

Perspectives on Data and Practices

Heidrun Allert

Christian-Albrechts-Universität zu Kiel
Kiel, Germany
allert@paedagogik.uni-kiel.de

Christoph Richter

Christian-Albrechts-Universität zu Kiel
Kiel, Germany
richter@paedagogik.uni-kiel.de

ABSTRACT

There is not something like *big data research*, but a variety of diverse *big data research practices* in different fields. They are based on different logics, rationalities, epistemological beliefs, types of data, and even different forms of objectivity as well as concepts of theory. Practices are specific regarding their temporalities and materialities. Furthermore, big data practices are not necessarily about doing research, but also about governance. To make these differences explicit is a step towards better collaboration in the many fields of the web sciences.

CCS CONCEPTS

• **General and reference** → *General literature*;

KEYWORDS

big data research, research practices, social science, objectivity, epistemology, epistemic beliefs, data

ACM Reference Format:

Heidrun Allert and Christoph Richter. 2018. Perspectives on Data and Practices. In *WebSci'18: 10th ACM Conference on Web Science, May 27-30, 2018, Amsterdam, Netherlands*. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/3201064.3201109>

1 INTRODUCTION

The notions of *web science* and *big data research* might suggest a set of unique and coherent approaches. In this paper we argue that in the web science community there is not a unique approach on data, objectivity and research rationality, but a plurality of scientific practices and epistemic beliefs. Researching and understanding them would allow for better interdisciplinary work. The proliferation of big data and computer-based procedures for the collection and analysis of large datasets not only gave rise to new methods and technologies in the humanities and social sciences but also introduced new players and epistemic beliefs. This already becomes evident in the rival conceptions of data as something that is either to be generated or collected. This paper will: 1. Shed light on different research practices and their enactments of theories (theories in data collection and analysis vs. programmed theories in data

generation). 2. Tell apart different perspectives on data and forms of objectivity.

2 THEORIES ENACTED

In this section we argue that data practices inevitably involve theories, but theories are enacted at various times and in different forms. The temporality of practices and the enactment of theory in big data practices differs from several social science research practices.

Some proponents of big data analytics have repeatedly proclaimed the end of theory in recent years. But how could this thesis catch ground in light of the extensive critique on any positivistic take on data in the social sciences and the humanities? Referring to the notion of "data-behaviorism" [18] it has for example been claimed that such approaches "would be mechanistic, reductionistic and authoritarian as they privilege the scientist in relation to the researched subject" ([20], p. 200, own translation). Big Data conveys the impression that the observer would have access to more extensive and more accurate information of a person and his/her behavior than s/he has access to her/himself (ibid.). The above quote however already insinuates, that processes of algorithmic data analysis should be understood as a form of scientific observation and hence would fit into the established debates in the theory of science. But we question this, as big data comprises not only research but also governance practices.

Qualitative research in the social sciences, in contrast to quantitative approaches, is not a coherent approach but refers to a huge variety of approaches [4]. As there are diverse approaches in the qualitative field, their use and development is accompanied by respective methodological discourses (ibid.), discussions on methods and techniques of data collection, analysis and interpretation. A major question is, whether the data gathered actually represents the phenomena observed or whether methods and research practices become performative regarding the phenomena, i.e. having an effect on the reality observed. [19] has a close look at qualitative research practices regarding their temporality. He states that research practices in social science are organized in a way that allows for separating data analysis from data collection. Transcribing interviews, arranging field notes, sharing material among inter-raters and so forth separates data collection from analysis and takes reasonable time. It is these practices and discourses, shared among researchers, which make it sound unreasonable that big data practices are claimed to be theory-free. From this perspective (performing research practices in the social sciences), big data practices seem to take a short-cut between data collection and data analysis (as in big data practices *actions* are taken immediately after *data generation*), and thus to cause a short circuit which is unjustifiable. Algorithms are designed to perform action (e.g. a search machine presenting search results). Considering temporality, big data practices allocate no time to actually bring in theory between data collection and

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

WebSci'18, May 27-30, 2018, Amsterdam, Netherlands

© 2018 Association for Computing Machinery.

