*Data & Policy* (2020), 2: e5
doi:10.1017/dap.2020.5

CAMBRIDGE UNIVERSITY PRESS

COMMENTARY

# Critical big data literacy tools—Engaging citizens and promoting empowered internet usage

Ina Sander[1,2,*]

[1]Data Justice Lab, School of Journalism, Media and Culture, Cardiff University, Cardiff, United Kingdom
[2]Center for Advanced Internet Studies (CAIS), Bochum, Germany
*Corresponding author. Email: Sanderi@cardiff.ac.uk

**Received:** 02 August 2019; **Revised:** 30 January 2020; **Accepted:** 20 April 2020

**Keywords:** critical data literacy; critical data studies; critical digital literacy; digital resignation; online privacy attitudes

## Abstract

Datafied societies need *informed public debate* about the implications of data science technologies. At present, internet users are often unaware of the potential consequences of disclosing personal data online and few citizens have the knowledge to participate in such debates. This paper argues that critical big data literacy efforts are one way to address this lack of knowledge. It draws on findings from a small qualitative investigation and discusses the effectiveness of online *critical big data literacy tools.* Through pre and post use testing, the short- and longer-term influence of these tools on people's privacy attitudes and behavior was investigated. The study's findings suggested that the tools tested had a predominantly positive initial effect, leading to *improved critical big data literacy* among most participants, which resulted in *more privacy-sensitive attitudes* and *internet usage.* When analyzing the tools' longer-term influence, results were more mixed, with evidence suggesting for some that literacy effects of the tools were *short-lived*, while for others they led to more *persistent and growing literacy.* The findings confirm previous research noting the *complexity of privacy attitudes* and also find that *resignation* toward privacy is multi-faceted. Overall, this study *reaffirms the importance* of critical big data literacy and produces new findings about the value of interactive data literacy tools. These tools have been under-researched to date. This research shows that these tools could provide a relevant means to work toward empowering internet users, promoting a critical internet usage and, ideally, enabling more citizens to engage in public debates about changing data systems.

## Policy Significance Statement

In our current "age of datafication," the collection of personal data online and the impact of big data systems are ubiquitous and ever-growing. In order to make responsible decisions on the regulation of data usage and big data systems in various areas of society, datafied societies need *informed public debate* about the implications of data science technologies. This research suggests that one way to enable such debate is to foster citizens' *critical big data literacy.* The article introduces this concept, presents ways to foster such literacy, and discusses the effect these efforts have on internet users. Overall, it argues that online data literacy tools can provide a substantial means to work toward empowering internet users and engaging citizens in public debates about data systems.

## Introduction

Given the ever-growing relevance of big data and automated systems in all areas of society, many argue that we have entered an age of "datafication" (e.g., Gray et al., 2018; Hintz et al., 2018; Mayer-Schönberger & Cukier, 2013; Redden, 2018). With private companies and governments worldwide collecting and analyzing vast amounts of—amongst others—personal data and the use of automated systems increasing rapidly, there has been a "profound transformation in how society is ordered, decisions are made, and citizens are monitored through 'big data'" (Hintz et al., 2018, p. 2f). Big data is a contested term and no universal definition has yet emerged (Zuboff, 2015, p. 75). However, many scholars emphasize not only big data's technical characteristics, but also its social origins, viewing big data as a "socio-technical phenomenon" that is "less about data that is big than it is about a capacity to search, aggregate, and cross-reference large data sets" (boyd and Crawford, 2012, p. 663). Such understanding of big data as an interplay of technology, analysis, and mythology—the "aura of truth, objectivity, and accuracy" (boyd and Crawford, 2012, p. 663) that is often associated with data systems—has guided this study. Academic research in recent years has challenged conceptions of big data as inherently benign and objective and has instead highlighted how these systems can be used to profile, target and socially sort citizens in ways that can affect their financial circumstances, ability to find employment, and access to essential services (e.g., Eubanks, 2018; O'Neil, 2016; Redden & Brand, 2017).

As large-scale data collection and related algorithmic systems increasingly transform the public sector, a critical public debate about data practices is essential. Citizens need to recognize some of the risks outlined above and should learn to critically reflect upon these practices of datafication. Only then are they able to assess new technological and societal developments, form opinions, and make informed decisions on their internet usage and about navigating their lives in a datafied society.

However, few citizens have this knowledge and are able to make their voices heard and fulfil their democratic role in datafied societies. Recent research has repeatedly emphasized a knowledge gap amongst American, European, and German internet users on data, algorithms, and online privacy (Grzymek and Puntschuh, 2019; Müller-Peters, 2020; Turow et al., 2018). For example, one study identified a "major understanding gap around technologies" with a quarter of British internet users having "no idea how internet companies make their money" (Doteveryone, 2018, p. 5). This emphasizes that while users may be aware of data collection practices, they often only have a vague idea of how "'the system'—the opaque under-the-hood predictive analytics regimes that they know are tracking their lives but to which they have no access" is operating (Turow et al., 2015, p. 20). Despite such vague knowledge, research shows that many consumers are uncomfortable with algorithmic practices (e.g., Bucher, 2017; Miller et al., 2018; Müller-Peters, 2020).

