

University of Groningen

Datafication and empowerment

Baack, Stefan

Published in:
Big Data & Society

DOI:
[10.1177/2053951715594634](https://doi.org/10.1177/2053951715594634)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2015

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Baack, S. (2015). Datafication and empowerment: How the open data movement re-articulates notions of democracy, participation, and journalism. *Big Data & Society*, 2(2), 1-11.
<https://doi.org/10.1177/2053951715594634>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Datafication and empowerment: How the open data movement re-articulates notions of democracy, participation, and journalism

Big Data & Society
 July–December 2015: 1–11
 © The Author(s) 2015
 DOI: 10.1177/2053951715594634
 bds.sagepub.com



Stefan Baack

Abstract

This article shows how activists in the open data movement re-articulate notions of democracy, participation, and journalism by applying practices and values from open source culture to the creation and use of data. Focusing on the Open Knowledge Foundation Germany and drawing from a combination of interviews and content analysis, it argues that this process leads activists to develop new rationalities around datafication that can support the agency of datafied publics. Three modulations of open source are identified: First, by regarding data as a prerequisite for generating knowledge, activists transform the sharing of source code to include the sharing of raw data. Sharing raw data should break the interpretative monopoly of governments and would allow people to make their own interpretation of data about public issues. Second, activists connect this idea to an open and flexible form of representative democracy by applying the open source model of participation to political participation. Third, activists acknowledge that intermediaries are necessary to make raw data accessible to the public. This leads them to an interest in transforming journalism to become an intermediary in this sense. At the same time, they try to act as intermediaries themselves and develop civic technologies to put their ideas into practice. The article concludes with suggesting that the practices and ideas of open data activists are relevant because they illustrate the connection between datafication and open source culture and help to understand how datafication might support the agency of publics and actors outside big government and big business.

Keywords

Open data, agency, civic technologies, open source, hacking culture, journalism

Introduction

Agency is deeply connected to the distribution of knowledge and power. If we understand agency as ‘the longer processes of action based on reflection, giving an account of what one has done, even more basically, making sense of the world *so as* to act within it’ (Couldry, 2014: 891), then the conditions under which we can make sense of our world and our own actions are crucial for our capacity to act with agency. With the expansion of the Internet, social media, and Big Data technologies, we can currently observe a number of fundamental transformations of knowledge production and distribution that raise urgent questions about public agency. To date, however, questions about agency have been ‘obscured by unnecessarily

generalised readings’ (Couldry and Powell, 2014: 1) of the supposed power of the new technologies. For this reason, Couldry and Powell (2014) recently called for ‘social analytics’ as a new research paradigm in relation to Big Data.¹ They stress that agency is still relevant and that we should study new forms of reflexive agency in increasingly datafied societies. This involves paying more attention to the social and cultural dimension of

Centre for Media and Journalism Studies, University of Groningen, Groningen, The Netherlands

Corresponding author:

Stefan Baack, Centre for Media and Journalism Studies, University of Groningen, Oude Kijk in ‘t Jatstraat 26, 9712 EK Groningen, The Netherlands.

Email: s.baack@rug.nl



Creative Commons CC-BY: This article is distributed under the terms of the Creative Commons Attribution 3.0 License (<http://www.creativecommons.org/licenses/by/3.0/>) which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<http://www.uk.sagepub.com/aboutus/open-access.htm>).

this transformation by examining how social actors respond to processes of data collection and analysis and how they use data to ‘meet their own ends’ (Couldry, 2014: 892). This article follows Couldry and Powell’s call and shows how activism around open data is a rich, but so far largely overlooked side of inquiry that allows us to think about the relationship between data and agency in new ways.

The open data movement is a particularly interesting case because it intersects with two ongoing transformations of knowledge and power that seem to contradict each other in terms of agency: datafication and the proliferation of hacking or open source culture. On the one hand, the practices and imaginaries of open data activists are centered around the distribution and use of data and thus linked to datafication, the ubiquitous quantification of social life (Mayer-Schönberger and Cukier, 2013: 78), for which Big Data is the most prominent expression. Big Data ‘reframes key questions about the constitution of knowledge’ (Boyd and Crawford, 2012: 665) and raises concerns about the agency of publics. As Couldry and Powell (2014: 4) note, Big Data technologies and the growing relevance of algorithms may disconnect ‘system and experience’ because the traces of data people leave behind are often unconscious and not meaningful to them, and the insights generated by companies or governments are not, or only partially, ‘folded back into the experience of everyday life’. The comprehensive surveillance of online activities made possible by Big Data technologies thus might impede our potential to act in an agentic manner. On the other hand, open data activists apply practices and values from open source culture to the creation and use of data. This links them to other initiatives rooted in open source culture, like Open Access, Wikipedia, Wikileaks, Anonymous or Creative Commons (Beyer, 2014a, 2014b; Coleman, 2014; Sauter, 2014). Similar to datafication, these phenomena raise fundamental questions about ‘the nature of knowledge and expertise, how information is organized and evaluated, and who decides’ (Lievrouw, 2011: 26). Different to datafication, however, open source culture is associated with a transparent and collaborative form of governance that might support agency. As Raymond (2001) famously pointed out when he contrasted the ‘bazaar model’ of open source with the ‘cathedral approach’, open source culture is fundamentally concerned with the rights to access and distribute knowledge. Open source is based on voluntary participation (Weber, 2004: 62) and collaboration, granting access to the source code of software and incorporating contributions from potentially everyone. The implications of transferring the ‘open source process’ (Weber, 2004: 16) and the values inherent in this process to new domains with different ways of organizing knowledge ‘reach directly into the heart

of the legitimacy, certainty, reliability and especially the finality and temporality of the knowledge and infrastructures we collectively create’ (Kelty, 2008: 6–7).

