What is critical big data literacy and how can it be implemented?

Ina Sander

Data Justice Lab, Cardiff University, United Kingdom

Abstract: This paper argues that data literacy today needs to go beyond the mere skills to use data. Instead, it suggests the concept of an extended critical big data literacy that places awareness and critical reflection of big data systems at its centre. The presented research findings give first insights into a wide variety of examples of online resources that foster such literacy. A qualitative multi-methods study with three points in time further investigated the views of citizens about the effects of these tools. Key findings are a positive effect particularly of interactive and accessible data literacy tools with appealing visualisations and constructive advice as well as highly insightful suggestions for future tools.

This paper is part of Digital inclusion and data literacy, a special issue of Internet Policy Review *guest-edited by Elinor Carmi and Simeon J. Yates.*

INTRODUCTION

With the increasing ubiquity of big data systems, awareness of citizens' need for 'digital skills' and 'data literacy' has been growing among scholars, activists and political decision-makers. What type of skills and knowledge do people need to 'be digital', to use the internet and related technologies in an informed manner? While media, digital and data literacy approaches have been refined and expanded to accommodate for the changing landscape of digital technologies

(see below), one aspect is only recently beginning to attract more attention: people's awareness and understanding of big data practices (e.g., report on 'Digital Understanding' by Miller, Coldicutt, & Kitcher, 2018). Even though various skills to use digital media, data sets and the internet for specific purposes are without doubt important, citizens in today's datafied societies need more than skills. They need to be able to *understand and critically reflect upon* the ubiquitous collection of their personal data and the possible risks and implications that come with these big data practices (e.g., Pangrazio & Selwyn, 2019), challenging common 'myths' around big data's objective nature (boyd & Crawford, 2012). This is essential to foster an informed citizenry in times of increasing profiling and social sorting of citizens and other profound economic, political and social implications of big data systems. These systems come with manifold risks, such as threatening individual privacy, increasing and transforming surveillance, fostering existing inequalities and reinforcing discrimination (e.g., Eubanks, 2018; O'Neil, 2016; Redden & Brand, 2017).

At the moment, however, people's knowledge and understanding of these issues seem fragmented. Recent research has shown Europeans' lacking knowledge about algorithms (Grzymek & Puntschuh, 2019); Germans' knowledge deficits with regards to digital interconnectivity and big data (Müller-Peters, 2019); and Americans' misunderstandings and misplaced confidence in privacy policies (Turow, Hennessy, & Draper, 2018). The British organisation Doteveryone identified a "major understanding gap around technologies", finding that a only a third of people know that data they have not actively chosen to share has been collected (Doteveryone, 2018, p. 5), and that less than half of British adult internet users are aware that apps collect their location and information on their personal preferences (Ofcom, 2019, p. 14).

Moreover, research has found that even when people are aware of the collection of their personal data online, they often only have a vague idea of the general system of big data practices and how this might impact their lives and our societies (Turow, Hennessy, & Draper, 2015, p. 20). Nevertheless, many are uncomfortable about the collection and use of their personal data (e.g., Bucher, 2017; Miller et al., 2018; Ofcom, 2019) and they find big data practices less acceptable the more they learn about them (Worledge & Bamford, 2019). Alongside many other scholars, I argue that it is essential that citizens in datafied societies learn more about and begin to understand big data practices in order to enable them "to form considered opinions and debate the issue in a factually informed way" (Grzymek & Puntschuh, 2019, p. 11). While aspects such as general media literacy, fake news and disinformation are slowly reaching more attention (e.g., the European Commission's 'European Media Literacy Week' or UNESCO's 'Global Alliance for Partnerships on Media and Information Literacy'), there is still lacking engagement with the impacts of big data systems. People's education about big data needs to be strengthened, with more "investment in new forms of public engagement and education" (Doteveryone, 2018, p. 6) that also "address structural features of media systems" (Turow et al., 2018, p. 461), such as the structural and systemic levels of big data practices. This study provides insight into ways to achieve these goals by: a) conceptualising critical big data literacy and b) drawing on empirical research to discuss the effects of data literacy tools.