Consequently, more and more scholars are calling for greater efforts to educate people about big data (Müller-Peters, 2020), invest in "new forms of public engagement and education" (Doteveryone, 2018, p. 6) and increase the public's *literacy* about big data practices (e.g., Grzymek and Puntschuh, 2019; Turow et al., 2018). Yet, there is still a lack of literacy efforts, too little knowledge on what kind of literacy efforts work best and a lack of constructive or comprehensive research on how to address people's lacking knowledge.

## Critical Big Data Literacy

As a response to this gap in research, this study suggests addressing this lack of knowledge and understanding through *critical big data literacy.* This concept builds on and combines approaches from three key fields of research: First, *Critical Data Studies* aim to understand the "importance of data" (Kitchin and Lauriault, 2014, p. 2) and examine the critical effects of datafication on individuals, society, industry, and governance. This work builds on the premise that data and algorithms are never as neutral as they seem (e.g., Dalton and Thatcher, 2014; Eubanks, 2018; Kitchin, 2017; O'Neil, 2016; Zuboff, 2015). Second, concepts of *critical media and digital literacy* originate in the field of education and pedagogy and aim to improve people's knowledge and skills of (digital) media (e.g., Garcia et al., 2015; Hammer,

https://doi.org/10.1017/dap.2020.5 Published online by Cambridge University Press

2011; Hinrichsen and Coombs, 2013). In recent years, an increased emphasis on critical approaches has emerged and some media, digital and also information literacy approaches include aspects like responsible internet usage and an understanding of data collection (Gorman, 2015; Leaning, 2017; Mihailidis, 2018). Particularly relevant are the "critical digital literacy" models including a "more nuanced understanding of power and ideology within the digital medium" by Pangrazio (2016, p. 168), and by Pötzsch (2019, p. 221), arguing for the "widest possible contextualization of technology, including issues of exploitation, commodification, and degradation in digital capitalism." Third, the small and emerging field of *data literacy* research predominantly employs an active and creative understanding of such literacy (e.g., Carretero et al., 2017; D'Ignazio, 2017; Frank et al., 2016; Khan et al., 2018). However, some concepts include a critical reflection of big data practices either while or as a result of working with data in classroom or workshop environments (Crusoe, 2016; François and Monteiro, 2018; Hautea et al., 2017). Some very relevant approaches extend beyond formal education and call for critical data literacy for all citizens, for example, D'Ignazio and Bhargava (2015), Gray et al. (2018), and Pangrazio and Selwyn (2019).[1]

Building on this, it can be argued that today's citizens of datafied societies require an extended *critical big data literacy.* They need an understanding of *datafication*, recognizing the risks and benefits of the growing prevalence of data collection, analytics, automation, and predictive systems, as well as being able to critically reflect upon these developments. This includes, but goes beyond the skills of, for example, changing one's social media settings, and rather constitutes an altered view on the pervasive, structural, and systemic levels of changing big data systems in our datafied societies. Thus, this study suggests that being *critically data literate* means being *aware of* and able to *critically reflect* upon *big data collection practices, data uses and the possible risks and implications* that come with these practices, as well as being capable of *implementing this knowledge for a more empowered internet usage.* Importantly, being critically data literate is not understood as necessarily taking a "negative" stance to all big data practices, but rather involving the ability to weigh the evidence, make informed decisions, and scrutinize and respond to the socio-technical systems of big data practices.

This focused qualitative study investigated examples of existing efforts to promote such literacy—web resources that aim to foster people's awareness and critical reflection of the increasing datafication of our societies as well as their ability to protect their data online. The study investigated such *critical big data literacy tools* and their influence on internet users' attitudes about data collection and online privacy as well as on users' internet usage in the short and longer term.

As identified by Critical Data Studies scholars, there are many aspects of contemporary life affected by datafication. However, much research to date has involved investigating how people's attitudes to datafication are changing specifically in relation to attitudes and behavior toward data privacy. This study seeks to advance previous research in this area in particular. Despite being a highly ambiguous concept which is often measured in different ways, most results on people's privacy attitudes are consistent in showing the public's general concern for privacy (Marreiros et al., 2017, p. 2). Despite this identified concern, there is nevertheless widespread disclosure of personal data online. This inconsistency is often referred to as the "privacy paradox" (e.g., Kokolakis, 2017, p. 122). Previous scholars have explained this by referring to the idea of a rational cost–benefit analysis (e.g., Draper, 2017; Westin, 2003); people's poor knowledge (Turow et al., 2014; see also above); and notions of a resignation toward privacy (Dencik and Cable, 2017; Draper and Turow, 2019; Hargittai and Marwick, 2016). These findings informed this study's research design and instruments.