An analysis of the open data movement offers a unique opportunity to *connect* datafication and open source culture, which raises interesting questions about agency: how do activists apply practices and values from open source culture to data, and what does this tell us, in return, about agency in datafied publics? To address these questions, I will present key findings from a study on the *Open Knowledge Foundation Germany* (OKF DE), a not-for-profit organization and one of the most influential and visible actors in the German open data movement.² First, I will address how we can trace the influence of open source culture on open data activists. This will be the foundation for the following analysis, in which I will show how open data activists modulate open source practices by applying them to data. Inspired by the social analytics approach, this analysis will ask, first, what their practices and imaginaries tell us about the *conditions* under which datafication might support the agency of publics and, second, how datafication supports the agency of activists themselves. In the conclusion, I will reflect on the broader relevance of this type of activism for the ongoing datafication of social life.

Tracing the influence of open source culture on open data activists

It is generally acknowledged that activism around open data is rooted in hacking culture (cf. Bates, 2012; Davies, 2010; Johnson, 2014), or more specifically open source culture as one of the most prominent genres of hacking (Coleman and Golub, 2008). However, while this connection is frequently pointed out, it is rarely examined in more detail. Authors usually refer to a set of broad ethical commitments taken directly from traditional hacker culture. These ethical commitments have been famously described by Levy (1984) as follows: access to computer technology and information should be free, centralized forms of power are rejected in favor of decentralization, hackers adhere to a meritocratic culture of technological excellence in which the hacker should only be judged by his or her code, and the belief that computers can create a ‘better world’. While these principles are indeed relevant, we run the risk of oversimplifying the relationship between open data activism and open source culture if we solely rely on them. As Coleman (2013: 17) points out, the frequent reference to Levy’s account is problematic because it ‘whitewashes’ the diversity among hackers. While hackers share some technical and ethical commitments (for which Levy’s description is still useful), hacker culture should not be treated as a ‘singular code

formulated by some homogeneous group called hackers but instead as a composite of distinct yet connected moral genres' (Coleman, 2013: 19).

To develop a more nuanced and differentiated picture of how open data activists draw from open source culture, we can turn to research on its broader cultural significance and influence beyond software development. Particularly helpful here is an approach developed by Kely (2008). While most attempts to grasp this phenomenon are primarily interested in making generalizations—for example by asking whether diverse initiatives rooted in open source culture are forming a coherent movement with a political project (cf. Beyer, 2014b; Clement and Hurrell, 2008; Kapczynski and Krikorian, 2010) or how the organizational features of open source software development can be generalized and applied to new domains (cf. Demil and Lecocq, 2006; Matei and Irimia, 2014; Weber, 2004)—Kely developed a model that can be used to trace the influence of open source culture for specific cases. In his study of the cultural significance of free software,³ Kely suggests that open source advocates associate with each other not just through a set of ethical commitments, but through a range of key practices and social imaginaries (Taylor, 2004). He understands open source as an experimental system made up of five key practices or 'components': sharing source code, defining openness, writing copyright licenses, coordinating collaborations, and forming a movement. Understood in this way, open source becomes 'a system of thresholds, not of classification' (Kely, 2008: 16):

Within each component are a range of differences in practice, from conventional to experimental. At the center, so to speak, are the most common and accepted versions of a practice; at the edges are more unusual or controversial versions. (Kely, 2008: 15)

Due to their flexibility, these components are not exclusive to the development of software: each of these practices can be adapted or 'modulated' to apply them to other domains. Therefore, Kely (2008: 246) calls initiatives like Wikipedia or Creative Commons *modulations* of open source that emerge 'out of a direct engagement with and exploration of Free Software' and are 'committed to experimenting with the given practices of Free Software'. Creative Commons, for example, paralleled some of the work of the Free Software Foundation in a different context (Garcelon, 2009: 1315): it modulates the practice of 'sharing source code' by applying it to 'content', it is writing copyright licenses for this new type of 'source code', and it has become a movement as well.

To trace the influence of open source culture, Kely (2008: 278) suggests treating its key practices as a

template that interacts with other forms of knowledge management: 'Where the practices match, no change occurs, and where they don't, it is the reorientation of knowledge and power'. Therefore, the proliferation of open source culture can be described as the proliferation and modulation of its key practices in order to alter the means of knowledge production and circulation. Tracing the influence of open source culture on open data activists then comes down to a set of specific questions: Which practices are modulated? How are they modulated? How does this change the domain to which they are applied? Answering these questions will help us to grasp how activists try to apply the more transparent and collaborative forms of governance associated with open source to politics, and how this might support the agency of datafied publics.

Practices and imaginaries of open data activists⁴

The following analysis is based on 10 semi-structured interviews with members of the OKF DE core team (including the chairman and founder, main developers, committee members and project managers) and a content analysis of nine relevant documents that were selected using a theoretical sampling, for example self-portraying descriptions from the official homepage. The data was collected in three rounds between September 2012 and January 2013 and analyzed using a grounded theory approach (Glaser and Strauss, 1967). I will structure the presentation of the findings in a way that shows how one modulation of open source culture leads activists to other, subsequent adaptations and interpretations.