After presenting the concept of critical big data literacy and its theoretical grounding, the paper details existing examples of initiatives and resources that aim to foster such literacy. An analysis of 40 data literacy tools is used to present a typology of their content and design approaches. A small selection of these tools is then evaluated by examining users' perspective on them: their perception of the tools' short and mid-term effects, reflections on their suitability to teach data

literacy and feedback on their content and design. Finally, based on the participants' feedback and ideas and complemented by further research findings, suggestions for the creation of future data literacy efforts are made.

1. WHAT IS CRITICAL BIG DATA LITERACY?

I argue here for *critical big data literacy*, that is literacy efforts that aim to go beyond the skills of using data. While learning to use digital media and data productively without doubt constitute relevant skills for today's digital citizens, only a small body of research around digital or data literacy concepts focuses on the many potentially problematic issues related to *big data*. As often highlighted before, big data constitute a "socio-technical phenomenon" that entails the "capacity to search, aggregate, and cross-reference large data sets" (boyd & Crawford, 2012, p. 663), to, among others, search for patterns and create categories, profiles or scores often used for decision-making and predictive analyses. The increasing use of these systems in areas such as banking, employment, policing, and social services comes with profound economic, political, and, importantly, social implications (Eubanks, 2018; O'Neil, 2016; Redden & Brand, 2017).

In today's datafied societies, citizens need to be aware of these systems that affect so many areas of their lives, and a critical public debate about data practices is essential. Therefore, I argue that data literacy today needs to go beyond data skills in order to foster such awareness and public debate of the datafication of our societies. Citizens need to recognise risks and benefits related to the increasing implementation of data collection, analytics, automation, and predictive systems and need to be able to scrutinise the structural and systemic levels of these changing big data systems. Thus, I suggest that *critical big data literacy* in practice should mean an *awareness, understanding* and ability to *critically reflect upon big data collection practices, data uses and the possible risks and implications* that come with these practices, as well as the *ability to implement* this knowledge for a *more empowered internet usage* (see also Sander, 2020).

The concept of critical big data literacy combines approaches from a variety of research fields. First, *critical data studies* not only aim to understand and critically examine the "importance of data" and how they have become "central to how knowledge is produced, business conducted, and governance enacted" (Kitchin & Lauriault, 2014, p. 2), but many critical data scholars also call for more societal and public involvement (Marwick & Hargittai, 2018, p. 14; O'Neil, 2016, p. 210; Zuboff, 2015, p. 86). They hope that more "reflexive, active and knowing publics" (Kennedy & Moss, 2015, p. 1) might not only empower citizens, but also "open up discussion of policy solutions to regulate such [big data] practices" (Marwick & Hargittai, 2018, p. 14). However, concrete suggestions on how to implement such a transfer of academic knowledge are rare.

Secondly, the concept of critical big data literacy builds on a long history of *critical media literacy* and *critical digital literacy* research. While literacy was long understood to include mainly the skills to use (digital) media productively, this understanding has been criticised as too uncritical in recent years. Scholars have questioned this "technocratic or functional perspective" and have called for a critical perspective that includes analysis and judgement of the "content, usage and artefacts" of digital technology as well as the "development, effects and social relations bound in technology" (Hinrichsen & Coombs, 2013, p. 2, p. 4). Thus, an increased emphasis on critical approaches has emerged (e.g., Hammer, 2011; Garcia et al., 2015; Pangrazio, 2016). Yet, only few literacy initiatives and concepts include critiquing the "platform or modality relationships to information and communication" (Mihailidis, 2018, p. 4), and aspects such as the impacts of datafication or an understanding of risks around privacy and

surveillance are often omitted. However, there are some particularly relevant concepts of critical digital literacy that "take structural aspects of the technology into account", considering "issues of exploitation, commodification, and degradation in digital capitalism" (Pötzsch, 2019, p. 221) and aiming for a "more nuanced understanding of power and ideology within the digital medium" (Pangrazio, 2016, p. 168).