## Methodology

In order to identify existing examples for critical big data literacy tools, this study used a snowball sampling method starting with the interactive web-series "Do Not Track" (Upian, National Film Board of

[1] For a more extensive discussion of relevant literature, see also Sander (2020).

https://doi.org/10.1017/dap.2020.5 Published online by Cambridge University Press

Canada, ARTE, Bayrischer Rundfunk [BR], 2015). This tool was developed by a wide range of "public media broadcasters, journalists, developers, graphic designers, and independent media makers from different parts of the world" (ibid., 2015). In seven episodes, users learn about various issues around big data and are often encouraged to interact with the series. Furthermore, Do Not Track provides a large number of follow-up links and further information. Following these links and mentions, the original sample of exemplary tools was compiled. The nearly 40 tools identified originated from diverse actors of different nationalities and applied a variety of different design approaches, ranging from websites or short videos to online games, graphic novels and many other formats. By using visual mapping techniques and conducting a comparative analysis of the tools, a typology was developed (see also Sander, 2019).

Through theory-driven selection criteria based on media literacy evaluation or effectiveness research (Bergsma and Carney, 2008; Burrows et al., 2013; Hindmarsh et al., 2015; Hobbs, 2010), three tools most suitable for this study were selected. Apart from the interactive web-series "Do Not Track", the multimedia website "Me and My Shadow," produced by the Tactical Technology Collective (n.d.), and the short video "Reclaim Our Privacy" by the NGO La Quadrature du Net (2014) were chosen. While using different design approaches, all three tools give users a broad introduction into the topic with no prior knowledge required, promote a critical reflection of data systems and their implications, and provide users with constructive advice on how to protect their data online.

As a second step, the influence of these three tools on people's privacy attitudes and behavior was investigated through pre and post use testing. Given the complexity of people's privacy attitudes as found by prior research (see above), this aimed at an in-depth insight into people's concerns and actions and how these are affected by learning more about data collection and big data systems. Participants were selected through purposive sampling, aiming for students—excluding as many confounding factors as possible and assuming that most students use the internet regularly and are fairly familiar with digital technologies. Ideally, the participants would possess a certain digital literacy without having any special previous knowledge on big data practices. Therefore, certain courses of study were excluded and a question was added that tested previous knowledge in the field. Participants were recruited through social media and university emails. Ultimately, the sample of this qualitative multi-methods study consisted of 10 students: 5 undergraduate and 5 postgraduate students, studying 10 different courses in Cardiff. The 6 women and four men, most aged in their 20s, were predominantly British, with 1 Canadian and 1 Hungarian participant.

Testing took place before using the tools, 1 week after the "intervention" and finally 8 months later. In the 40-min "intervention," the participants used the 3 selected tools and were also invited to additionally navigate freely around further links and resources they found. This free design was implemented to cater for the needs of different learning types and allow for a natural browsing behavior. To be able to reconstruct the resources used, how much time spent with each, and how people's browsing behaviors differed, a screen recording tool was used. The participants' initial reactions on the tools during and after the intervention were also observed, which allowed for additional insights.

At each of the three points in time, questionnaires tested the participants' concern for privacy as well as various aspects of their internet usage—representing their "privacy behavior" (for a list of questions, see Appendix S1). The concern for privacy was measured through agreement (using a Likert scale) to 19 statements about privacy and data protection, which were conceptualized around different topics and designed building on established instruments that measure privacy attitudes and concerns, adapting these instruments according to this study's research question (Chellappa and Sin, 2005; Malhotra et al., 2004; Smith et al., 1996). Moreover, the questionnaires included open questions in order to invite participants to reflect on the tools used. Finally, the second questionnaire asked participants if they were interested in taking part in follow-up interviews at a later point in time. Five agreed and took part in the qualitative 1-h interviews that were conducted 8 months later. These aimed at a more in-depth understanding of people's concern for privacy, the changes made to their internet usage, their reflections on the tools, and a clarification of ambiguities that arose before. The interviews were structured around an interview guide, which was adapted to the individual participant based on the prior findings (see Appendix S1).

https://doi.org/10.1017/dap.2020.5 Published online by Cambridge University Press

## Influences on Privacy Attitudes and Behavior

This study revealed interesting examples of existing critical big data literacy tools, their design, and content approaches as well as findings related to the participants' reflections on these tools that are discussed in a forthcoming publication.[2] This paper discusses the core question: What influence do the tools studied have on people's privacy attitudes and behavior? As outlined above, critical big data literacy includes more than mere knowledge about big data and thus is difficult to measure. This study aimed to gain an understanding of how people's critical big data literacy is developed and may change through using the selected tools by investigating if and how participant's concern for privacy changed over time and also if and how their internet usage behaviors changed. Moreover, through qualitative interviews, this study gained insight into participant's awareness of privacy issues as well as their motivations and their concerns regarding big data. The findings, as related to privacy attitudes and behavior, demonstrate that after engaging with the data literacy tools, most participants showed an increase in concern for privacy and more privacy-sensitive internet usage. These findings suggest a relationship between tool usage and increased knowledge, understanding, and critical reflection of big data's influence as related to privacy. More research into the impact of data literacy tools on concerns about other aspects of big data, including rights, civic participation, and bias, is necessary.