ACM ISBN 978-1-4503-5563-6/18/05...\$15.00

<https://doi.org/10.1145/3201064.3201109>

analysis (or better: between data generation and performing action). Social media applications and search machines directly translate data generation into action (e.g. displaying search results). In contrast to this, the social scientists aim to perform methodologically controlled steps to go from *collecting data* to *analyzing data*. Furthermore, moving from *data analysis* to *taking action* requires for *creative design/synthesis* (this step is most often referred to as *intervention*, not as *research*). But does that mean that big data practices are theory-free? Here we state that the theory is implemented in the instrument. We call this: *programmed theories*. Be it a theory on what a good shopping experience looks like, a theory on the user, and so on. This theory is materialized (programmed) in an online shop, a social media platform, an application, a research instrument, a search machine, and so on. Big data practices are not theory-free, but theory comes in at a different stage (at the design phase). [21] demonstrate this for *big data in the physical sciences* where theory is in the Large Hadron Collider. According to [4], programmed theory is also known in social science: He states that theories are in instruments used in qualitative data analysis, i.e. research software. Thus, theory becomes performative in big data practices as well as in social science practices.

3 DIGITALE DATA AND DIFFERENT FORMS OF OBJECTIVITY

Closely related to the idea of analytic procedures that operate independent of theoretical commitments is the question of our understanding of the "nature" of digital data and their "objectivity". Again, we are facing a peculiar coexistence of two contradictory positions in the current discourse on digital data. While on the one hand digital data is conceived as a "natural resource", that can be mined, exploited and processed (e.g. [15]), scholars in the social sciences and the humanities on the other hand have stressed the inevitable subjectivity and social constitution of data and its interpretation (e.g. [3],[16]). But how can the persistence of these disparate conceptions of (digital) data be explained? How can it be that not only software engineers and vendors of respective technologies but also decision makers and managers assume the objectivity of data to be given while the discourse in the social sciences and humanities points into the exact opposite direction? A way to overcome this (apparent) contradiction is not to ask whether one of these positions is right and the other wrong but to have a closer look at the way data is actually used and the functions that data fulfill in different contexts. This change in perspective is in line with the observation that there is hardly any agreement on what data actually is, neither in science nor in applied disciplines (cf. [7]). If we shift the focus towards the practical usage of data we also have to reconsider the question of when and under which conditions data qualifies as "objective". Instead of asking whether "objectivity" can principally be achieved or ensured, the primary focus is on how different *forms objectivity* are produced and established within different contexts of practical action (cf. [7]). In the following we aim to roughly outline three basic perspectives on (digital) data and the related *processes of objectification* we spot in the current discussion on big data and the algorithmic processing of data. Tracing these three perspectives, we aim to trigger a more differentiated discussion on data in the web sciences and provide a framework for critical data studies.

3.1 Objectivity as a Representational Function – the Observers' Perspective

The first perspective, that can also be understood as the *observers' perspective*, is closely related to the methodological traditions and discourses grounded in empirical (social) research. From this perspective (digital) data is primarily conceived of as a means for the description of empirical matters. Data in this view is first and foremost something that depicts or represents something else. The use of digital data, from this perspective, is conceived of in reference to the usage of measurement or observational data. Data, hence, are not given, but have to be collected, elicited or extracted by means of dedicated procedures. For example, from an observer' perspective data are collected from Twitter in order to explore the linguistic conventions applied. The matters, things or phenomena to be represented and analysed thereby are conceived, at least ideally, as independent of the data collection procedures¹. Due to their role as depictions and recordings of an empirical matter, the collected data allows for a systematic, repeated and collective analysis, an analysis functionally decoupled from the matter itself. Towards this end it does not matter whether the actor adopt a quantitative approach, trying to ensure objectivity by means of experimental control, sampling and standardized data collection techniques or whether they opt for the collective and methodologically reflected interpretation of qualitative data. As data in this perspective is relevant only to the extent that it refers to something else, objectivity is a representational function. Objectivity, in both cases, is established by disciplinary legitimized forms of signification, backed up by the respective procedures and standards of data-collection, -analysis and - interpretation. As a consequence, objectification from the observer' perspective is primarily a concern for a common understanding of some empirical matter in a situation where the actors are relieved from any direct action on this very subject matter. The materiality of data and of the instruments used for data-elicitation and processing plays a marginal role at best (cf. [7]). To put it brief: data, in this perspective, stands in for something else.

