

Jessica Heesen:

Big Data for a Fairer Democracy?

Abstract:

Big data-analysis is linked to the expectation to provide a general image of socially relevant topics and processes. Similar to this, the idea of the public sphere involves being representative of all citizens and of important topics and problems. This contribution, on one side, aims to explain how a normative concept of the public sphere could be infiltrated by big data. On the other, it will discuss how participative processes and common wealth can profit from a thorough use of big data analysis. As important parts of the argument, two concepts will be introduced: the numerical public (as a public that is constituted by machine-communication) and total politicisation (as a loss of negative freedom of expression).

Agenda:

Introduction.....	16
Information technologies as tools and media	16
Technocracy instead of democracy.....	17
Numerical public and total politicisation.....	19
How to use big data in a smart way	20

Author:

PD Dr. Jessica Heesen:

- University of Tübingen, Centre for Ethics in the Sciences and the Humanities, Head of Junior Research Group Media Ethics, Wilhelmstraße 19, 72074 Tübingen, Germany
- ☎ + 49 – 7071 2977516, ✉ jessica.heesen@uni-tuebingen.de, 🌐 <http://www.izew.uni-tuebingen.de/>
- Relevant publications:
 - Computer and Information Ethics. In: Ruth Chadwick (ed.): Encyclopedia of Applied Ethics. Second Edition, Vol. 1, Academic Press: San Diego 2012, 538 – 546 p.
 - Ein Fels in der Brandung? Positionen der Medienethik zwischen verflüssigtem Medienbegriff und schwankender Wertebasis, in: Marlis Prinzing/Matthias Rath/Christian Schicha/Ingrid Stapf (Hg.): Neuvermessung der Medienethik. Bilanz, Themen und Herausforderungen seit 2000, Beltz Juventa: Weinheim 2015, 86 – 98 p.

Introduction

The concept of “big data” indicates the interplay of different technological developments. On the one hand, big data stands for efficient storage and management of huge amounts of data. On the other hand, it depends on the always growing sources for digital information due to mobile applications, an “always on”-mentality, and the pervasion of machines, communication, and the whole environment with information technology (internet of things, ambient intelligence). But big data does not only mean that more data is gathered from more sources – additional important aspects are new methods for the evaluation of data pools. The outcome will be demonstrations of new correlations or - as expressed by the technical term “knowledge discovery”⁴² - a new “knowledge” that is generated from data after a process of interpretation and modelling.

The ability to analyze large amounts of data is being discussed more and more often in the context of its usefulness for the improvement of political decision making and social welfare; for example for the prediction of epidemics⁴³, for smart policing (<http://www.smartpolicinginitiative.com>), traffic management, energy supply or the identification and publication of administrative affairs (e.g. <http://www.dublindashboard.ie>, <http://www.d4d.orange.com/en/Accueil>).

The MIT researcher Alex Pentland mentions in particular the richness of detail of big data analysis and the breadth of information it provides for all sectors of society. He has great expectations concerning the future benefits of big data: “We’re going to end up reinventing what it means to have a human society.”⁴⁴ That implies: potentially everything could be important and should be gathered and analyzed with the new comprehensive methods. At this point, an interesting linkage to the concept of the public sphere is emerging: big data-analysis is linked with the expectation of providing a general image of socially relevant topics and processes. Similar to this, the idea of the public sphere involves being representative of all citizens and of important topics and problems. This article wants to point out how a normative concept of the public sphere is challenged when a technology is used to support or substitute parts of its functions.

Information technologies as tools and media

Big data analysis uses a variety of data which provides information on everyday activities, the usage of internet-services and the communication behaviour of individuals. Because of their mobility and their intertwining with objects of every-day live, information technologies (IT) are ubiquitous. Yet at the same time, they keep their function as a “window to the world” – but in two directions: on the one hand, for example television, telephone and internet communication in general serve as media for information and communication; on the other, IT potentially reports the individual behaviour to an unknown circle of machines and/or human beings or, in the case of social networks, to a broader public sphere. Big data methods profit from the features of computers as universal machines with numerous users and contextual usages.

In this context, information technologies can be described as hybrids between tools and media. Therefore, sources for the gathering of data are, both, the witting respectively unwitting interaction with a technical system and the interaction between human beings by use of natural language. Only this latter “natural” communication is part of a discursive conception of the public sphere as it is justified in normative concepts of the public in philosophy and political theory. This normative concept includes the supposition that public discourse – to a certain extent – is guided by rational justification, refers to reciprocal expectations, and is focused on certain

⁴² Cf. Fayyad et al. (1996).

⁴³ Cf. Ginsberg u.a. (2009).