1. By regarding data as a prerequisite for generating knowledge, activists transform the sharing of source code to include the sharing of raw data. Sharing raw data would allow others to make their own interpretation of it and generate their own knowledge, which represents a 'democratization of information' for activists.
2. Seeing information as a necessary precondition for political participation, activists connect this idea to an open and flexible form of representative democracy by applying the open source model of participation (the 'bazaar model') to political participation, which should lead to more participation of citizens in political decision-making processes and more active and engaged local communities.
3. A third set of practices refers to activists' acknowledgment that raw data needs to be 'refined' to create knowledge for citizens, which is why they seek to create, and become, 'data intermediaries' for the

public. This leads them to a special interest in journalism.

In the following, I explain each of these modulations and their implications in more detail.

Raw data as source code

The overall mission of the *Open Knowledge* Foundation is already implicated in its name. The organization adapts a hierarchical understanding of the relationship between data, information, and knowledge that is common in knowledge management literature (Tuomi, 1999). According to this model, (raw) data are understood as symbols that have not been interpreted; data becomes information when it is structured and put into context; and information becomes knowledge when it is interpreted, meaningful, and actionable. As the OKF explains, open knowledge *'is what open data becomes when it's useful, usable and used'* (Open Knowledge, n.d. a). With its name, the OKF thus indicates that it aims at spreading not just open data, but open knowledge. However, data is seen as a *prerequisite* for generating knowledge. This hierarchical understanding leads activists to the first and most fundamental modulation of open source culture: to conceive raw data as source code that should be shared openly to allow others to interpret it and to generate their own knowledge from it.

Implicit here is that activists do not simply modulate the practice of sharing source code by replacing code with data. They also adapt the metaphors and concepts behind this practice. To execute human-readable source code on a computer, it has to be translated into binary instructions that are only readable by machines. These binary instructions cannot be retranslated back into the source code from which they have been generated. Having only the binary code without the source code (which is the case for most proprietary software) means that it is not possible to understand or modify the 'inner workings' of the software. Similarly, open data activists treat raw data as source code and interpretations—or knowledge—as binary code. As one activist explains, raw data 'is not really neutral' but it allows more interpretation than a 'summary or a press conference' (Interview: Developer 1).⁵ That is because summaries are *already* interpretations of raw data. Only offering an interpretation of raw data without allowing access to it would make it difficult for others to understand how this interpretation was developed. Governments would then maintain a 'monopoly of interpretation'. Sharing raw data makes the process of interpreting it transparent and breaks governments' monopoly, which means that everybody could make his or her own interpretation of the data that governments use to make and justify their decisions—allowing people to examine

biases in government's data collection and interpretation. For activists, open data therefore represents a democratization of interpretation or—as they put it—a 'democratization of information' (Interview: Chairman & Founder).

It is interesting to contrast the notion of 'raw data' developed by activists with the way the term is used in discussions about Big Data. 'Raw data is an oxymoron' (Gitelman, 2013) is one of the most common critiques of Big Data advocates' belief in 'objective quantification' (Van Dijck, 2014: 198) or Big Data's 'aura of truth, objectivity, and accuracy' (Boyd and Crawford, 2012: 664). In their critique, authors point out that data is always prefigured through gathering mechanisms (Van Dijck and Poell, 2013: 10) and collected data has to be interpreted to make it meaningful and actionable, a process guided by specific interests and rationalities and not something that can be considered as objective. Essentially, this questions whether something like 'raw' data actually exists when we understand it as something 'pure' beyond human influence. However, members of the OKF DE adapt a different understanding of 'raw' data. For them, 'raw' simply means 'as collected'. Accordingly, sharing data in 'raw' form—as collected—is not about revealing an unbiased and objective truth, but about making the biases of this data *transparent* and allowing 'more interpretation of truth' (Interview: Chairman & Founder).

Using this understanding as a basis, members of the OKF DE are also concerned with the conditions that must be met to ensure this type of transparency, i.e. with the way raw data has to be provided to fulfill their vision of a democratization of information. This leads them to another modulation of open source practices: defining openness. More specifically, they define both the legal and technical characteristics of openness in relation to data in order to delineate open data from 'closed' data. For legal openness, the OKF developed the international Open Definition (Open Knowledge, n.d. b), according to which data is 'open' when it can be accessed, modified and shared by anyone for any purpose without restrictions. Technical openness is about ensuring that these rights can be exercised without too much effort. Key here are the principles developed by the Sunlight Foundation (2010) and the rating system developed by Tim Berners-Lee (2010). According to these guidelines, datasets should be complete, released in a timely fashion, accessible, machine readable, and available in open formats. While activists acknowledge that personal data and data crucial to security should not be made available in this way, they suggest that these legal and technical conditions are necessary to effectively break the interpretative monopoly of governments.

Given the importance of knowledge for agency, this type of transparency has the potential to support the

agency of datafied publics. As activists acknowledge themselves, however, the mere provision of raw data is insufficient and only represents ‘the first step’ (Interview: Chairman & Founder). As I will explain in the following sections, this provision should go along with more continuous and flexible forms of participation and ‘data intermediaries’ that make raw data accessible to the public.

Data and democracy

The democratization of information described above is not regarded as an end in itself by activists. Ultimately, this form of transparency is taken as a means through which ‘the people should be considered again as the sovereign’ (Interview: Project Manager 1). Even though they do not explicitly talk about agency themselves, activists’ articulations of their broader aims are interesting for understanding *how* and under what *conditions* the democratization of information they envision could support the agency of datafied publics. The overall ‘vision’ of OKF DE members is essentially a vision of citizen empowerment: sharing raw data should help citizens to better understand and control their governments and to be more active and engaged in their local communities.