3.2 Objectivity as a Product of Articulation – the Users' Perspective

In the case of social matters, however, data are often not just means of representation but also of transformation. Data cannot only be depictions or recordings of something but also for something. The transformation of the world by the production and usage of data is the central motive of the second perspective, that can also be read as the *users' perspective*. From the users' perspective data is an integral element of (social) interaction. Data are not just means for the description of certain matters but, an especially in digital contexts, constitutively entangled with the matters itself. The creation of a Tweet is coextensive with the production of data, data that can be processed, stored, reused and (re-)interpreted by computers in very different contexts. As this is a fact the users are principally aware of, it consequently (directly or indirectly) shapes the way the respective technologies are used. The data and the subject matter are therefore

¹Concerns for the reactivity of data collection procedures or the occurrence of measurement artifacts can be understood as practical problems to achieve the ideal of purely representational data.

reflexively entangled from the users' perspective. As [14] has argued, these kinds of circular effects shape the instrumental, social, and aesthetic qualities of social media usage. For example, Twitter's 140-characters limit has not simply been a technical restriction but was a co-constitutive moment for the emergence of specific linguistic conventions, memes, genres and collaborative practices unfolding on Twitter. From the users' perspective, the production of qualitative and quantitative data is embedded in processes of articulation and social positionings (cf. [9]). Data here are not just carriers of meaning but enter into practice as material entities that come along with specific forms of resistance. It is only by means of data that the respective forms of sociotechnical interaction become possible while at the same time they produce a surplus of meaning, which is neither intended or controlled by those evoking the data [10]. The medium leaves its marks: it co-constitutively shapes what we are doing. The production of data, from this perspective, is not so much focused on description and retrospection but geared towards the initiation and opening of options for future interactions. The objectification of data does not depend on its (retrospective) interpretation but is achieved in the process of articulation itself. The actors, in this perspective, are "reflexive fellow players" [8] that cannot escape their active involvement in the production of data but are constitutively entangled with the data they are producing. The observers' perspective and the corresponding forms of distant critique is no option for them. Critical engagement, from the users' perspective, instead requires an experimental take on the established practices geared towards the exploration of new forms of socio-technical intercourse.

3.3 Objectivity as Operationalization – the Processors' Perspective

The third perspective, that we deem relevant in the current discussion on big data and algorithmic data processing, is focused on the utilization and usage of data from a technical and information processing point of view. This *processors' perspective* does not approach data as representations of or for something but conceives data as objects that are amenable to algorithmic processing in the first place. Data is defined here as "reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or processing", which is organized in discrete "data objects" [1]. Data, from this perspective, is the primary *material* on which digital and computer-based technologies operate. The materiality and objectness of data is central to this position, as it are the data objects that are processed and operated on. The data of an email is the thing itself - the email only exists as a data object. As objects they are also objectively given, there is hardly any doubt whether a data object exists or not. From an informatics point of view, data can be understood as signs that are reduced to their material dimension, as signs that are stripped of their relationality, as Nake (quoted in [13]) put it. Accordingly, data and sets of data are a resource that exists independent of their interpretability. The read- and interpretability of data is not given but has to be explicitly produced and therefore can also get lost so that data might become "rotten" (see [2]). The read- and interpretability of data, this is another important specific of this perspective, has to be ensured both for the algorithmic processor as well as the human

user. The development of digital technologies hence takes place at the interface of two systems, the computational system on the one and a socio-material reference system on the other hand and shapes the relation between these two. The interpretation of data however is fundamentally different among humans and computers [12]. While human users inevitably make sense of digital data in relation to the situation and context they find themselves in as well as their motives and interest, computers can interpret and operate on data only on the basis of the formalized procedures available to them. In this sense computers are determined by the data they are provided with. In this sense, a piece of software such as Google Translate does not need to *understand* a natural language utterance in order to generate a translation but can do so by the algorithmic processing of a respective dataset. The core challenges from the processors' perspective is to devise "auto-operational forms" [5] that allow to abstract from the specifics of a particular situation in a way that algorithmic data processing becomes possible while ensuring a meaningful interpretation of the data in relation to some sociomaterial practice. Digital technologies, therefore are not just closely related to the development of respective algorithms but also to the establishment of formalized systems of description, models and ontologies. From the processor' perspective objectivity is essentially achieved to the extent data is made accessible to operational processing, both on the semantic as well as the ontological layer. Data have an instrumental function and are means in order to do something. To achieve this, respective technologies effectively *short-circuit* the processes of data elicitation (production), analysis and intervention.