### *The positive influence of critical big data literacy tools*

Overall, this study found a positive influence of the examined online tools in fostering critical big data literacy, which resulted in an increased concern about privacy and more privacy-sensitive internet usage of the participants. When examining the short-term impact of the tools 1 week after the intervention, first, an *increase in concern for privacy* could be measured for 8 of the 10 participants through stronger agreement with the 19 privacy statements given in the questionnaires[3]. Second, all but 2 participants used the open text fields to emphasize their *perceived increase* in privacy awareness and concern. Third, distinct changes in the participants' *internet usage* could be identified. Half of the participants checked their privacy settings, many started to use ad and tracking blockers, restricted the use of location services, increased their number of passwords, several cleared their cookies, and one even started to use the Tor Browser. Fourth, the majority of participants used a final open text field to emphasize their newly gained awareness and understanding of big data and privacy, the critical view they developed through the tools and to express their *appreciation of this new literacy*, for example: "Thanks for doing this as it has made me better-versed in online privacy and the way companies use my data, and I know I don't like it" (P02).

While this first part of the study also revealed some ambiguities, such as a decrease in measured concern for two participants, the overall finding of a positive influence was nevertheless unequivocal. When investigating the *longer-term influence* of the tools after 8 months, some findings clearly reaffirmed this positive influence, yet no distinct general trend could be identified. The five participants who took part in this in-depth follow-up study showed *diverse longer-term developments*, yet the majority indicated an increased concern about at least *some aspects of privacy* (see section "Complexity of privacy concerns") and all participants stated an increase of caution in at least *some situations of data disclosure online.* Moreover, several participants expressed a guilty conscience of not having done more.

However, despite the diversity of the findings, this follow-up study also provided strong evidence for a positive longer-term influence of the tools on people's critical big data literacy. Two of the five participants showed very distinct changes in their privacy concern and behavior, which not only remained, but even *intensified* in the months after using the tools. Both were very aware and concerned about the

[2] More detailed findings on these tools, their content and design approaches and the participants' suggestions for prospective tools are included in Sander (2020). What is critical big data literacy and how can it be implemented? *Internet Policy Review*, 9(2). DOI: 10.14763/2020.2.1479.

[3] While some participants showed only small increases, 5 out of the 10 participants scored substantially higher in the second questionnaire.

https://doi.org/10.1017/dap.2020.5 Published online by Cambridge University Press

issue of privacy and the collection of their personal data online and have taken extensive measures to restrict their data disclosure online after using the tools. Both further repeatedly emphasized the importance of more education on these issues and one even considered teaching data literacy in the future. Before taking part in the study, both participants had no special interest in or knowledge about online privacy but had a general curiosity about internet-related topics or a general drive to learn more. Importantly, both participants clearly linked the increase in their awareness and concern for privacy and the changes they made to their internet usage *directly to the tools* used in this study. P04, for example, explained that the tools "made me prioritize it [data protection]" and they are now "really trying to get my own checklist on what I need to do to protect myself." This constitutes a highly suggestive outcome about the positive influence data literacy efforts can have.

Also the other participants whose privacy concerns and behaviors were examined in depth showed interesting longer-term developments. While two participants expressed concern about *some aspects* of online privacy, the sample also included one case of "defaulting back" to their original attitude and behavior (P07). This participant had a very strong initial reaction to the tools, which they described as: "Oh my goodness, everyone knows everything about me, I need to delete myself" (P07). However, while this strong reaction led to a few initial small changes in their internet usage, this attitude faded with time and did not lead to a lasting change in privacy concern or behavior.

Thus, examining the influence of critical big data literacy tools on people's privacy attitudes and their internet usage led to evidence of an *only short-lived effect* of the tools for one participant, as well as a *persistent and even growing* increase in concern for privacy for others, expressed both in attitude and actions.

***Reasoning and motivations behind changes***

In the second stage of the study, some patterns emerged that hinted at the participants' reasonings when making changes to their internet usage. For example, they seemed hesitant to make changes to the e-mail providers and instant messengers they used, but made various other, lower-threshold changes. These patterns of *convenience* and a *network effect* could be clearly confirmed through the qualitative interviews in the third stage of the study. Many participants directly referred to these notions as a reasoning for their attitudes but also as strong motivators for making decisions on their internet usage. One example is P04, who elaborated: "So the search engine only affects me and I don't need to convince everyone I know to change to start using it and I think it's more effort to convince everyone to use a [certain instant messaging] app."