This means that more possibilities for citizens to participate in political decision-making processes is a major goal for members of the OKF DE: ‘to participate, people need information’ (Interview: Project Manager 2). In this respect, they regard themselves as part of an ‘Internet generation’ that is not content with periodic voting: ‘[I want] a higher degree of participation...a more continuous form of participation’ (Interview: Committee member). This does not, however, necessarily translate into a demand for more direct democracy. Instead, the open source model of participation is taken as a paradigm:

What is powerful about open source development is that people can elect themselves as participants. I mean people can find my project and then decide for themselves to participate in its development and contribute to it. I think this model of self-selective participation is extremely powerful and I believe it can be applied to politics. (Interview: Developer 1)

This means that everybody who wants to participate in the decision-making process of a particular issue should have the opportunity to do so in a meaningful way. Here, activists explicitly modulate another practice from open source culture: coordinating collaborations, the organization of open source projects (Kelty, 2008, ch. 7). As mentioned above, this organizational model has been described as the ‘bazaar

model’ (Raymond, 2001) because it encourages and incorporates contributions from potentially everyone. Just as there is not one standard model for coordinating collaborations in open source—larger and more well-known projects like the Linux kernel, the Apache servers, or the Debian project have all developed distinct organizational models over time (Coleman, 2013; Kelty, 2008; Weber, 2004)—activists reject clearly prescribing a specific model of participation. For them, applying the bazaar model of open source to governance is first and foremost about experimentation. There will not be ‘this one solution that you just need to apply. I think public authorities will need to have the courage to experiment’ (Interview: Chairman & Founder). This illustrates that more participation is not seen as a natural outcome of open data. Activists argue that it requires a cultural change within public institutions: a change towards a ‘beta culture’ that is willing to experiment and risk failure (Schwegmann, 2012), and a more collaborative and less authoritative relationship with citizens. Public institutions, it is argued, should promote the use of data and actively include citizens in decision-making processes: ‘It is not just about opening data...but also about investments from public institutions to ensure that this data is used’ (Interview: Committee member). Activists think that this cultural change will mainly happen at a local level, where issues are ‘closer’ to the people and institutions can experiment with ‘less resources’ (Interview: Chairman & Founder).

Taken together, the way activists apply the open source model of participation to governance results in a notion of a more open and flexible form of representative democracy. ‘Open’ refers to a higher degree of transparency (by sharing raw data) and the openness of political decision-making processes for public participation. ‘Flexible’ means that activists think that the inclusion and coordination of citizens’ voluntary, ‘self-selective participation’ should be adapted to the issue at hand and to the local context. At the same time, activists do not question representative democracy as such and are rather skeptical about elements of direct democracy, e.g. referendums: ‘I don’t know if direct democracy is always the right answer...but I definitely want more mechanisms to involve people more often’ (Interview: Chairman & Founder). From the perspective of democratic theory, they negotiate between representative models of democracy—in which participation is mainly limited to periodic voting—and direct models of democracy, where entire electorates vote on certain proposals. This is similar to Barber’s (2004) model of ‘strong democracy’, a more explicit attempt to develop an alternative to representative and direct democracy. Put briefly, strong democracy is based on a ‘creative consensus’ that is meant to

recognize the diversity of interests and ‘is premised on citizens’ active and perennial participation in the transformation of conflict through the creation of common consciousness and political judgment’ (Barber, 2004: 224). Similarly, the diverse and flexible modes of organizing voluntary participation envisioned by OKF DE members require the active involvement of citizens and imply a consensus building process that is ‘creative’ in negotiating diverse interests *and* in its organization.

We can summarize the ideas and aims of open data activists described thus far to articulate—as an intermediary result—a first proposal about the conditions that must be met to support the agency of datafied publics: the transparency created through the sharing of raw data should be accompanied by a cultural change within public institutions to support voluntary and flexible forms of participation similar to those found in open source projects. As I will detail in the next section, activists not only emphasize the importance of public institutions, but also of other intermediaries to facilitate this participation.

Creating empowering intermediaries: Complementing or replacing journalism?

Even though the idea behind the democratization of information is to *potentially* allow *everybody* to interpret raw data, activists are well aware that the average citizen does not have the time and expert knowledge to do so. They recognize that their vision of empowerment through open data can only be realized with intermediaries that make raw data accessible to the public.⁶ Different to the modulations discussed above, this is not directly reflected by open source culture. However, it emerged out of activists’ engagement, exploration, and modulation of open source, i.e. of modulating the sharing of source code to include the sharing of raw data and of using the open source model of participation as a paradigm for political participation. Because activists realize the importance of suitable intermediaries for their goals, they actively seek to ‘create’ them—which makes this aspect also interesting for understanding how datafication supports the agency of activists themselves.