⁴⁴ Pentland (2012).

topics.⁴⁵ In contrast, big data starts from other premises: Its “knowledge” is constituted from data material that is generated through the everyday routines of people.

Nevertheless, there could be good reasons for intensified usage of big data for decision making procedures. In the following section four reasons will be mentioned why big data analysis could have advantages over politics as usual.

(1) It might be the case that knowledge from data analysis can claim a higher proximity to reality. In many empirical investigations one finds that people often give a subjective or strategic evaluation of their behaviour. For example, when asked about their highest priority interest with respect to internet communication, many people answer “information”, whereas statistics tell us that in reality, pornography is significantly more popular.⁴⁶ In this way analysis of data can produce a more accurate, “truer” empirical knowledge.

(2) Moreover, one could imagine that a comprehensive gathering of data takes into account all those people that are not part of public discourse and the publicity of (mass)media and its own laws (for example, so called marginal groups). Therefore a more intensive use of big data for politics could lead to more social equality.

(3) Also a general lack of interest of the population in the democratic process could be balanced by behavioural analysis. In many western democratic systems we have come to notice a decrease in voter turnout. Thus data analysis could offer a compensation for election and active participation.

(4) And as a last point it should be mentioned that big data includes a promise for more neutrality and rationality in political decision making, as pointed out for example by many advocates of so called information governance technology.⁴⁷

Technocracy instead of democracy

A consistent application of big data analysis might therefore have advantages in terms of policy-making. Possibilities for datafying or respectively mediatization offer a lot of potential advantages regarding government and regulation. For example there is talk of “evidence-based policy”⁴⁸ and the aforementioned “information governance”. These are phenomena of a new technocratic movement where governments and organisations try to strictly follow findings of science and research. Daniel Gillick, a Google Senior Research Scientist, states “Big data’ represents a cultural shift in which more and more decisions are made by algorithms with transparent logic, operating on documented immutable evidence.”⁴⁹

In this respect, one could speak of a newly optimized technocratic process. Initially, the term technocracy indicated a concept for the regulation of societal organization. According to that, “mechanical” elements of scientific and technological civilization assume the role of political arrangements. From this perspective, the advantage lies in an efficient social regulation that is immunized against inappropriate ideological decisions. In a technocracy, instead of the political sovereign, the optimized self-organization of the people through science and technology gains centrality.

Due to the established values of constitutional democracies, technocratic forms of government are refused; in particular because they undermine the role of the political sovereign and the principle of democratic self-organization.⁵⁰ Yet, constitutional democracies use significant technocratic elements of administration as well.

⁴⁵ Cf. Habermas (1996).

⁴⁶ [Daily Infographic \(2013\)](#).

⁴⁷ Information Governance Initiative, <http://iginitiative.com/>

⁴⁸ Head (2010).

⁴⁹ Dutcher (2014).

⁵⁰ Cf. Habermas (2013).

This becomes obvious for example in the definition of the goals for climate control: They are based on scientific forecasts, but they have heavy consequences for economic and political processes. Another example is the development of new forms of participation like e-democracy or open data, which are driven by new technological possibilities.

There is no doubt that it makes sense to use empirical and general scientific knowledge for political processes and decisions. How else could one come to an informed decision? Despite this, the advanced forms of data analysis develop their own dynamics that connect to the past lines of discussion about technocracy.⁵¹ The special authority of the "pure number", the apodictic evidence of calculation models and the apparent neutrality of quantification and statistics were always reason for disputes about their validity in terms of political action (but also within the epistemological dimension).⁵² Moreover, quantification is one of the methods to give a topic a high news value and, according to this, to gain a high level of public attention.⁵³ But what distinguishes the use of data analysis for social regulation on one side from a societal self-regulation by a political public sphere on the other?

Indicators for a normative concept of public communication are:

1. **Conscious participation:** this is connected to a specific cognitive mind-set of the participating persons in relation to their communication as public communication. On the one hand this attitude is reflected in the selection of the topic of communication. Accordingly, it is visible in a supra-individual choice of subject, or rather its relevance for a discussion of the *res publica* and public welfare. On the other hand this attitude is also reflected in the awareness of the individual behaviour: that means persons in a public situation will reflect on their impact on others. For example they will ask themselves if their behaviour is commendable and if they wish that anybody would act like they do.
2. **Reciprocity:** associated with this are also mutual expectations of communicators as they are formulated by discourse ethics. These expectations involve truth, rightness, and truthfulness of communication, which is oriented towards reaching understanding.⁵⁴
3. **Orientation:** media publicity produces a "knowledge from the knowledge of the other". This orienting knowledge is given through the ability to monitor what is happening in mass media communication. This knowledge includes not only knowledge of facts or historical contexts, but also knowledge on the validity of normative rules. Not least it gives information about the emergence of social understanding as such. Thus, the orienting function of public communication is a prerequisite for the realization of an individual's successful social participation.