Overall, this analysis repeatedly found a *discrepancy* between the envisaged "ideal state" of privacy versus the more difficult implementation of this into reality. The notions of convenience and the network effect are examples of this discrepancy. While many participants of this study emphasized their concern for privacy, they also explained being drawn by the convenience of, for example, personalization mechanisms or felt unable to change their internet usage because their friends and family were not willing to. For example, P10 called this a "sacrifice" they made in order to stay in touch with loved ones. Moreover, both the two very concerned participants but even those who seemed unconcerned at first glance expressed a *guilty conscience* about their lacking concern or activism when asked further, for example P07, who explained "I know I'm wrong" and that this "should bother me more than it does."

Besides, new and unexpected notions could be identified. *Security concerns* constituted a strong justification for certain behaviors online and some participants expressed having *given up on protecting certain data* and focusing on restricting the collection of other data instead. Further, many still felt *unable to protect their privacy* online and called for easier means of data protection. Interestingly, these arguments correspond with this study's findings on the complexity of privacy concerns, which identified a prevalence of concern about *data security* as well as further evidence for a *resignation* toward privacy (see section "Complexity of privacy concerns"). Finally, what nearly all participants agreed on and strongly emphasized was that the problematic aspects of big data systems are currently *too removed from individuals* and that the protection of one's data online should be *made easier*: "I wish the way to opt out [of data collection] was easier" (P04).

https://doi.org/10.1017/dap.2020.5 Published online by Cambridge University Press

***Complexity of privacy concerns***

As already indicated above, this study confirmed the complexity of people's privacy concerns. Categorizing the participants' privacy attitudes into "concerned" or "not concerned" proved difficult as they were often only worried about certain aspects of online privacy, seemed unconcerned but then expressed a guilty conscience or a resignation toward their online privacy, or they explained being torn between concern and convenience. While an elaborate discussion of the often contradictory and fluctuating nature of these concerns is beyond the scope of this short paper, the following paragraph gives a first overview of these findings, which will be discussed in more detail in Sander (2020).

Overall, *nuanced themes* emerged both in relation to a concern about privacy but also with regard to a feeling of not "being bothered" by online privacy. Interestingly, many participants expressed concern but nevertheless felt *unable to imagine specific negative consequences* of data disclosure online. Also further *contradictions within* the individual participants' privacy attitudes became apparent. For example, some participants stated they felt at the same time concerned about their personal security and felt safer through big data, or they explained that it is easy to protect one's privacy online, yet also expressed that they sometimes feel unsure about how to protect their privacy. One participant spent the majority of the interview emphasizing the importance of privacy and the risks of big data, but admitted that even he sometimes cannot imagine the negative consequences data disclosure might have: "In a sense, I'm the same as the attitude that I described, that okay—even if they have all of this data, what are they going to do with it?" (P10).

Apart from this, *distinct differences between* the participants' understandings of privacy could be identified. The findings suggest the notion of *different dimensions of online privacy*, with a key finding of a prevalent concern about the primarily technological issue of *data security.* While participants seemed very aware and concerned about the security of their data and some also about their *personal security* and the *social implications* of data-driven technologies, they seemed far less concerned about the *impact* data disclosure might have, such as tracking, scoring, or surveillance. Finally, the notion of *resignation* was examined. The feeling of having "given up" on one's data was discussed by several participants and P10 particularly elaborated on the *origin* of the feeling as well as suggesting possible *solutions* to prevent resignation in the future (see also Sander, 2020).

## Engaging Citizens and Promoting Empowered Internet Usage

Overall, this paper problematizes citizens' lacking knowledge and understanding of big data systems and argues that the researched tools could provide relevant means to promote critical big data literacy. While not without ambiguities, the study's findings suggested a predominantly strong initial reaction to the tools, which lead to more concern for privacy and more privacy-sensitive internet usage for most participants. When analyzing the longer-term influence of the tools, results were more mixed with evidence of both a short-lived effect of the tools for some and also of a substantial longer-term impact for others. The differences between participants could be attributed to individual characteristics such as a general curiosity or interest in the area. However, these first findings are highly suggestive and more research is required in order to better understand factors that determine the longer-term success of data literacy efforts.