Thirdly, many *data literacy* approaches and concepts have developed in recent years. While a large part of these studies understand data literacy in an active and creative way and aim to teach citizens to read, use and work with data – often in an empowering way (e.g., D'Ignazio, 2017), fostering people's "data mindset" (D'Ignazio & Bhargava, 2018) – there are a number of concepts that include a critical reflection of (big) data systems either while or as a result of working with data. For some, this critical approach falls under the term "data literacy" (e.g., Crusoe, 2016), others use concepts such as "big data literacy" (D'Ignazio & Bhargava, 2015; Philip, Schuler-Brown, & Way, 2013), "digital understanding" (Miller et al., 2018), or "algorithmic literacy" (Grzymek & Puntschuh, 2019). Particularly relevant approaches are "data infrastructure literacy" that aims to promote "critical inquiry into datafication" (Gray, Gerlitz, & Bounegru, 2018, p. 3) and an "extended definition of Big Data literacy" by D'Ignazio and Bhargava, aiming to educate about "when and where data is being passively collected", "algorithmic manipulation" and the "ethical impacts" of "data-driven decisions for individuals and for society" (2015, p. 1, p. 3).

These concepts focus on fostering a critical view through *active usage* of data tools in *classroom and workshop environments*. Thus, while they add highly relevant insights into conceptualising critical big data literacy, they lack focus on a *broader audience* as well as critical inquiry of the *structural level* of big data systems. However, some very relevant approaches go beyond formal education and argue for a broader conceptualisation of critical data literacy. For example, Fotopoulou conceptualises data literacy for civil society organisations and argues for "data literacies as agentic, contextual, critical, multiple, and inherently social", raising "awareness about the ideological, ethical and power aspects of data" (2020, p. 2f). Similarly, Pangrazio and Selwyn call for "personal data literacies" that "include conceptualisations of the inherently political nature of the broader data assemblage" and aim to build "awareness of the social, political, economic and cultural implications of data" (2019, p. 426), while Pybus, Coté and Blanke work towards a "holistic approach to data literacy", including an understanding of meaning making through data, of big data's opaque processes and an "active (re)shaping of data infrastructures" (2015, p. 4). A recent report by the 'Me and My Big Data'-Project further suggests "data citizenship" as a new data literacy framework that combines skills with critical understanding. This framework consists of "Data thinking" – critically understanding the world through data; "Data doing" – learning practical skills around everyday engagements with data; and "Data participation" – aiming to "examine the collective and interconnected nature of data society" (Yates et al., 2020, p. 10).

Critical big data literacy builds on such conceptualisations as well as contributes to research from all of these fields: It aims to communicate critical data studies' findings to the public; to learn from critical media and digital literacy approaches; and to build on and advance data literacy concepts by working to foster citizens' critical understanding of datafication. Following Mulnix's considerations on critical thinking, critical big data literacy aims at the "development of autonomy", or the "ability to decide for ourselves what we believe", through our own critical deliberations on the use of our data and the impact of datafied systems (2012, p. 473). Importantly, just as critical thinking is "not directed at any specific moral ends" and does not intrinsically contain a certain set of beliefs as its natural outcomes (ibid., p. 466), so is being

critically data literate not understood as necessarily taking a 'negative' stance to all data practices.

The aim of critical big data literacy is not to affect internet users in a way that leaves them feeling negatively about all data collection and analysis or even resigned about such big data practices (for resignation, see below and, e.g., Turow et al., 2015). Rather, the goal is to foster users' awareness and understanding about what happens to their data and thus to enable them to question and scrutinise the socio-technical systems of big data practices, to weigh the evidence, to build informed opinions on current debates around data analytics as well as to allow them to make informed decisions on personal choices such as which data to share or which services to use. While it is important to not merely shift responsibility to individuals as citizens' agency in this field and possibilities to opt out of these systems are limited, it is nevertheless crucial that citizens are able to learn more about these debates and how to use the internet in a more critical and empowered way, for example through using alternative services.

2. HOW TO RESEARCH CRITICAL BIG DATA LITERACY

This study investigates online resources that foster critical big data literacy and focuses on two perspectives: to gain insight into existing examples of such online resources and to understand how they affect those who use them as well as how these users perceive these resources.

To investigate the first research question – which and what kind of examples for online critical big data literacy tools already exist? – I conducted a snowball sampling of available tools, using the interactive web-series 'Do Not Track' (2015) as a starting point. 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" ('About Do Not Track - Who We Are', 2015). 'Do Not Track' is often being used in teaching and offers a great amount of follow-up links for every episode, including further information and articles, but also other resources that inform or teach about big data, constituting the ideal starting point for this snowball sampling. In snowball sampling, a group of respondents – or, in this case, online data literacy resources – is initially selected because of their relevance to the research objectives (Hansen & Machin, 2013, p.217). These are then asked to identify other potential respondents "of similar relevance to the research objectives" (ibid.). In this case, this meant following the great amount of links and mentions of other resources on the 'Do Not Track' website as well as further links and mentions on any website identified in this way.

