Towards understanding the duality of datafication and culture in current society

Daniel Houben & Bianca Prietl

1. Introduction: datafication and digitalization of society as recent phenomena

Datafication and digitalization are recent phenomena that – according to our perception – are of a strangely ambivalent appearance. On the one hand everyone is somehow aware of these developments; heated discussion primarily concerning issues of privacy and control over data not only arise within popular scientific discourses (see for instance Morgenroth 2014; Schreier 2015). On the other hand, however, scientific engagement – especially within sociology – falls somehow short. Digitalization and datafication are still mostly discussed and researched in the context of media studies and cultural studies, recently predominantly under the catch-term “Big Data” (see for instance boyd & Crawford 2012). While these studies gain fruitful insights in the specifics of mediatized communication and social interaction as well as knowledge production, consumption, and distribution, they also remaining restricted to a rather narrow scope of questions. Further, the term “Big Data” leads itself to a limited perspective on this far more complicated phenomenon. Big Data is usually defined by the so-called 3-Vs of Big Data – volume, velocity and variety – focusing our attention at the sheer amount of data, at the speed at which it increases and at which it can be processed as well as at the diversity of existing and collected data. Although these characteristics bear important insights, they do not account for the novel qualities which render society and culture deeply changed (see among others Lupton 2015; Orton-Johnson & Prior 2013; Reichert 2014). Ancient bureaucracies as well as the state apparatus of the former German Democratic Republic are known for the immense amount – and variety – of data collected about their people; a mere increase of ‘the same’ would not justify our increased attention, but what does, are insights and findings arising in the context of digitalization and datafication that are far less often discussed and brought together:

1. Changing social practices and institutions in the context of datafication
   No longer are data simply collected about us, but we constantly – willingly or unwillingly – produce data ourselves. The idea and practice of sharing data about almost everything with friends but also with an unknown public is for example a fairly recent development.

2. Data are ubiquitous
   The digitalization of ever more social processes, the development of diverse tracking technologies, and the practice of sharing data have allowed for the datafication of almost, if not all, spheres of society. In contrast to former times there are hardly any data-free areas – be it intimate relationships or the relation to our own bodily selves, data are actually everywhere, even in what is still called the ‘private’ sphere.
3. **Digital data are connected and traceable**

The binary form and algorithmic processing of digital data allows for it to have a „new dimension of connectivity“ (Baecker 2013, 156; also Miller 2011) that again allows for the accumulation of data traces that make our every doing traceable (Latour 2013, 120). The new possibilities of storing, aggregating, processing and so forth of digital data (Manovich 2012), thus, result in data profiles that are more dense and extensive than what we have ever seen before.

4. **Digital data are increasingly generated and processed automatically**

Or as Miller (2011) puts it: „[D]igital products and media can be automatically modified or even created through software and programs instead of being specifically created or modified by people. In short, much of what we experience as media objects in digital culture are created out of databases by machines as opposed to being the result of human endeavor.“ (19)

5. **Rising importance of data banks**

As some argue, data banks are on the brink of becoming the new dominant objectification of our culture, since increasing forms of social interaction and cultural reproduction at large depend on them (Burkhardt 2015). As Manovich (2001, 128) states, „[…] almost every practical act involves choosing from some menu, catalogue, or database“.

Departing from these insights, we argue for at least two desiderata that are inextricably intertwined and that need to be taken into consideration when sociologically examining the phenomena of digitalization and datafication: First, it is necessary to do some terminological groundwork in order to gain a sociological idea of the term data that considers the multiplicity and diversity of social developments associated with the digitalization and datafication of society. One way of defining data in sociology is to understand it in relation to the terms information and knowledge. Starting from this idea, we will depict some of the limitations of such a discussion and propose an analytical framework that draws on insights from Science and Technology Studies as well as on social constructivist perspectives. Second, while media studies and cultural studies have pointed out the implications of digital forms of communication – mostly with respect to social media – the developments sketched above ask for a more fundamental approach that aims at conceptualizing digital transformations and datafication as deeply ingrained into societal reproduction. As we will argue, data must be understood as both being socially constructed and therefore culturally embedded on the one hand and as ever more taking part in the reproduction of society on the other. Hence, we think it necessary to address what we call the duality of data and culture as a co-constitutive interplay, which renders our society deeply changed.
2. Terminological groundwork: data – information – knowledge