Moreover, this study's findings confirmed the theory that people might know about the collection of their data online but that they have little awareness and understanding of the implications the disclosure of their data might have. Together with the many calls for more education on big data systems both from researchers and from this study's participants, this reemphasizes the importance of critical big data literacy and the necessity for more research in this field. One key message of this study is that this literacy may not only empower internet users and constitute an important asset of contemporary citizenship, but is also highly appreciated by the participants themselves. Even though the study suggested that privacy concerns are incredibly complex and often restricted to certain aspects of online privacy, the participants were predominantly concerned when learning more about the use of their data. At least in the context of this very small sample, this outcome debunks claims of the indifferent internet user who does not care about privacy and feels like they have "nothing to hide."

https://doi.org/10.1017/dap.2020.5 Published online by Cambridge University Press

Thus, the findings advance research on people's knowledge of data collection and analysis practices, but also on attitudes toward privacy. The complexity of privacy concerns was repeatedly confirmed and the different understandings and aspects of privacy, such as a prevalence of data security issues or a discrepancy between people's envisaged "ideal state" of privacy versus the more difficult implementation of this into reality need to be considered and better understood through future research.

Finally, this research provided valuable in-depth findings on people's data literacy and a first insight into the influence of critical big data literacy tools on people's privacy concern and behavior. With the increasing datafication of the public sector, it is essential to address public concerns, foster digital trust and enable citizens to protect their data online. Only if citizens understand basic aspects of the technology, analysis, and mythology around big data, they are able to recognize the increasing datafication of our societies, assess these new developments and make informed decisions on their internet usage and about navigating their lives in a datafied society. Despite its small scale and some ambiguities in the findings, the research suggests that these not previously researched critical big data literacy tools could provide a substantial means to work toward empowering internet users, promoting a critical internet usage and, ideally, enabling more citizens to engage in public debates about changing data systems.

More theoretical and empirical research on the concept of critical big data literacy, the implementation of such literacy into practice and on existing tools is required. This study makes novel and in-depth contributions to this field, informs future work and provides various leverage points to address this gap in research. It further argues that more education on big data systems is necessary and the protection of personal data online needs to be made easier. Thus, besides more research, also more resources to teach critical big data literacy are required. Forthcoming publications based on this study will present first suggestions on 'what works' in terms of such tools' design and content strategies.

**Acknowledgments.** I would like to thank Dr Joanna Redden, Cardiff University, for her continuous support and her greatly helpful advice and comments. I would also like to show my gratitude to the *Center for Advanced Internet Studies* (CAIS) for their generous support and a pleasant and productive working atmosphere.

**Funding Statement.** This work was partially supported by a research fellowship at the *Center for Advanced Internet Studies* (CAIS) in Bochum, Germany. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests.** The author declares no competing interests exist.

**Data Availability Statement.** Data can be made available to interested researchers upon request by email to the corresponding author.

**Author Contributions.** Conceptualization, I.S.; Data Curation, I.S.; Formal Analysis, I.S.; Funding Acquisition, I.S.; Investigation, I.S.; Methodology, I.S.; Project Administration, I.S.; Resources, I.S.; Validation, I.S.; Writing—original draft, I.S.; Writing—review and editing, I.S.

**Supplementary Materials.** To view supplementary material for this article, please visit http://dx.doi.org/10.1017/dap.2020.5.

## References

**Bergsma LJ and Carney ME** (2008) Effectiveness of health-promoting media literacy education: a systematic review. *Health Education Research 23*(3), 522–542.

**boyd d and Crawford K** (2012). Critical questions for big data: provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society 15*(5), 662–679.

**Bucher T** (2017) The algorithmic imaginary: exploring the ordinary affects of Facebook algorithms. *Information, Communication & Society 20*(1), 30–44.

**Burrows R**, **Johnson H and Johnson P** (2013) *Influencing Values, Attitudes and Behaviour via Interactive and Social-media technology: The Case of Energy Usage (Technical Report)*. Bath, UK: Department of Computer Science, University of Bath.

**Carretero S**, **Vuorikari R and Punie Y** (2017) *DigComp 2.1: The Digital Competence Framework for Citizens with eight proficiency levels and examples of use*. Luxembourg: Publications Office of the European Union.

**Chellappa RK and Sin RG** (2005). Personalization versus privacy: An empirical examination of the online consumer's dilemma. *Information Technology and Management 6*(2), 181–202.

**Crusoe D** (2016) Data Literacy defined pro populo: to read this article, please provide a little information. *The Journal of Community Informatics 12*(3), 27–46.

**D'Ignazio C** (2017) Creative data literacy: bridging the gap between the data-haves and data-have nots. *Information Design Journal 23*(1), 6–18.

**D'Ignazio C and Bhargava R** (2015) *Approaches to Building Big Data Literacy.* New York, NY. Available at http://rahul-beta.connectionlab.org/wp-content/uploads/2011/11/ Edu_DIgnazio_52.pdf.

**Dalton C and Thatcher J** (2014) What does a critical data studies look like, and why do we care? *Society & Space*. Available at http://societyandspace.org/2014/05/12/what-does-a-critical-data-studies-look-likeand-why-do-we-care-craig-dalton-and-jim-thatcher/.