In terms of agency, more interesting than the basic acknowledgment that intermediaries are necessary is what kind of intermediaries are deemed necessary to empower citizens. Three criteria can be identified that constitute an ‘empowering intermediary’ in the eyes of activists. First, they should be *data-driven*, which means that they should be able to handle large and complex datasets to make them accessible to others. Second, empowering intermediaries should be *open*, which means that they should make the data from which they generate stories or build applications available to

their audiences—the principle of sharing raw data applies here as well. The fact that professional journalists or NGOs often do not give access to their sources is therefore frequently criticized, one activist calling it a ‘fundamental bug of newspapers’ (Interview: Developer 1). Third, empowering intermediaries should be *engaging*, which means that they should actively involve citizens in public issues. This implies that such intermediaries should not only be information providers and that they should have a cooperative relationship with their audiences: ‘journalism also needs to change to be closer to citizens’ (Interview: Developer 2). Taken together, the three criteria articulated by activists are clearly related to their goal of an open and flexible democracy described above. Empowering intermediaries are a necessary prerequisite in this sense, or, as we might say in relation to agency, they are important supporters of agency in datafied publics.

To ‘create’ these intermediaries, activists try to cooperate with other NGOs and professional journalists and offer teaching. Here, they are part of a larger phenomenon: the increased interaction between the social worlds of technology and journalism, or more specifically between hackers and journalists (Karlsen and Stavelin, 2014; Lewis and Usher, 2013; Parasie and Dagiral, 2013; Royal, 2010). Members of the OKF DE are involved, for example, in Hacks/Hackers events (Lewis and Usher, 2014), where hackers and journalists come together to innovate news; the News Challenge of the Knight Foundation, an open-to-all contest rewarding projects that aim to transform news and information distribution (Lewis, 2012a); or the Knight-Mozilla Fellowships, which bring together hackers and technologists ‘to spend 10 months working on open source code with partner newsroom[s]’ like *The New York Times* or *Der Spiegel* (OpenNews, n.d.). The existing research on the interaction between these groups shows that activists’ goal of a more equal and cooperative relationship between citizens and professional journalists essentially questions the professional boundaries of journalism. As Lewis (2012b) describes, journalism is shaped by a professional logic of (exclusive) control over content that does not align easily with a more participatory form of journalism favored by activists. The research shows an ongoing process of negotiation of different values, imaginations, and practices: news organizations try to re-interpret the distinct way activists think about technology and data ‘into the language of news’ (Lewis and Usher, 2013: 604), while at least some activists tend to believe the new possibilities are ‘capable of altering the very nature of journalism’ (Parasie, 2011).

However, activists do not only try to influence journalism by interacting with professional journalists or by becoming programmer-journalists in newsrooms. They also act as intermediaries outside the profession and

develop independent, non-profit applications to ‘implement’ their ideas. Key here are so called ‘civic technologies’—small-scale, specialized applications that aim to ‘connect people’ (Interview: Developer 1). These applications are either about improving government services for citizens, or about helping citizens to coordinate with each other to solve problems together. Often these are relatively simple web applications that focus on one task. For example, there are civic technologies that help people to exchange deposit bottles, that show how and where to engage in local building projects, that inform people about the local air quality, visualize which parts of the city are barrier-free and which are not, and so forth.⁷ Even though civic technologies do not always depend on open data, data is key to their functioning in two ways: first, the availability of open data creates more opportunities to develop civic technologies (for example, when they require traffic data); second, they often datafy the activities they are concerned with, i.e. they often create new data. For example, FragDenStaat.de (inspired by the British WhatDoTheyKnow⁸) makes it easier to submit freedom of information requests to public authorities and tracks both the requests and the responses from institutions. This crowdsourcing approach created a database that can be used to analyze and compare how different institutions react to these requests, what kind of requests are more likely to get refused and so forth. This illustrates that the development of civic technologies is not only interesting because it could support the agency of citizens. It also shows how activists use or create data to meet their own ends by developing tools to put their ideas into practice. For OKF DE members, the purpose of these applications is two-fold. On the one hand, they are supposed to help citizens to be more active and engaged in their local communities in a general sense—for example by helping people with disabilities to move around the city. On the other hand, they hope to create new communities or ‘alternative publics... with a controlling function’ (Interview: Developer 1). An often cited example is Ushahidi,⁹ which was originally developed by a group of citizen journalists to track violent outbreaks after a disputed election in Kenya (Giridharadas, 2010). Because journalists received threats about their work, Ushahidi was designed as a crowdsourcing application that maps incidents reported anonymously by users. Both in the sense of more active and engaged citizens and of ‘controlling publics’, civic technologies are linked to a notion of ‘self-empowerment’ (Interview: Chairman & Founder) or ‘do-it-yourself-empowerment’ through data, understood as the ability of citizens to solve issues without the help of governments or businesses.

In terms of agency, the development of civic technologies by activists is interesting for another, less

obvious reason. Civic technologies can be described as alternative ways of fulfilling functions traditionally described as ‘journalistic’ (making governments more transparent and accountable and engaging citizens in public issues) or of accessing and using public services (e.g. with an easy-to-use website to submit freedom of information requests). In other words, these applications are developed *independently outside* professional journalism or public institutions, but at the same time are trying to fulfill *similar functions*. This ability to create ‘actually existing alternatives’ is characteristic of the political power of hacking in general, as Kelty (2008) illustrates. He argues that open source advocates (and modulations like the open data movement) create independent ‘recursive publics’ through their key practices:

A recursive public is a public that is vitally concerned with the material and practical maintenance and modification of the technical, legal, practical, and conceptual means of its own existence as a public; it is a collective independent of other forms of constituted power and is capable of speaking to existing forms of power through the production of actually existing alternatives. (Kelty, 2008: 3)