All English-language online resources that aim to educate about big data and data collection and that were identified through this snowball sampling were included in the initial sample. Thus, when I talk about critical big data literacy tools, the term 'tool' is used in a very open way and the sampling included various different kinds of informational and educational resources on issues related to big data (see below). However, individual news articles on the topic, sites that cover only one aspect (e.g., cookie tracking), data visualisation tools, and resources that do not include any constructive suggestions for ways to protect one's data online were excluded. It is further important to emphasise that, as there exists very little research on these tools as yet, this study did not aim at a comprehensive, but rather an adequate overview and a first insight into the field.

To deepen this insight, I conducted a comparative analysis of the identified tools, focusing on the tools' formats and their production origins. Based on this, I developed a typology reflecting

the different content and design approaches that tools in the sample apply. This typology further provided an initial basis on which to select tools to test in the second part of the study. The three selected tools should: (1) apply different design approaches according to this typology; (2) not only outline critical aspects of big data, but also provide easy-to-follow constructive suggestions for ways to protect one's data online; (3) be oriented towards findings of media literacy evaluation or effectiveness research.

The second part of this study tested the three selected tools, researching the questions: how do critical big data literacy tools change people's privacy attitudes and behaviour short and mid-term? How do people perceive these tools? In this qualitative multi-methods approach, critical big data literacy was operationalised by investigating changes in people's concern for privacy and in their privacy-sensitive internet usage in three points in time: before using the tools, one week after this 'intervention' and finally eight months later. While concern for privacy and privacy-sensitive internet usage are likely insufficient to measure *all* aspects of the complex concept of critical big data literacy as outlined above, an increase in these two areas certainly indicates a generally increased awareness and critical reflection of issues around big data and online privacy as well as the ability to implement these concerns into one's daily internet usage. For example, critically data literate citizens should be aware of the way Google and WhatsApp, among others, collect, use and sell their data, what it could be used for, and they should know about alternative and privacy-sensitive search engines and messaging apps (e.g., DuckDuckGo, Telegram or Signal) to be able to make an informed decision on which service to use in the future.

The study's sample was selected in consideration of 'data adequacy' (Vasileiou et al., 2018). Both the sample size and the participants' demographic were aimed to be *adequate* for this study's design and research objectives. As many studies with larger sample sizes have indicated the complexity of people's attitudes towards data usage (see above), the aim of this study was to complement such quantitative studies by conducting a small and focused qualitative analysis that provides in-depth information for each participant. Thus, considering the length of the study and the depth of the analyses, a small sample size was adequate for this design. As each participant contributed to the study at several points in time and the study's instruments included various open questions and investigated different perspectives on data literacy, each participant's responses and reflections consisted of very information-rich data. Therefore, a small sample size of ten participants was adequate for this multi-methods study.

Also in terms of the demographics of the sample, data adequacy was considered. Here, two aspects played a key role. First, as the goal of this study was to identify if data literacy tools would lead to a change in *online* privacy attitude and *online* behaviours, and the interactive tools employed in the study required a certain level of digital ability, it was decided to recruit participants with some level of digital skills. Second, given the small sample size and the study's objective to understand individual privacy concerns, potentially intricate reasonings for online behaviour and individual users' perceptions of data literacy tools in-depth, the sample aimed for as few variables in the sample composition as possible. In light of these two aspects, focussing on a narrow demographic with a certain level of digital skills was adequate for this study. Therefore, university students constituted an ideal group of participants for this study as many of them already possess the basic level of digital ability required to engage with the tools and recruiting them was resource-friendly through formal and informal university networks. However, this narrow sample also came with some limitations, which will be discussed in the conclusion.