**Dencik L and Cable J** (2017) The advent of surveillance realism: public opinion and activist responses to the Snowden leaks. *International Journal of Communication 11*, 763–781.

**Doteveryone** (2018). *People, Power and Technology: The 2018 Digital Attitudes Report.* London: Doteveryone. Available at https://attitudes.doteveryone.org.uk/files/People%20Power%20and%20Technology%20Doteveryone%20Digital%20Attitudes%20Report%202018.pdf.

**Draper NA** (2017) From privacy pragmatist to privacy resigned: challenging narratives of rational choice in digital privacy debates. *Policy & Internet 9*(2), 232–251.

**Draper NA and Turow J** (2019) The corporate cultivation of digital resignation. *New Media & Society 21*(8), 1824–1839.

**Eubanks V** (2018) *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York, NY: St. Martin's Press.

**François K and Monteiro C** (2018) Big data literacy. In Sorto MA, White A and Guyot L (eds), *Looking Back, Looking Forward. Proceedings of the Tenth International Conference on Teaching Statistics, Kyoto, Japan*. Voorburg, The Netherlands: International Statistical Institute.

**Frank M**, **Walker J**, **Attard J and Tygel A** (2016) Data literacy—what is it and how can we make it happen? *The Journal of Community Informatics 12*(3), 4–8.

**Garcia A**, **Mirra N**, **Morrell E**, **Martinez A and Scorza D** (2015) The Council of Youth Research: critical literacy and civic agency in the digital age. *Reading & Writing Quarterly 31*(2), 151–167.

**Gorman GE** (2015) What's missing in the digital world? Access, digital literacy and digital citizenship. *Online Information Review 39*(2). Available at: http://www.emeraldinsight.com/doi/10.1108/OIR-02-2015-0053.

**Gray J**, **Gerlitz C and Bounegru L** (2018) Data infrastructure literacy. *Big Data & Society 5*(2), 1–13.

**Grzymek V and Puntschuh M** (2019) *Was Europa Über Algorithmen Weiß und Denkt. Ergebnisse Einer Repräsentativen Bevölkerungsumfrage*. Barcelona, Spain: Bertelsmann Stiftung.

**Hammer R** (2011) Critical media literacy as engaged pedagogy. *E-Learning and Digital Media 8*(4), 357–363.

**Hargittai E and Marwick A** (2016) "What Can I Really Do?" Explaining the privacy paradox with online apathy. *International Journal of Communication 10*, 3737–3757.

**Hautea S**, **Dasgupta S and Hill BM** (2017) *Youth Perspectives on Critical Data Literacies*. Denver, CO: ACM Press, pp. 919–930. Available at http://dl.acm.org/citation.cfm?doid=3025453.3025823.

**Hindmarsh CS**, **Jones SC and Kervin L** (2015) Effectiveness of alcohol media literacy programmes: a systematic literature review. *Health Education Research 30*(3), 449–465.

**Hinrichsen J and Coombs A** (2013) The five resources of critical digital literacy: a framework for curriculum integration. *Research in Learning Technology 21*, 1–16. Available at https://journal.alt.ac.uk/index.php/rlt/article/view/1433.

**Hintz A**, **Dencik L and Wahl-Jorgensen K** (2018) *Digital Citizenship in a Datafied Society.* Medford, OR: Polity Press.

**Hobbs R** (2010) *Digital and Media Literacy: A Plan of Action*. Washington, DC: The Aspen Institute. Available at https://www.aspeninstitute.org/events/digital-media-literacy-plan-action/.

**Khan HR**, **Kim J and Chang H-C** (2018) Toward an understanding of data literacy. In *iConference 2018 Proceedings.* Sheffield, England. Available at https://www.ideals.illinois.edu/bitstream/handle/ 2142/100243/Khan-Hammad_20180417_V01.pdf?sequence=1.

**Kitchin R** (2017) Thinking critically about and researching algorithms. *Information, Communication & Society 20*(1), 14–29.

**Kitchin R and Lauriault TP** (2014) *Towards Critical Data Studies: Charting and Unpacking Data Assemblages and Their Work. The Programmable City Working Paper 2*. Lincoln, NE: University of Nebraska Press. Available at https://ssrn.com/abstract=2474112.

**Kokolakis S** (2017) Privacy attitudes and privacy behaviour: a review of current research on the privacy paradox phenomenon. *Computers & Security 64*, 122–134.

**La Quadrature du Net** (2014) *Reclaim Our Privacy.* Available at https://www.youtube.com/watch?v=TnDd5JmNFXE (accessed 10 May 2019).

**Leaning M** (2017) *Media and Information Literacy. An Integrated Approach for the 21st Century.* Cambridge, MA: Chandos Publishing is an imprint of Elsevier. Available at https://linkinghub.elsevier.com/retrieve/pii/B978008100170700007X.