By being able to maintain their own terms of existence (to a certain degree at least), recursive publics can act as ‘actually existing alternatives’. In this sense, civic technologies developed by activists could to some degree act as ‘actually existing alternatives’ to professional journalism or (ways of accessing) public services. Activists are well aware of this potential: The ultimate goal of developing alternative services with civic technologies is to pressure established institutions to adapt them. ‘Flagship projects’ (Interview: Chairman & Founder) are intended to demonstrate what is possible and to invite (or provoke) established institutions to imitate them. As one member notes: ‘We have discovered software as a lobbying tool’ (Interview: Developer 2). Let me illustrate this with another example: Frankfurt-Gestalten.de (‘Shaping-Frankfurt’) monitors information provided by local parliaments in the city of Frankfurt and illustrates them on a map. Users can check what is currently discussed in their street or district (e.g. building projects), comment on it or initiate new discussions. Activists use this project to advocate for easier access to local parliamentary data, and for local public institutions to offer similar services. Moreover, I suggest that applications like Frankfurt-Gestalten.de represent a data-driven form of local journalism that is focused on engaging citizens on a local level. As such, Frankfurt-Gestalten.de has a complex relationship with professional journalism: First, it could complement professional journalism

because local journalists can use it as a research tool. Secondly, however, it also represents a potential threat for professional local journalism—if people use an application like Frankfurt-Gestalten.de instead of consulting their local news media. Yet it is also conceivable, thirdly, that news media develop and maintain similar applications themselves, offering them as services to their audience and using them as research tools for their own investigations—Bell (2014) recently made a similar suggestion. This example illustrates how activists attempt to directly or indirectly influence established institutions on many different levels through the development of civic technologies, and shows that acting as intermediaries themselves is as much about directly putting ideas into practice as it is about transforming existing institutions. It not only shows how activists use data to directly meet their own ends, but also how they attempt to influence the conditions of the wider public to support the agency of ordinary citizens.

Conclusion: Data hacking and new forms of agency?

I conclude by returning to the questions raised in the beginning of this article. What do the practices and values developed by members of the OKF DE tell us about the conditions under which datafication can support agency?

When we look at activists themselves, datafication obviously does not undermine, but rather supports their agency in important ways: their technological expertise enables them to utilize or create data to meet their own ends. They even use the applications they create as lobbying tools that pressure institutions by offering actually existing alternatives. These findings emphasize the connection between datafication and the proliferation of hacking culture. The ability to ‘hack’ and to create recursive publics fundamentally depends on the availability and modifiability of the underlying technology (Kelty, 2008: 10–11): participants have to be able to access and modify the technology needed to build their own, independent infrastructures. Otherwise, the expressive use of technology—the expression of imaginaries, values and rationalities *through* technology—would not be possible. The process of ‘datafying’ a phenomenon—of transforming it into quantifiable information—can be an integral part of recursive publics in itself, as illustrated by civic technologies that collect data via crowdsourcing. More importantly, to datafy a phenomenon is to re-materialize it into a highly modifiable form: in its essence, data is structured information that can be analyzed, edited, and combined with other data. This is why the availability of data creates more opportunities

for the development of software that utilizes it in new ways, like activists do with the civic technologies they create. By rendering phenomena into data that have never been quantified before, datafication can make the key practices of recursive publics applicable to them in ways that have not been possible before—given that the data created in these processes is accessible and modifiable. Therefore, datafication also has the potential to extend both the possibilities for and the scope of recursive publics, i.e. of creating ‘actually existing alternatives’ to established forms of knowledge production and circulation.

Moreover, members of the OKF DE are primarily concerned with how they can support democratic values and the agency of citizens through open data. As I showed in this article, three interrelated conditions must be met in their eyes: raw data should be shared openly to make decision-making processes more transparent, public institutions should actively include citizens in these decision-making processes to create a more open and flexible form of representative democracy, and ‘empowering intermediaries’ are needed to make raw data accessible to the wider public. It seems clear that these propositions have a *potential* to remedy the issues identified by Couldry and Powell (2014), i.e. the danger that Big Data technologies undermine agency by disconnecting system and experience. However, it is of course important and necessary to critically examine these ideas. For example, how to ensure that the raw data provided by governments does not violate privacy and is free of manipulations? Can voluntary participation of citizens work at larger scales? How to ensure that these processes do not end up ‘empowering the empowered’ (Gurstein, 2011)? We also need to be wary about the idealism of activists and the high level of technical literacy required to get involved in data activism. It might be easy to criticize activists as naive or techno-deterministic when we point out all these potential issues. However, while there is much to be critical of, I suggest that it is equally important to study and understand the practices and ideas of activists in order to evaluate what we can learn from them. Activists aim to develop new rationalities and alternative social imaginaries around datafication to *connect* system and experience in new ways and to create a new sense for the legitimacy of collective knowledge creation and distribution in democratic, datafied publics. For all their shortcomings, these attempts are relevant and deserve attention because they provide a vital starting point to discuss how we can counter the threats of Big Data and utilize the potential of these new technologies in ways that do not damage democratic values and the agency of those not in big government or big business. As Couldry and Powell (2014: 4–5) point out, we should not only highlight the risks of creating and sharing data, but also

the opportunities for forms of social organization that take ‘into account agents’ practices of giving an account of themselves and their conditions of life’.

