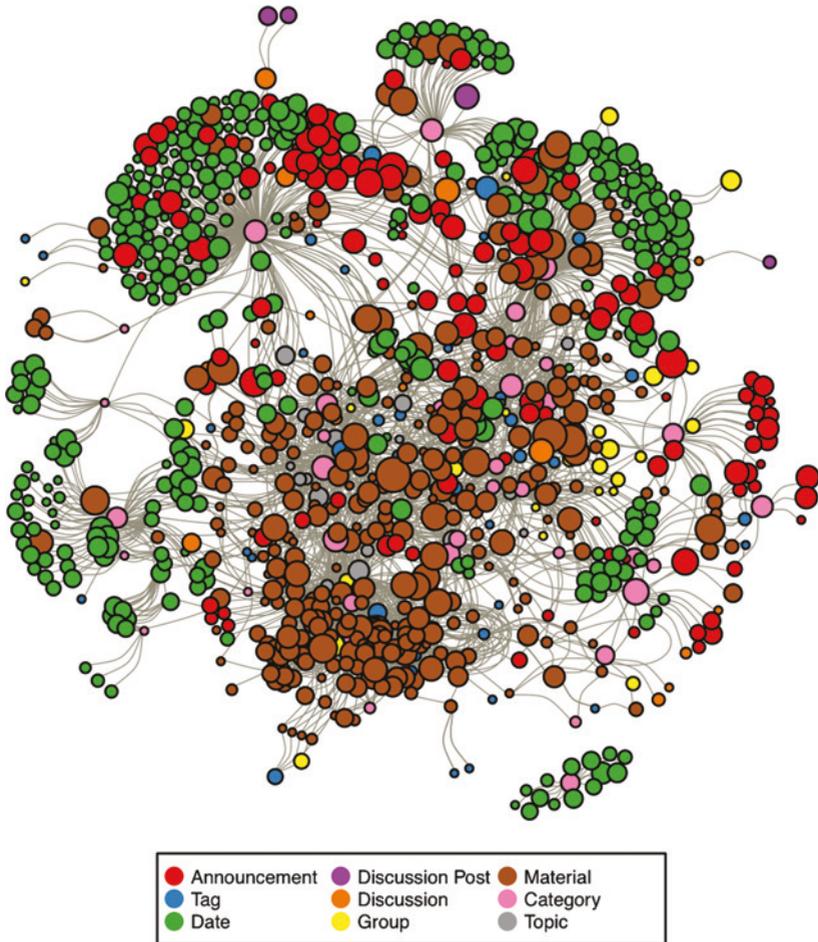


Fig. 16.2 Data from learning management system as network graph. *Source* Schulz and Breiter (2013)

teachers were recorded and coded according to standards of qualitative data analysis. Based on both data sources, we could find clusters of activities as well as subject-specific communicative practices. In both cases, the usage of the school information system was an integral part of the data collection. This helped us to identify patterns which could be reconstructed in the log-file analysis.

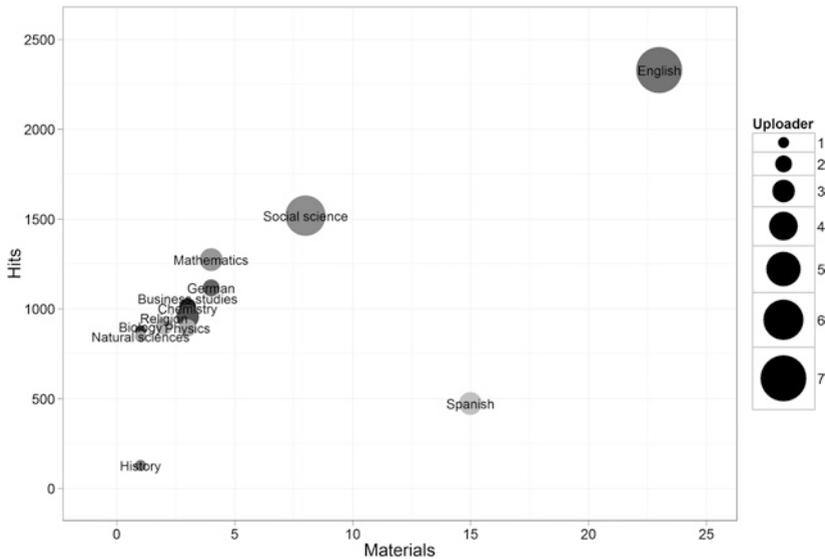


Fig. 16.3 Materials and hits by subject. Data from log-files of learning management system. *Source* Schulz and Breiter (2013)

This offered a different and more detailed view on the organizational processes of a school beyond the data in the system. Spanish is a small subject; the teachers usually teach at different schools and need to be virtually present at different locations. Time management is difficult and the learning management system with calendar function allows the scheduling of meetings and book resources online any time and anywhere. The link to their subject community is mainly organized through web-based systems. The English subject group has a long-standing tradition of exchanging classroom materials. Years before the introduction of the learning management system, they arranged their exchange via paper folders in their subject-related staff room.