These forms of orientation, reciprocity, and conscious participation require a special normative definition of the public. This normative concept is concerned with the function of the public sphere in a democratic system and the preconditions for making the public a valuable element of a liberal constitutional state. In spite of the importance of a normative concept of the public for political theory, empirically existing public spheres are often shaped by the prevalence of power struggles and the dynamics of media communication. In this sense a normative definition of the public sphere is oriented to an often counterfactual ideal.

⁵¹ In this regard, cf. the controversy between positivism and dialectic approaches (Adorno et al. 1989) as well as the debate about technocracy (Bell 1980, Ellul 1953, Lenk 1973).

⁵² Cf. Porter (1996).

⁵³ Cf. Luhmann (1990: 178).

⁵⁴ Cf. Habermas (1991).

Numerical public and total politicisation

But in the case of big data, things are different: In contrast to the development of a counterfactual ideal, big data is based on other conditions. The use of big data for policymaking builds on the numerous traces of individual action that are to be detected following the penetration of everyday life with information technologies. Unlike in public discourse – where citizens can regulate their participation in the public sphere –, big data-analysis causes the gathering and publication, especially, of every day-activities and behaviour. Because of this, big data gains a numerical picture of the general public *and* the general private that is connected by correlations, not by interactions. In this sense, one can speak of a numerical public here, thus, a public that does not reflect on its own publicness and public responsibility.

On the one hand the individual involvement in the constitution of a big data pool can be called participative in a way, but on the other hand this kind of “participation” opens a technical way for the pervasion of private and everyday life with a political or rather ideological component. Total participation, enabled by context-aware media, ambient intelligence and big data, implies, at the same time, a total politicisation. In this case, politicisation must not be understood as reciprocal relationship with discourse and critique in terms of the definition of public welfare, but as an anonymous communitization⁵⁵ in relation to different social contexts like economy, traffic management or insurance business. This means that the actions of every single person are arranged and in the end assessed with respect to specific patterns. The single action is no longer an individual one but part of a “typical situation”, the fulfilling of a specific stereotype in a similar group of people. On the one hand, this anonymous communitization is always combined with a comparison of individual behaviours with the current notion of normality and is an important factor for a feeling of surveillance and its negative consequences.⁵⁶ On the other hand, individual (private) actions gain importance for supra individual decision making.

To give a simple example: Maybe you have a weakness for superficial entertainment shows and from time to time you watch one of these shows with your smart television. Your internet provider, the broadcasting company and/or the smart TV-company will collect your data or rather your selection and maybe in the future more superficial entertainment will be offered for you and for the general public. But if somebody would have asked you individually for your opinion about the offers of TV-channels probably you would not have supported an increasing amount of entertainment shows. On the contrary you could have voted for more documentary films, news programmes or auteur movies because you think public communication should support intellectual involvement – even if you personally love superficial entertainment.

So, it can be noted that not the motivation for action is important for a data analysis, but the fact that you conduct it. In such a kind of data analysis it is not about finding reasons but about recognising patterns in data agglomerations. In this respect, a systematic use of big data for political action implies the exclusion of counterfactual and probably moral reasons from political reflection. Big data confines societal analysis to what is empirically existent. Instead of this, in public discourse citizens, journalists and stakeholders (and unfortunately also spin doctors) have the possibility for a deliberate (innovative or critical) discussion of topics. In this respect, data analysis only reproduces what exists anyway and reduces the possibilities of information technology to its affirmative – the societal conditions confirming – forms of usage.

Moreover, total politicisation implies that every individual action is registered, and, as a consequence, every individual person could start thinking about the political and ethical dimension of his or her action. In a society with a comprehensive use of big data, all areas of life gain importance for the public and the political and we get a “participation” by the analysis of behaviour. As another example: data from your private fitness program can be used by medical insurances to change their concept of solidarity.

⁵⁵ Cf. in a different context Hubig (2003).

⁵⁶ Cf. Haggerty/Ericson (2006).

Furthermore, for a total politicisation it is typical to have a dissolution of nonpolitical private and social spaces. Which in turn leads to the loss of a negative freedom of expression. But: in a liberal state you have the right not to express your opinion or attitude. So it can be concluded that (roughly) total behavioural analysis is quite the opposite of a seemingly non-ideological and neutral approach to reality – the aspect that is stressed as an advantage by the advocates of information governance. With total behavioural analysis, there are no longer any non-political and in this sense non-ideological actions.