The study's participants were identified through purposive sampling, a strategy where particularly "information-rich cases" are selected (Vasileiou et al., 2018, p. 2). In this case, the purposive sampling aimed for Cardiff students with an average 'digital literacy' but no previous knowledge on big data. Therefore, certain courses of study were excluded from the sampling and a question that tested previous knowledge in the field was added in the first questionnaire. There was also an effort to balance gender, course and degree type. The final sample consisted of five undergraduate and five postgraduate students who each studied a different university course in Cardiff. The six women and four men, most aged in their twenties, were predominantly of British nationality, with one Canadian and one Hungarian participant.
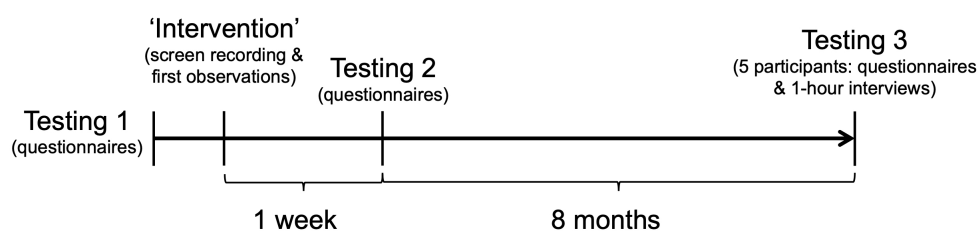

Figure 1: The three stages of research.

At each of the three points in time when testing took place, the participants completed questionnaires with open and closed questions, which examined their concern for privacy and various aspects of their internet usage. Furthermore, after the intervention, open questions asked for a reflection of the tools applied in the study. The questionnaires were designed building on established instruments that measure privacy attitudes and concerns, adapting them to this study's research question (Chellappa & Sin, 2005; Malhotra, Kim, & Agarwal, 2004; Smith, Milberg, & Burke, 1996). In the 40 minutes intervention, the participants were invited to use the three tools and also navigate freely around further links and resources they found. This took place individually and was not closely supervised, aiming at a 'natural browsing behaviour' and catering for the needs of different learning styles (see for example Pritchard, 2008, p. 41ff). The intervention was examined by noting particularly striking initial reactions of the participants to the tools (e.g., exclamations of surprise about certain methods of tracking online) as well as using a screen recording tool, both took place transparently with the participants' knowledge and consent. Even though the screen recording may have restricted the natural browsing situation described above because some participants may have felt observed, it also allowed for interesting analyses of which tools were used, how much time was spent with each, and how people's browsing behaviours differed that would not have been possible otherwise.

Finally, qualitative one-hour interviews were conducted with five of the participants eight months after they had used the tools. These aimed at gaining a more in-depth understanding of the participants' critical big data literacy and patterns found by prior research; the potential effect the tools had on their privacy concern and behaviour; and the participants' perceptions of and their reflections on the three tools. Using a structured interview guide, which was partly adapted to each participant based on their prior findings, I was further able to clarify minor ambiguities that arose before and inquire about intentions participants had expressed earlier.

3. HOW COULD CRITICAL BIG DATA LITERACY BE IMPLEMENTED?

3.1 EXAMPLES FOR DATA LITERACY TOOLS

Aiming at an overview of examples of existing data literacy tools, the main finding of the first part of this study was the *large amount and variety* of tools identified. The snowball sampling process proved effective as many organisations and websites seem to be highly connected and interrelated. In total, nearly 40 examples of data literacy tools could be identified. A comparative analysis gave further insight into how critical big data literacy is being implemented in each tool's design, revealing a variety of different approaches.
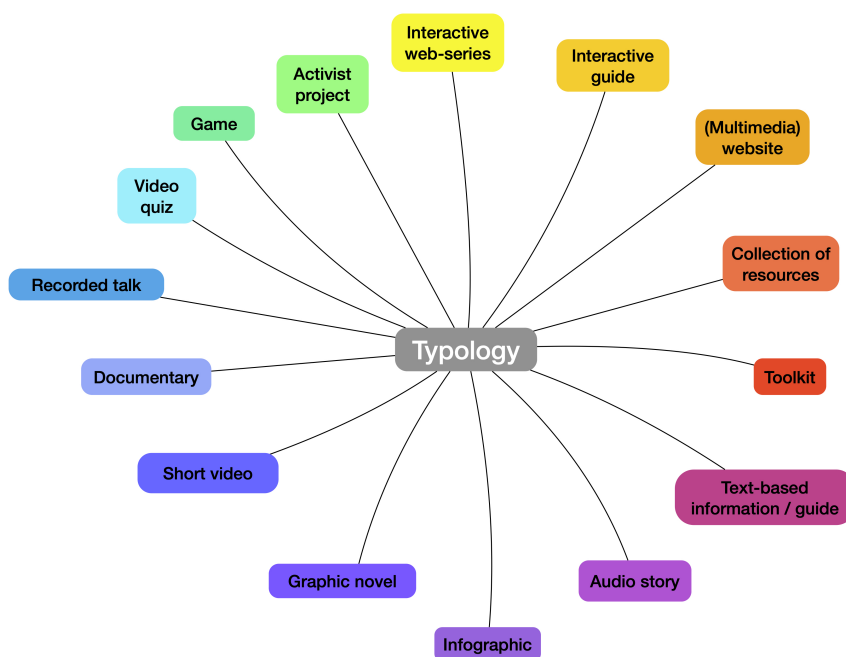