**Malhotra NK**, **Kim SS and Agarwal J** (2004) Internet users' information privacy concerns (IUIPC): the construct, the scale, and a causal model. *Information Systems Research 15*(4), 336–355.

**Marreiros H**, **Tonin M**, **Vlassopoulos M and Schraefel MC** (2017) "Now that you mention it": a survey experiment on information, inattention and online privacy. *Journal of Economic Behavior & Organization 140*, 1–17.

https://doi.org/10.1017/dap.2020.5 Published online by Cambridge University Press

**Mayer-Schönberger V and Cukier K** (2013) *Big data: A Revolution that will Transform How We Live, Work and Think*. London, UK: Murray.

**Mihailidis P** (2018) Civic media literacies: re-imagining engagement for civic intentionality. *Learning, Media and Technology 43*(2), 1–13. 10.1080/17439884.2018.1428623.

**Miller C**, **Coldicutt R and Kitcher H** (2018) *People, Power and Technology: The 2018 Digital Understanding Report.* London, UK: Doteveryone. Available at http://understanding.doteveryone.org.uk/.

**Müller-Peters H** (2020) Big data: Chancen und Risiken aus Sicht der Bürger. In Knorre S, Müller-Peters H and Wagner F (eds), *Die Big-Data-Debatte. Chancen und Risiken der Digital Vernetzten Gesellschaft*. Berlin, Germany: Springer, pp. 137–193.

**O'Neil C** (2016) *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy.* London, UK: Allen Lane, Penguin Books.

**Pangrazio L** (2016) Reconceptualising critical digital literacy. *Discourse: Studies in the Cultural Politics of Education 37*(2), 163–174.

**Pangrazio L and Selwyn N** (2019) 'Personal data literacies': a critical literacies approach to enhancing understandings of personal digital data. *New Media & Society 21*(2), 419–437. https://doi.org/10.1177/1461444818799523.

**Pötzsch H** (2019) Critical digital literacy: technology in education beyond issues of user competence and labour-market qualifications. *tripleC: Communication, Capitalism & Critique. Open Access Journal for a Global Sustainable Information Society 17*(2), 221–240. https://doi.org/10.31269/triplec.v17i2.1093.

**Redden J** (2018) Democratic governance in an age of datafication: lessons from mapping government discourses and practices. *Big Data & Society 5*(2), 1–13.

**Redden J and Brand J** (2017) *Data Harm Record*. Cardiff, UK: Data Justice Lab, Cardiff University. Available at https://datajusticelab.org/data-harm-record/.

**Sander I** (2019) *A Critically Commented Guide to Data Literacy Tools.* http://doi.org/10.5281/zenodo.3241421 (accessed 5 July 2019).

**Sander I** (2020) What is critical big data literacy and how can it be implemented? *Internet Policy Review* 9(2). DOI: 10.14763/2020.2.1479.

**Smith HJ**, **Milberg SJ and Burke SJ** (1996) Information privacy: measuring individuals' concerns about organizational practices. *MIS Quarterly 20*(2), 167–196.

**Tactical Technology Collective** (n.d.) *Me and My Shadow.* Available at https://myshadow.org (accessed 10 May 2019).

**Turow J**, **Bleakley A**, **Bracken J**, **Carpini MXD**, **Draper NA**, **Feldman L**, **Good N**, **Grossklags J**, **Hennessy M**, **Hoofnagle CJ**, **Howard-Williams R**, **King J**, **Li S**, **Meltzer K**, **Mulligan DK**, **Nir L** (2014). *Americans, Marketers, And The Internet: 1999–2012*. Philadelphia, PA: The Annenberg School for Communication, University of Pennsylvania. Available at http://ssrn.com/abstract=2423753.

**Turow J**, **Hennessy M and Draper N** (2018) Persistent misperceptions: Americans' misplaced confidence in privacy policies, 2003–2015. *Journal of Broadcasting & Electronic Media 62*(3), 461–478.

**Turow J**, **Hennessy M and Draper NA** (2015) *The Tradeoff Fallacy: How Marketers are Misrepresenting American Consumers and Opening them up to Exploitation*. Philadelphia, PA: The Annenberg School for Communication, University of Pennsylvania. Available at https://www.asc.upenn.edu/sites/default/files/TradeoffFallacy_1.pdf.

**Upian, National Film Board of Canada, ARTE, Bayrischer Rundfunk [BR]** (2015). *Do Not Track.* Available at https://donottrack-doc.com/en/ (accessed 10 May 2019).

**Westin AF** (2003) Social and political dimensions of privacy. *Journal of Social Issues 59*(2), 431–453.

**Zuboff S** (2015) Big other: surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology 30*(1), 75–89.

**Cite this article:** Sander, I. 2020. Critical big data literacy tools—Engaging citizens and promoting empowered internet usage. *Data & Policy*, 2: e5. doi:10.1017/dap.2020.5