A guidance for future research provided by this analysis is to look at the way activists’ practices and ideas are institutionalized, i.e. how they are adapted by other NGOs, news media, or public institutions. As activists acknowledge themselves with their emphasis on the importance of empowering intermediaries, their influence on the wider public—and therefore their potential to support the agency of datafied publics—depends on transforming existing institutions rather than on building new, alternative ones. To study these processes, we can take further inspiration from Couldry’s (2010: 1) concept of ‘effective voice’—the insurance that ‘my voice matters’, which is a crucial aspect for both agency and democratic legitimacy. We can argue that activists describe important preconditions for processes of effective voice in datafied societies. Yet we have to be critical about whether the adaption of activists’ practices and ideas really leads to *effective* voice, or only to more opportunities to *raise* voice. What matters is how ‘people’s practices of voice are *sustained* and the outcomes of those practices *validated*’ (Couldry, 2010: 113). We have to ask whether and how the adaption of activists’ practices and ideas by news media, public institutions, or others does or does not lead to structures that create and sustain the conditions necessary for effective voice in increasingly datafied societies. Such research can form the basis to further examine, refine, and extend the practices and imaginaries of activists to formulate in more detail the conditions necessary to support agency in the ‘age of Big Data’.

Acknowledgements

I wish to thank Tamara Witschge for her thoughts and feedback on earlier versions of this article, Andreas Hepp and Andreas Breiter for their supervision and support of the original study, as well as the editors and anonymous reviewers for their constructive and helpful critique.

Declaration of conflicting interest

The author declares that there is no conflict of interest.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Notes

1. The social analytics approach was first developed in the Storycircle project to study how organizations use analytics to meet their goals. See <http://storycircle.co.uk/>
2. <http://okfn.de/>. The OKF DE was founded in 2011 as the first international chapter of the British Open Knowledge Foundation, which was founded in 2004 in Cambridge. It

was recently renamed ‘Open Knowledge’. See <https://okfn.org/>

3. ‘Free software’ and ‘open source’ generally refer to the same practices. While the term free software emphasizes social and cultural values (‘free as in speech’), open source emphasizes the practical advantages for developing software (Kelty, 2008). For convenience, I will use ‘open source’ to address both strands.
4. Some of the empirical findings presented here were published in German by the author (Baack, 2013).
5. Quotes from the interviews were translated from German by the author.
6. I adapt activists’ usage of the term intermediary in this article. However, ‘intermediaries’ in their sense are more likely to act as mediators according to Latour’s (2005: 39) distinction of the terms.
7. See a list of projects supported by the OKF DE at <http://codefor.de/projekte/>
8. <https://www.whatdotheyknow.com/>
9. <http://www.usahidi.com/>

References

- Baack S (2013) Die Open-Data-Bewegung. Das Verhältnis von Praktiken, Zielen und Selbstbild der Open Knowledge Foundation Deutschland. University of Bremen. Available from: <http://nbn-resolving.de/urn:nbn:de:0168-ssoar-363745> (accessed 3 March 2014).
- Barber BR (2004) *Strong Democracy: Participatory Politics for a New Age*. Berkeley, CA: University of California Press.
- Bates J (2012) ‘This is what modern deregulation looks like’: Co-optation and contestation in the shaping of the UK’s Open Government Data Initiative. *The Journal of Community Informatics* 8(2). Available at: <http://ci-journal.net/index.php/ciej/article/view/845> (accessed 3 October 2014).
- Bell E (2014) Silicon valley and journalism: Make up or break up? *Reuters Institute for the Study of Journalism*. Available at: <http://reutersinstitute.politics.ox.ac.uk/news/silicon-valley-and-journalism-make-or-break> (accessed 22 November 2014).
- Berners-Lee T (2010) Linked data. Available at: <http://www.w3.org/DesignIssues/LinkedData.html> (accessed 26 January 2015).
- Beyer JL (2014a) *Expect Us: Online Communities and Political Mobilization*. Oxford: Oxford University Press.
- Beyer JL (2014b) The emergence of a freedom of information movement: Anonymous, WikiLeaks, the Pirate party, and Iceland. *Journal of Computer-Mediated Communication* 19(2): 141–154.
- Boyd D and Crawford K (2012) Critical questions for Big Data. *Information, Communication & Society* 15(5): 662–679.
- Clement A and Hurrell C (2008) Information/Communications rights as a new environmentalism? Core environmental concepts for linking rights-oriented computerization movements. In: Elliott MS and Kraemer KL (eds) *Computerization Movements and Technology Diffusion: From Mainframes to Ubiquitous Computing*. Medford, NJ: Information Today, Inc.,