Figure 2: Typology with categories.

Firstly, the analysis demonstrated the sample's diversity in the tools' formats. Fifteen distinct categories of content and design approaches could be identified (see figure 2), which included approaches that were to be expected, such as '(multimedia) websites', 'short videos', or 'text-based information', but also some emerging findings, such as a graphic novel, a game, and an audio story used to communicate a critical view on big data[1]. Many of those are interactive in design, meaning when using these tools, they will require an action from the user such as the click of a button or entering some information before the content will continue to the next section.

Moreover, the typology identified several 'collections of resources': websites that collate a variety of self-produced and external resources that educate about big data. 'Toolkits', on the other hand, describe resources that provide constructive advice such as services and software to use or easy to follow steps to take to protect one's data online. As figure 3 depicts, some tools apply more than one of the identified approaches. For example, the website Privacy International includes short videos as well as text-based information. An overview of all identified tools including a brief assessment of each tool's content, design and suitability for different purposes

of teaching data literacy, geared towards practitioners, can also be found in the "Critically Commented Guide to Data Literacy Tools" (Sander, 2019).
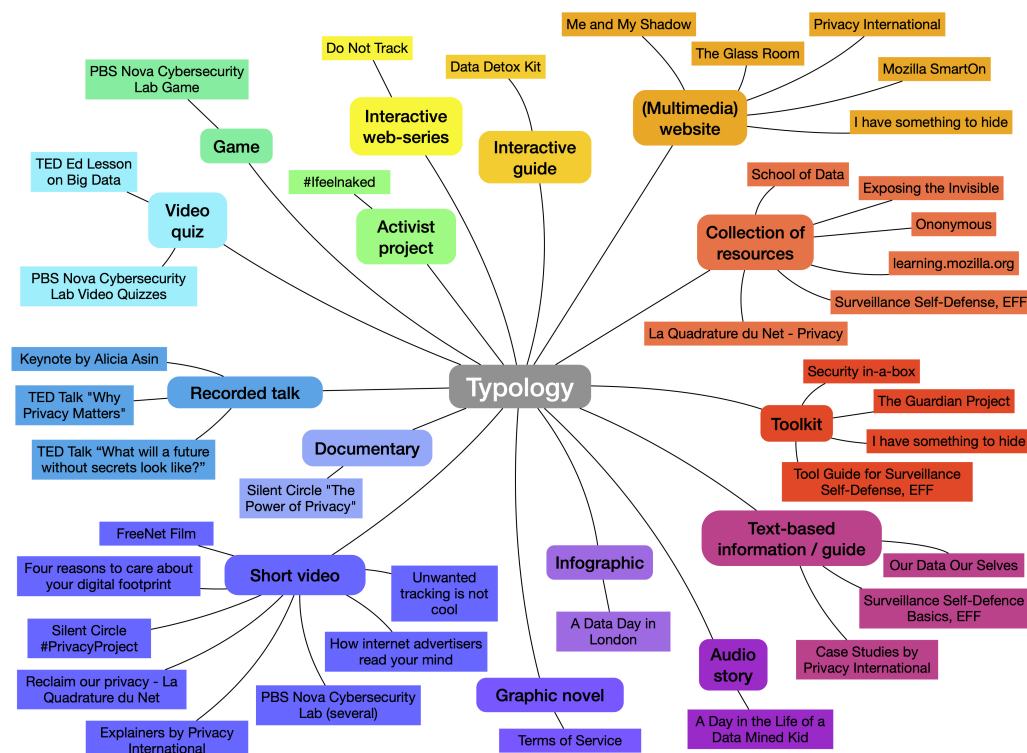

Figure 3: Typology of all identified data literacy tools and associated categories.