3.4 Options for Practical Data Critique

In the contrasting juxtaposition of the three perspectives on digital data it should have become obvious that the understanding of data, the way objectivity is established but also the ways in which data can be criticized, is dependent on the way data is used in different social arenas. Depending on whether the observers', the users' or the processors' perspective is adopted different conceptions of data come to the fore. At the same time these different perspectives do not exist in isolation but are closely interwoven in practical encounter, even though they cannot be reduced to one another. As a consequence, empirical scientists cannot simply take a person's posting on a social media site as an authentic articulation of this person's opinion but have to account for the reflexivity of data-production on these platforms (cf. [11]). Similarly, the scientific critique of digital research methods also has to reflect on the material qualities of the technologies used in the research process (e.g. [6]). By the same token users of digital technologies are not just producers but also recipients of data and therefore also entangled in processes of observation, reflection and collective meaning making. In addition, they are also forced to deal with the resistance of digital technologies and to position themselves in relation to these technologies and the regimes of governance they are implicating. Finally, those engaged in the development of the technologies, to some extent have to draw on or rely on observational data in order to devise model, ontologies and auto-operational form, while at the same time they articulate themselves in the development process and position themselves in relation to other social actors,

be it vendors, users or some other public entity. To foster the multidisciplinary discourse on digital data and technologies it seems important to overcome a reductionist understanding of how data is *produced, made sense of* and *objectified*. Instead it appears important to have a closer look at the practices and the rationalities in and through which data are made up and used. This also requires theoretical perspectives and methodological approaches, that account for symbolic and material nature of digital data. Towards this end we concur with [6] and [16], that timely forms of critical data studies have to have a close look on the actual practices through which data is generated, interacted with and used. It also implies that we have to analyze the historical, cultural and societal processes that gave rise to but are also implicated by today's digital data practices.

4 FURTHER WORK

This paper shows that there are different logics and epistemic beliefs concerning research, data, objectivity and theory-based data analysis. A consequence is, that in the web science community we should not talk about big data research as a unique approach and epistemological stance but refer to big data research as various big data practices. Based on this, further work in the science and technology studies may be on investigating a variety of big data practices. Ethnographic methods and practice-based accounts are appropriate to take into account the doings and sayings, the materiality and temporality of specific practices. A respective approach on different rationalities of creative design is taken in [17]. We would like to present examples where the perspectives come together as well as examples that address each perspective only. Addressing a reviewer comment: How do these three perspectives, alone or together, shed light on different issues? How do these perspectives become useful for conceptualizing and articulating research problems in action?