Empirical social research – for sure the very branch of social sciences which dedicates itself to the reflection on data of all kinds – tells us to understand data as (numerical) representations of ‘real life’ observations (Schnell/Hill/Esser 2012). With data – at least within the paradigm of quantitative social research – being thought of as representative of some matter of interest in the social world, the major goal then is to obtain data that corresponds to certain models of reality and causality and at the same time to minimize “biases” that is flaws in these representations (Diekmann 2003, 40-76). Following this perception, data is then to be viewed as processed representations and therefore necessarily reductions with respect to their ‘real life’ relatives.

While data is considered as a means to understand the social world in empirical social research, it serves primarily as a contrasting concept in thinking about information and knowledge within the sociology of knowledge. Most commonly within this strand of research information is understood as the meaningful interpretation of a signal (Ott 2007, 389) that is of some sort of data being transmitted between a transmitter or sender and a receiver or recipient. Hence, information as a term as well as a phenomenon is strictly relational, relative, situational, and dependent on the very context it is generated and interpreted in (Faßler 2001, 281). Holbein (1993), for example, conceptualizes information as already presented, sorted, and therefore processed data. In other words, information can be seen as the smallest entity of social order, for it already incorporates different processes, configurations, decisions, social constructions, legitimations, or interpretations (Faßler 1997, 332). Knowledge is further conceptualized as building on information and is thereby thought of as meaningful translation or transformation of information (with the term translation not necessarily understood in an ANT-sense). Information then is versatile and the building block of knowledge without being taken up in it completely. Hence, knowledge emerges when information – mediated or not – is translated into language or into action. Put differently, information is knowledge in waiting (Faßler 2001, 337). As different strands within the sociology of knowledge argue, knowledge is to be understood as highly interdependent with social structure (see in the tradition of phenomenology Berger & Luckmann 2004; in discourse theory Foucault 2012 [1976]).

Both of these discussions of data have considerable limitations, of which we want to consider two in more detail. Looking at empirical social research – especially its quantitative paradigm – it needs to be considered that while it is acknowledged for so-called “bias” in data, the very notion of data biases refers to a positivistic understanding of the social world and, thus, the ideal of a non-biased, objective and neutral representation of this world. Recent work within social accounting research, however, challenges this idea by radically questioning the objectivity and neutrality of any data and, instead, emphasizing the socially
constructed character of every data. Following this line of reasoning the rather programmatic claim according to which „raw data is an oxymoron“ (Gitelman 2013) can be taken as a starting point in revealing the complex operations of measurement and calculation that make up the highly abstract (digital) data we usually rely on (Hansen 2015; Heintz 2007; Porter 1996). Whereas the latter appear to adhere to the formal logics of mathematics, it is those very rules, categories and classifications that need to be analysed as the result of socially negotiated orders and legitimized forms of social constructions of reality. This also means to take into account the particular context within which certain data makes sense and is taken as legitimate representation of ‘reality’ (Barlösius 2005; Bowker/Star 1999). The cascading approach of differentiating data from information and knowledge, further, takes data to be at the basis of information respectively knowledge and therewith tends to stabilize it as allegedly a-social input of social processes that only then transform it into information or knowledge.