- pp. 337–358. Available at: <http://hdl.handle.net/1807/35202> (accessed 19 October 2014).
- Coleman G (2013) *Coding Freedom: The Ethics and Aesthetics of Hacking*. Princeton, NJ: Princeton University Press. Available at: http://codingfreedom.com/buy_download.html.
- Coleman G (2014) *Hacker, Hoaxer, Whistleblower, Spy: The Many Faces of Anonymous*. 1st ed. London: Verso.
- Coleman G and Golub A (2008) Hacker practice: Moral genres and the cultural articulation of liberalism. *Anthropological Theory* 8(3): 255–277.
- Couldry N (2010) *Why Voice Matters: Culture and Politics After Neoliberalism*. Los Angeles: Sage.
- Couldry N (2014) Inaugural: A necessary disenchantment: Myth, agency and injustice in a digital world. *The Sociological Review* 62(4): 880–897.
- Couldry N and Powell A (2014) Big Data from the bottom up. *Big Data & Society* 1(2). Available at: <http://bds.sagepub.com/content/1/2/2053951714539277> (accessed 7 July 2014).
- Davies T (2010) *Open data, democracy and public sector reform: A look at open government data use from Data.gov.uk*. MSc Dissertation, University of Oxford. Available at: <http://www.opendataimpacts.net/report/> (accessed 12 November 2014).
- Demil B and Lecocq X (2006) Neither market nor hierarchy nor network: The emergence of bazaar governance. *Organization Studies* 27(10): 1447–1466.
- Garcelon M (2009) An information commons? Creative commons and public access to cultural creations. *New Media & Society* 11(8): 1307–1326.
- Giridharadas A (2010) Ushahidi – Africa’s gift to Silicon Valley: How to track a crisis. *The New York Times*, 13 March. Available at: <http://www.nytimes.com/2010/03/14/weekinreview/14giridharadas.html> (accessed 12 February 2015).
- Gitelman L (2013) *Raw Data is an Oxymoron*. Cambridge, MA: MIT Press.
- Glaser BG and Strauss AL (1967) *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Chicago: Aldine Pub. Co.
- Gurstein M (2011) Open data: Empowering the empowered or effective data use for everyone? *First Monday* 16(2). Available at: <http://firstmonday.org/ojs/index.php/fm/article/view/3316> (accessed 11 November 2014).
- Johnson JA (2014) From open data to information justice. *Ethics and Information Technology* 16(4): 263–274.
- Kapczynski A and Krikorian G (eds) (2010) *Access to Knowledge in the Age of Intellectual Property*. New York, NY: Zone Books. Available at: http://www.zonebooks.org/titles/KRIK_ACC.html (accessed 3 October 2014).
- Karlsen J and Stavelin E (2014) Computational journalism in Norwegian newsrooms. *Journalism Practice* 8(1): 34–48.
- Kelty CM (2008) *Two Bits: The Cultural Significance of Free Software*. *Experimental Futures*. Durham: Duke University Press. Available at: <http://twobits.net/read/> (accessed 4 March 2014).
- Latour B (2005) *Reassembling the Social: An Introduction to Actor-Network-Theory*. Clarendon lectures in management studies. Oxford: Oxford University Press.
- Levy S (1984) *Hackers: Heroes of the Computer Revolution*. 1st ed. Garden City, NY: Anchor Press/Doubleday.
- Lewis SC (2012a) From journalism to information: The transformation of the knight foundation and news innovation. *Mass Communication and Society* 15(3): 309–334.
- Lewis SC (2012b) The tension between professional control and open participation. *Journalism and its boundaries. Information, Communication & Society* 15(6): 836–866.
- Lewis SC and Usher N (2013) Open source and journalism: Toward new frameworks for imagining news innovation. *Media Culture Society* 35(5): 602–619.
- Lewis SC and Usher N (2014) Code, collaboration, and the future of journalism. *Digital Journalism* 2(3): 383–393.
- Lievrouw LA (2011) *Alternative and Activist New Media*. Digital media and society series. Cambridge, UK: Polity.
- Matei A and Irimia SI (2014) Open source governance. A more ambitious cousin of collaborative governance. *International Journal of Public Administration* 37(12): 812–823.
- Mayer-Schoenberger V and Cukier K (2013) *Big Data: A Revolution that will Transform How We Live, Work and Think*. London: Murray.
- Open Knowledge (n.d. a) What is open? Available at: <https://okfn.org/opendata/> (accessed 11 February 2015).
- Open Knowledge (n.d. b) The open definition. Available at: <http://opendefinition.org/> (accessed 30 January 2015).
- OpenNews (n.d.) The Knight-Mozilla fellowships. Available at: <http://opennews.org/what/fellowships/> (accessed 8 May 2015).
- Parasie S (2011) ‘Hacker’ journalism – A new utopia for the press? *Books & Ideas*. Available at: <http://www.booksandideas.net/Hacker-Journalism-A-New-Utopia-for.html> (accessed 3 March 2014).
- Parasie S and Dagiral E (2013) Data-driven journalism and the public good: ‘Computer-assisted-reporters’ and ‘programmer-journalists’ in Chicago. *New Media & Society* 15(6): 853–871.
- Raymond ES (2001) *The Cathedral & the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary*. Cambridge, MA: O’Reilly.
- Royal C (2010) The journalist as programmer: A case study of the New York Times interactive news technology department. In: *The International Symposium in Online Journalism*, Austin, TX, April.
- Sauter M (2014) *The Coming Swarm: DDOS Actions, Hacktivism, and Civil Disobedience on the Internet*. New York, NY: Bloomsbury Academic.
- Schwegmann C (2012) It’s a fad! 7 reasons against open data. In: *Open Knowledge Foundation Deutschland*. Available at: <http://okfn.de/blog/2012/11/7-reasons-against-open-data/> (accessed 4 July 2015).
- Sunlight Foundation (2010) Ten principles for opening up government information. Available at: <http://sunlight>

- foundation.com/policy/documents/ten-open-data-principles/ (accessed 30 January 2015).
- Taylor C (2004) *Modern Social Imaginaries*. Durham: Duke University Press.
- Tuomi I (1999) Data is more than knowledge: Implications of the reversed knowledge hierarchy for knowledge management and organizational memory. *Journal of Management Information Systems* 16(3): 103–117.
- Van Dijck J (2014) Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. *Surveillance & Society* 12(2): 197–208.
- Van Dijck J and Poell T (2013) Understanding social media logic. *Media and Communication* 1(1): 2–14.
- Weber S (2004) *The Success of Open Source*. Cambridge, MA: Harvard University Press.