How to use big data in a smart way

The foregoing analysis tried to describe the theoretical differences between a social organisation through big data on the one hand and public discourse on the other. At any rate, on an empirical level we cannot see a prevailing comprehensive use of data analysis yet. Nevertheless applications from information technology already determine areas of politics like security and energy politics. Hence public deliberation about the chances and risks of data analysis is essential to reflect on these new opportunities and to justify them as well informed decisions.

Big Data entails a lot of critical points to consider but on the other side there are a lot of valuable data for societal self-reflection and regulation that we should use. As a conclusion, the following section lists four assumptions that are important for a prudent “self-tracking of democracy”:

1. Awareness of the advantages and limitations of data analysis:
Don't have too much respect for “pure” data
2. Use of privacy by design to avoid a totalitarian accumulation of data:
Don't forget the principles of data protection like informed consent, limited purpose, relevance
3. Provision of governmental data as open data:
“Make public data available, protect private data” (CCC)
4. Integration of big data in public discourse:
With the aim of ensuring well informed decisions and showing the possibilities and capabilities of data gathering for politics, civil society and public welfare, data journalism⁵⁷ starts to be established.

References:

- Adorno, Theodor W. et al (1989): Der Positivismusstreit in der deutschen Soziologie 13th Edition. Darmstadt/Neuwied: Luchterhand.*
- Bell, Daniel (1980): Veblen and the Technocrats: On the Engineers and the Price System. In: The Winding Passage: Sociological Essays and Journeys 1960 - 1980, Cambridge: Harvard University Press, 69-90.*
- Chaos Computer Club (CCC): Hackerethik. <https://www.ccc.de/de/hackerethik> (13.1.2016)*
- Daily Infographic (2013): The Stats on Internet Pornography. <http://www.dailyinfographic.com/the-stats-on-internet-pornography-infographic>.*
- Dutcher, Jennifer (2014): What Is Big Data? (3.9.2014). <https://datascience.berkeley.edu/what-is-big-data/> (19.11.2015)*
- Ellul, Jacques (1954): La technique ou l'enjeu du siècle. Paris: Armand Colin.*

⁵⁷ Gray et al. (2012).

- Fayyad, Usama M. et al. (ed.) (1996): *Advances in Knowledge Discovery and Data Mining*. Amer. Assn. for Artificial Intelligence, Menlo Park, CA, USA, *AI-Magazin*, 37–54.
- Ginsberg, Jeremy et al. (2009), *Detecting influenza epidemics using search engine query data*. In: *Nature* 457, 1012 - 1014, 19 February 2009, doi: 10.1038/nature07634; (Published online 19 November 2008; Corrected 19 February 2009).
- Gray, Jonathan et al. (2012): *The data journalism handbook*. O'Reilly Media, Inc.
- Habermas, Jürgen (1991): *Erläuterungen zur Diskursethik*. Frankfurt a. M: Suhrkamp.
- Habermas, Jürgen (1996): *Strukturwandel der Öffentlichkeit. Untersuchungen zu einer Kategorie der bürgerlichen Gesellschaft. Mit einem Vorwort zur Neuauflage 1990 5th Edition 1996*. Frankfurt a. M: Suhrkamp..
- Habermas, Jürgen (2013): *Im Sog der Technokratie*. Berlin: Suhrkamp.
- Haggerty, K./Ericson, R. (2006): *The New Politics of Surveillance and Visibility*. Toronto: University of Toronto Press.
- Head, Brian (2010): *Evidence-based policy: principles and requirements*. In: *Strengthening evidence based policy in the australian federation, volume 1: proceedings01/2010*. https://www.researchgate.net/publication/242709855_2_Evidence-based_policy_principles_and_requirements (10.12.2015)
- Hubig, Christoph (2003): *Selbstständige Nutzer oder verselbstständigte Medien – Die neue Qualität der Vernetzung*. In: F. Mattern (ed.): *Total vernetzt, Szenarien einer informatisierten Welt*, Springer: Berlin/New York, 211-230.
- Information Governance Initiative.he. <http://iginitiative.com/> (13.1.2016)
- Lenk, Hans (ed.) (1973): *Technokratie als Ideologie. Sozialphilosophische Beiträge zu einem politischen Dilemma*. Kohlhammer: Stuttgart.
- Luhmann, Niklas (1990): *Gesellschaftliche Komplexität und öffentliche Meinung*. In: *Soziologische Aufklärung 5. Konstruktivistische Perspektiven*. Westdeutscher Verlag: Opladen, 170-183.
- Pentland, Alex (2012): *Reinventing Society in the wake of big data, A Conversation with Alex (Sandy) Pentland* [8.30.12]. <http://edge.org/conversation/reinventing-society-in-the-wake-of-big-data> (30.9.2015)
- Open Data City, <https://opendatacity.de/> (28.10.2015)
- Porter, Theodore M. (1996): *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton University Press: Princeton.