We want to shortly hint at two recent studies that convincingly demonstrate that data – especially in a digital and algorithmically processed form – is everything but an objective and neutral input of information or knowledge. Both studies are experimental explorations of Google’s search algorithm in combination with Google Ads, done in software departments of the U.S.: Datta, Tschantz and Datta (2015) show that Google Ad Settings can lead to discriminatory ads, in particular that setting the gender in the user profile to female results in getting fewer ads related to high paying jobs than setting it to male; Sweeney (2013) shows that a Google search for a person’s name is 25 percent more likely to get an ad suggesting an arrest of the person in search – such as “Ebony Bookman, arrested?” – when the person’s first name was one that is associated with an African American origin than when it is a name that is associated with white people. Concluding with Langdon Winner’s (1999 [1980]) claim that “artefacts have politics” it seems fair to assume that also digital data technologies can be racist and sexist and, thus, need to be analysed as highly social phenomena. In doing so, we think, that Science and Technology Studies (STS) have much to offer, since analysing digital data technologies – such as data banks, self-tracking apps and algorithms – as specific artefacts would allow for inquiries on how relations of social inequality become constitutive for the specific construction – and later use – of those artefacts. (Take for instance the gender script approach proposed by Rommes, Van Oost & Oudshoorn 1999 or the approach of co-materialization of ICT-artefacts and gender as proposed by Bath 2011). Taking up STS-literature on the social construction of scientific knowledge, especially in the tradition of laboratory studies (Knorr-Cetina 1984; Latour & Woolgar 1986 [1979]), guides our focus on the fundamentals of computer science such as the rules and processes of formalization, standardization and abstraction that form the basis of digital data technologies. This would mean to scrutinize the operations of classification and
categorization that transform complex and heterogeneous data into aggregated, comparable and processed data (Crutzen 2007; Star 2002).

3. Conclusion: duality of datafication and culture as co-construction of digital data (technology) and society

To summarize, it can be noted that mainstream positivist conceptualizations of data are themselves somehow distorted and tend to confirm everyday notions of (numerical) data as being objective and neutral (Heintz 2007). Symptomatically, then, we can find much more research on how datafication and digitalization transform social practices and institutions than research on how data – and therewith also the developments of datafication and digitalization – are themselves socially constructed and the effect of specific social conditions. Thus, the current research on “Big Data” and society not only reproduces the idea of data as input for social and cultural transformations. Thereby it loses track of the multiple social influences and dynamics that society puts into their very coming to existence. We therefore argue for an analytical framework that considers data itself as socially constituted as well as for research that analyses the ways in which data and digital data technologies are socially constituted.

We propose – as a working definition – a relational framework to conceptualize data, information, and knowledge (see figure 1). Hence, data as well as information and knowledge must be seen as boundary phenomena which cannot be apprehended satisfactory without considering their relatedness to each other. Even more important, this perspective emphasizes the imperative to focus on the processes of transformation and translation that constitute the constellations of data, information, and knowledge. From this also follows that data or datafication need to be considered in relation to what might be called culture or society.
More concrete, we propose to think of datafication and culture as presupposing each other. Such a dual view of datafication, digital technologies, and culture keeps in mind their high mutability and regards them rather as processes than as steady states. It makes little sense, then, to talk about datafication and digitalization as such, since they are only realized in actual use. Hence, data mediated communication cannot be grasped outside of everyday appropriation. Thus, any sustained effort to grasp the interplay between culture and digital development has to conceptualize them as "dual", i.e. as co-constituted. A proper understanding of our current society – and by that, the potential to emancipation – cannot be obtained without considering the interdependence of cultural forms and their digital bases, or vice versa, the cultural bases of digital forms. Therefore, the co-construction of datafication and culture should be at the heart of further development not only in Science and Technology Studies, but in any appraisal about the scientific use of and trust in "Big Data".

Applying a social constructivist perspective, datafication, then, describes how processes of the construction of reality by digitalized communication become manifest in technologies and media and how, in turn, existing specifics of digital technologies have a contextualized influence on the process of the communicative construction of socio-cultural reality. Any form of data, any technical artefact and any mediatized form of data processing, then, must be understood as only part of a not only consecutive but mutually independent part of what we propose to call cultural construction regime of data.

References


Morgenroth, Markus (2014), Sie kennen dich! Sie haben dich! Sie steuern dich!: Die wahre Macht der Datensammler. Droemer HC.


---

I This paper is based on a presentation given at the 15th Annual STS Conference in May 2016 in Graz, Austria.

II Contact: Daniel Houben & Bianca Prietl, RWTH Aachen University, Department of Sociology; address: Eilfschornsteinstraße 7, 52062 Aachen. E-mail: dhouben@soziologie.rwth-aachen.de; bprietl@soziologie.rwth-aachen.de