This example of digital traces in school data systems highlights the relevance of digital traces in context, which can be very rich empirical data if analyzed interdependently. Dealing with log-files entails additional concerns about research ethics and privacy. Users cannot give their consent a priori.

16.4 CONCLUSION: CHALLENGES PUTTING DIGITAL TRACES IN CONTEXT

Taking the example of school data systems, we could demonstrate what it means to put digital traces in context: the data collected by the respective systems have to be linked with further, detailed information to make them socially meaningful. Only in this way do such data become a source of describing our present complex social world of datafication. For empirical analysis, this is related to three challenges which we consider as fundamental for any social science analysis of digital traces.

The first challenge is to find a way to grasp digital traces with reference to a defined social entity. Very often, digital traces are understood as a phenomenon of a single actor, an individual who left the traces through the use of digital media and services. While this is correct for a basic definition of digital traces as well as for many procedures of data generation (it is the single user of an online system who leaves the footprints that are collected by this system—often because the individual is of interest as a customer), our example demonstrates that we rather have to consider these individuals as social actors whose practices are located and embedded in the figurations of further institutional contexts and groups of people (in the case of our example the organization of the school and the different groups of teachers). Only by reflecting this does the data become meaningful. Therefore, we have to have the whole communicative figurations in mind in which the individual who is the ‘originator’ of the respective traces acts. The challenge here is to find a way to link the data being automatically generated with a social analysis of such a figuration. To achieve this, it seems to be appropriate to start the research with the frame of relevance of this figuration and locate the analysis of the digital traces. By so doing, there is a chance of finding helpful ways of contextualizing.

As a second challenge, we are confronted with the triangulation of quantitative ‘digital methods’ (in the case of our example the log-file analysis) with forms of qualitative analysis that offer the context information which is needed. As we have seen, combining automatized collected and processed data, on the one hand, with various forms of qualitative interviews or focus groups, on the other, is a promising method. But this again refers back to the first challenge: only if the actor constellation of the figuration under consideration is known does it become possible to conduct such interviews and focus groups. This offers rich data, which

because of their richness at the same time might become critical from a research ethics point of view.

Therefore, research ethics are the third challenge. Any approach which puts digital traces in context in such a way entails linking digital data that are left (partly without any detailed knowledge of this on the part of the persons concerned) with further information about certain persons. The knowledge being gathered in this way can be very far reaching—at many times much more far reaching than the knowledge a person has about him or herself. For research ethics, one consequence of this is the necessity to inform the investigated persons in detail about such possibilities of data collection and analysis (and to offer them, for example, the opportunity to have such unknown information communicated back to them). Another consequence is that as researchers we must be very careful of how we publish such results because the publication of data on digital traces in triangulation with further information about individuals might offer others the chance to isolate these persons. Anonymization becomes an important and complicated task.

With respect to these three challenges, it becomes obvious how far the meaningful analysis of digital traces is more than just a new field for media and communication research. As a new field, it is necessary not only to reflect in a new way about the relations of qualitative and quantitative data but also about our (digital) research ethics. This is essential if we want to conduct a form of media and communication research that addresses the complexity of the present social world, which is increasingly characterized by datafication. We hope that this chapter offers some stimulation for further steps in such a direction.

NOTES

1. The term ‘trace’ collects numerous meanings and appendices (to trace, track, traceable, traceability, tracing, etc.) and seems to connote an isolated object as well as an action or a process (Serres 2002: 1; Reigeluth 2014: 249). Because of this semantic richness of ‘trace’ in general, there is some ambiguity determining ‘digital traces’ in a proper way, which we want to clarify with the definition above.
2. Here the general problem of the idea of the social world as the sum of assemblages becomes replicated (for a critique of such an approach see Couldry and Hepp 2016: 57–78).

3. Funded by the Deutsche Forschungsgemeinschaft (BR-2273/10-1). Acknowledgements to Arne Hendrik Ruhe for data processing and producing the figures in this project, which we quote in the following.
4. A visit is defined as a sequence of hits from a unique user. A visit ends after 30 minutes of inactivity.
5. The network visualization represents the links within the SIS. Edges are (bi-directional) links, nodes are single entries (dates, appointments, etc.). The linking was done by hand by the teachers. The arrangement was done by a ‘force-driven’ algorithm (Fruchterman-Reingold), based on which closely related (linked) nodes were grouped.
6. Research on knowledge management often describes informal communication practices at the workplace with the ‘water cooler’. Usually located at a central place in a US office, staff gather spontaneously to chat (e.g. Earl 2001).