REFERENCES

- [1] ISO/IEC 2382:2015. 2015. *Information technology – Vocabulary*. Technical Report. <https://www.iso.org/standard/63598.html>
- [2] Tim Boellstorff. 2014. Die Konstruktion von Big Data in der Theorie. In *Big Data – Analysen zum digitalen Wandel von Wissen, Macht und Ökonomie*, Ramon Reichert (Ed.). transcript, 105–131.
- [3] danah boyd and Kate Crawford. 2011. Six provocations for big data. In *A decade in internet time: Symposium on the dynamics of the internet and society*. Oxford Internet Institute, 0–17. <https://ssrn.com/abstract=1926431orhttp://dx.doi.org/10.2139/ssrn.1926431>
- [4] Rainer Diaz-Bone. 2011. Die Performativität der qualitativen Sozialforschung. *Forum Qualitative Sozialforschung/Forum Qualitative Social Research* 12, 3 (2011). <http://www.qualitative-research.net/index.php/fqs/article/view/1750/3254>
- [5] Christiane Floyd. 2002. Developing and Embedding Auto-Operational Form. In *Social Thinking - Software Practice*, Yvonne Dittrich, Christiane Floyd, and Rolf Klischewski (Eds.). MIT Press, Cambridge, 5–28.
- [6] Sebastian Gießmann and Markus Burkhardt. 2014. Was ist Datenkritik? Zur Einführung. *Mediale Kontrolle unter Beobachtung* 3, 1 (2014), 0–13. <http://www.medialekontrolle.de/wp-content/uploads/2014/09/Giessmann-Sebastian-Burkhardt-Marcus-2014-03-01.pdf>
- [7] Lisa Gitelman and Virginia Jackson. 2013. Introduction. In *„Raw Data“ Is an Oxymoron*, Lisa Gitelman (Ed.). MIT Press, Cambridge, 1–14.
- [8] Karl-Heinz Hörning. 2004. Soziale Praxis zwischen Beharrung und Neuschöpfung. Ein Erkenntnis- und Theorieproblem. In *Doing Culture – Neue Positionen zum Verhältnis von Kultur und sozialer Praxis*, Karl-Heinz Hörning and Julia Reuter (Eds.). transcript, Bielefeld, 19–39.
- [9] Benjamin Jörissen. 2014. Artikulationen - Bildung in und von medialen Architekturen. In *Schriften zur Medienpädagogik: School's out? - informelle und formelle Medienbildung*, Sandra Aßmann, Dorothee M. Meister, and Anja Pielsticker (Eds.). Kopaed, München, 13–27.
- [10] Sybille Krämer. 1998. Das Medium als Spur und als Apparat. In *Medien Computer Realität – Wirklichkeitsvorstellungen und Neue Medien*, Sybille Krämer (Ed.). Suhrkamp, Frankfurt a.M., 73–94.
- [11] Lev Manovich. 2011. Trending: The promises and the challenges of big social data. *Debates in the digital humanities* 2 (2011), 460–475. http://www.manovich.net/DOCS/Manovich_trending_paper.pdf
- [12] Frieder Nake. 2001. Das algorithmische Zeichen. *GI Jahrestagung* 2 (2001), 736–742.
- [13] Erhard Nullmeier. 2007. Wissensbasierte Systeme. In *Wissensmanagement in der Wissenschaft: Wissenschaftsforschung Jahrbuch 2004*, Klaus Fuchs-Fittkowski, Walther Umstätter, and Roland Wagner-Döbler (Eds.). GeWiF, Berlin, 43–61.
- [14] Johannes Paßmann. 2014. From Mind to Document and Back Again. In *Big Data – Analysen zum digitalen Wandel von Wissen, Macht und Ökonomie*, Ramon Reichert (Ed.). transcript, Bielefeld, 259–285.
- [15] Bob Picciano. 2014. Why Big Data is the New Natural Resource. *Forbes Brand Voice* (2014). <https://www.forbes.com/sites/ibm/2014/06/30/why-big-data-is-the-new-natural-resource/#55a8b3656628>
- [16] Florian Püschel. 2014. Big Data und die Rückkehr des Positivismus: Zum gesellschaftlichen Umgang mit Daten. *Mediale Kontrolle unter Beobachtung* 3, 1 (2014), 0–23. <http://www.medialekontrolle.de/wp-content/uploads/2014/09/Pueschel-Florian-2014-03-01.pdf>
- [17] Christoph Richter and Heidrun Allert. 2017. Different Rationalities of Creative Design – A Comparative Case Study. In *Research into Design for Communities*, Vol. 1. Springer Nature, Singapore, 3–13.
- [18] Antoinette Rouvroy. 2013. The end(s) of critique: data-behaviorism vs. due process. In *Privacy, Due Process and the Computational Turn: The Philosophy of Law Meets the Philosophy*, Katja de Vries and Mireille Hildebrand (Eds.). Routledge, Abingdon, 143–165.
- [19] Tim Seitz. 2017. *Design Thinking und der neue Geist des Kapitalismus: Soziologische Betrachtungen einer Innovationskultur*. transcript, Bielefeld.
- [20] Felix Stalder. 2016. *Kultur der Digitalität*. Suhrkamp, Berlin.
- [21] Anna Wilson, Terrie-Lynn Thompson, and Cate Watson. 2017. Big Data and Learning Analytics: Singular or Plural? *Firstmonday* 22, 4 (2017). <http://firstmonday.org/ojs/index.php/fm/article/view/6872/6089#author>