REFERENCES

- Anagnostopoulos, Dorothea, Stacy A. Rutledge, and Rebecca Jacobsen. 2013. *The infrastructure of accountability: Data use and the transformation of American education*. Cambridge, MA: Harvard Education Press.
- Andrejevic, Mark. 2014. The big data divide. *International Journal of Communication* 8: 17.
- Baum, Constanze, and Thomas Stäcker. 2015. Die Digital Humanities im deutschsprachigen Raum. Methoden – Theorien – Projekte. *Sonderband der Zeitschrift für digitale Geisteswissenschaften* 1. doi:[10.17175/sb001_023](https://doi.org/10.17175/sb001_023).
- Borgman, Christine L. 2015. *Big data, little data, no data: Scholarship in the networked world*. Cambridge, MA: MIT Press.
- Bowker, Geoffrey C. 2014. The theory/data thing. *International Journal of Communication* 8: 1795–1799.
- danah, boyd, and Kate Crawford. 2012. Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society* 15: 662–679.
- Breiter, Andreas. 2014. Schools as mediatized organizations from a cross-cultural perspective. In *Mediatized worlds*, ed. Andreas Hepp and Friedrich Krotz, 288–303. London: Palgrave.
- Couldry, Nick, and Andreas Hepp. 2016. *The mediated construction of reality*. Cambridge: Polity Press.
- Couldry, Nick, Aristeia Fotopoulou, and Luke Dickens. 2015. Real social analytics: A contribution towards a phenomenology of a digital world. *British Journal of Sociology* 67 (1): 118–137.
- Crampton, Jeremy W., Mark Graham, Ate Poorthuis, Taylor Shelton, Monica Stephens, Matthew W. Wilson, and Matthew Zook. 2013. Beyond the

- geotag: Situating 'big data' and leveraging the potential of the geoweb. *Cartography and Geographic Information Science* 40: 130–139.
- Damkjær, Maja S. 2015. Becoming a parent in a digitized age: Facebook as an agent of change?—Performative, dialogical, and preservative Facebook strategies in the transition to parenthood. Paper presented at Nordmedia 2015, Copenhagen, Denmark, 13 August 2015.
- De Smet, C., J. Bourgonjon, B. De Wever, T. Schellens, and M. Valcke. 2012. Researching instructional use and the technology acceptance of learning management systems by secondary school teachers. *Computers & Education* 58 (2): 688–696. <https://doi.org/10.1016/j.compedu.2011.09.013>.
- Earl, Michael J. 2001. Knowledge management strategies: Towards a taxonomy. *Journal of Management Information Systems* 18: 215–233.
- Elmer, Greg, Ganaele Langlois, and Joanna Redden. 2015. Introduction: Compromised data—from social media to big data. In *Compromised data: From social media to big data*, 1st ed, ed. Greg Elmer, Ganaele Langlois, and Joanna Redden, 1–14. London: Bloomsbury Academic.
- Ferguson, Rebecca. 2012. Learning analytics: Drivers, developments and challenges. *International Journal of Technology Enhanced Learning* 4: 304–317.
- Gardiner, Eileen, and Ronald G. Musto. 2015. *The digital humanities*. Cambridge: Cambridge University Press.
- Gitelman, Lisa, and Virginia Jackson. 2013. Introduction. In "Raw data" is an oxymoron, ed. Lisa Gitelman, 1–14. Cambridge, MA: MIT Press.
- Golding, Peter, and Slavko Splichal. 2013. New media, new research challenges. An Introduction. *Javnost—The Public* 22: 5–10.
- Hepp, Andreas. 2016. Kommunikations- und Medienwissenschaft in datengetriebenen Zeiten. *Publizistik* 61 (3): 225–246.
- Hutchinson, Jonathon. 2016. An introduction to digital media research methods: How to research and the implications of new media data. *Communication Research and Practice* 2: 1–6.
- Ifenthaler, Dirk. 2012. Learning management systems. In *Encyclopedia of the sciences of learning*, vol. 11, ed. Norbert M. Seel, 1689–1692. Boston, MA: Springer.
- Kitchin, Rob. 2014. *The data revolution: Big data, open data, data infrastructures and their consequences*. London: Sage.
- Kitchin, Rob, and Gavin McArdle. 2016. What makes big data, big data? Exploring the ontological characteristics of 26 datasets. *Big Data & Society* 3: 1–10.
- Latour, Bruno. 2007. Beware, your imagination leaves digital traces. *Times Higher Literary Supplement* 6. <http://www.bruno-latour.fr/sites/default/files/P-129-THES-GB.pdf>. Accessed 27 Apr 2017.
- Latour, Bruno, Pablo Jensen, Tommaso Venturini, Sébastien Grauwin, and Dominique Boullier. 2012. The whole is always smaller than its parts:

- A digital test of Gabriel Tarde's monads. *British Journal of Sociology* 63: 590–615.
- Liu, Bing. 2011. *Web data mining: Exploring hyperlinks, contents, and usage data*. Berlin: Springer.
- Lohmeier, Christine. 2014. The researcher and the never-ending field: Reconsidering big data and digital ethnography. In *Studies in qualitative methodology*, ed. Martin Hand and Sam Hillyard, 75–89. Bingley: Emerald Group Publishing.
- Markov, Zdravko, and Daniel T. Larose. 2007. *Data mining the web: Uncovering patterns in web content, structure and usage*. Hoboken, NJ: Wiley.
- Marr, Bernard. 2015. *Big data: Using smart big data, analytics and metrics to make better decisions and improve performance*. Chichester: Wiley.
- Mayer-Schönberger, Viktor, and Kenneth Cukier. 2013. *Big data: A revolution that will transform how we live, work and think*. New York: John Murray.
- Merzeau, Louise. 2009. Présence numérique: Les médiations de l'identité. *Les Enjeux de l'information et de la communication* 1: 79–91.
- Mosco, Vincent. 2014. *To the cloud: Big data in a turbulent world*. St Paul, MN: Paradigm Publishers.
- Neuendorf, Kimberly A. 2017. *The Content Analysis Guide*. Los Angeles, CA: Sage.
- O'Connor, Daniel, Rob Shields, Suzan Ilcan, and Edwina Taborsky. 2001. Data, information, and knowledge. In *Organizational semiotics. Evolving a science of information systems*, ed. Kecheng Liu, Rodney J. Clarke, Peter Bøgh Andersen, Ronald K. Stamper, and El-Sayed Abou-Zeid, 21–40. Boston: Kluwer.
- Oliner, Adam, Archana Ganapathi, and Wei Xu. 2012. Advances and challenges in log analysis. *Communications of the ACM* 55: 55–61.
- Papamitsiou, Zacharoula, and Anastasios A. Economides. 2014. Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence. *Educational Technology & Society* 17: 49–64.
- Perrotta, Carlo, and Ben Williamson. 2016. The social life of learning analytics: Cluster analysis and the 'performance' of algorithmic education. *Learning, Media and Technology*: 1–14. doi:[10.1080/17439884.2016.1182927](https://doi.org/10.1080/17439884.2016.1182927).
- Puschmann, Cornelius, and Jean Burgess. 2014. Metaphors of big data. *International Journal of Communication* 8: 20.
- Reigeluth, Tyler B. 2014. Why data is not enough: Digital traces as control of self and self-control. *Surveillance & Society* 12: 243–254.
- Rogers, Richard. 2013. *Digital methods*. Cambridge, MA: MIT Press.
- Ruppert, Evelyn, John Law, and Mike Savage. 2013. Reassembling social science methods: The challenge of digital devices. *Theory, Culture & Society* 30: 22–46.

- Savage, Mike. 2013. The 'social life of methods': A critical introduction. *Theory, Culture & Society* 30: 3–21.
- Savage, Mike, and Roger Burrows. 2007. The coming crisis of empirical sociology. *Sociology* 41: 885–899.
- Schulz, Arne H., and Andreas Breiter. 2013. Monitoring user patterns in school information systems using logfile analysis. In *Next generation of IT in educational management*, ed. Don Passey, Andreas Breiter, and Adrie J. Visscher, 94–103. Berlin: Springer.
- Serres, Alexandre. 2002. Quelle(s) problématique(s) de la trace? Texte d'une communication prononcée lors du séminaire du CERCOR (actuellement CERSIC), 13 December 2002.
- Suneetha, K.R., and R. Krishnamoorthi. 2009. Identifying user behavior by analyzing web server access log file. *International Journal of Computer Science and Network Security* 9: 327–332.
- Townsend, Anthony M. 2013. *Smart cities: Big data, civic hackers, and the quest for a new utopia*. New York: WW Norton.
- Turow, Joseph. 2011. *The daily you: How the new advertising industry is defining your identity and your worth*. New Haven, CT and London: Yale University Press.
- van Dijck, José. 2014. Datafication, dataism and dataveillance: Big data between scientific paradigm and ideology. *Surveillance & Society* 12: 197–208.
- Venturini, Tommaso. 2012. Building on faults: How to represent controversies with digital methods. *Public Understanding of Science* 21: 796–812.
- Venturini, Tommaso, and Bruno Latour. 2010. The social fabric: Digital traces and quali-quantitative methods. In *Proceedings of Future En Seine*, 15–30.
- Welling, Stefan, Andreas Breiter, and Arne H. Schulz. 2015. *Mediatisierte Organisationswelten in Schulen: Wie der Medienwandel die Kommunikation in den Schulen verändert*. Wiesbaden: VS.

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.