Secondly, the analysis revealed a variety of different production origins. As was to be expected, a large amount of tools originated from NGOs such as the 'Tactical Technology Collective' or the 'Electronic Frontier Foundation'. Yet, also further less expected actors were identified, such as the "secure mobile communications" company 'Silent Circle' that produced the documentary "The Power of Privacy" (Silent Circle, n.d.), or several private individuals who developed the website "I have something to hide". Also somewhat surprisingly, the sample did not include any resources from governmental or public service institutions, nor from traditional educational avenues and only very few efforts on the part of academia. Moreover, even though this sampling only included English-language tools, the identified online resources were not only developed by actors from English-speaking countries, but also, for example, by the French 'La Quadrature du Net', Berlin-based 'Tactical Technology Collective', and the already mentioned Portuguese individuals.

Thirdly and finally, the sampling and first analysis of tools also revealed that while many useful tools for educating about big data were identified, not all ideally suited the above definition of critical big data literacy and many required a certain level of previous knowledge. Thus, I decided to differentiate between the wider category of *data literacy tools* and more specific *critical big data literacy tools*. While the latter were found to implement all aspects of critical big data literacy (as defined above) and address a general public, data literacy tools would, for example, focus on providing resources such as teaching material about big data or technological tools to help users improve their digital security and protect their data online. Particularly the targeted audience of the tools was one of the key distinguishing factors between the two categories. Many data literacy tools did not seem to address the general public as they often lacked a general introduction to big data and directly skipped to issues such as encryption, digital security or online tracking. These tools seemed to aim at already interested individuals

with a pronounced prior knowledge on issues related to big data (e.g., 'Ononymous'; 'Exposing the Invisible') or those planning to teach about big data (e.g., 'learning.mozilla.org'). Thus, they often lacked the broader perspective and awareness-raising aspect that resources categorised as critical big data literacy tools provided.

3.2 TESTING THE TOOLS

The second part of this study tested three critical big data literacy tools: the interactive web-series 'Do Not Track', the website 'Me and My Shadow' and the short video 'Reclaim our Privacy'. Overall, the three tools predominantly had an awareness-raising and privacy-enhancing influence on the participants, leading to a generally increased concern for privacy and distinctly more privacy-sensitive internet usage[2]. This effect was nearly unambiguous in the short term (one week after using the tools), with more diverse findings mid-term. After eight months, some participants showed interest and concern only about *some* aspects of privacy (mainly about data security, see also 3.3, as well as their personal security in relation to their traceability), one had 'defaulted back' to their original attitude and behaviour but two others showed a *persistent and even growing* increase in privacy concern and behaviour. Furthermore, all participants stated increased caution in at least some situations of data disclosure online. Thus, this first testing demonstrated that while such literacy resources are of course no ultimate and perfect solution to solve all broader problems around big data, critical big data literacy tools nevertheless *can* be great ways to increase internet users' awareness, understanding and critical reflection of big data practices as well as their ability to protect their data online.

This effect was further confirmed by the participants' reflections on the tools, stating their insights through the tools, that they needed to "think about and reflect" (Participant 01) on what they learned and that they wanted to change their internet usage: "I'm going to go home and clear all my cookies now" (P09). They further stressed that they may have known about data collection but had not been aware of the impact this can have on them, arguing that issues around big data are too removed from individuals and that people "just don't necessarily realise the weight of this" (P10) or "don't think it impacts their lives, but it really, really does" (P05). This confirms people's lacking knowledge as outlined above and reaffirms that even if they are aware of data collection, internet users 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).

Appreciation for the tools
In general, the participants felt very positive about the tools. Both in the questionnaire after one week and also in the follow up interviews after eight months, they expressed their excitement about the tools, praised the tools' *accessibility*: they are "accessible to people" (P10) and "easy to read through" (P05), and found them well-suited for "*educational purposes*" (P10) as they constitute "a good way to reach people" and to "spark a little interest in them, and a little concern" (P05). Moreover, two participants emphasised how great it is that the tools "gave you all the technical information in a way that *wasn't technical*" (P05). Finally, the tools' *interactivity* was praised repeatedly: "They were all good, it's just the interactive parts of it that were better than any of the other ones" (P07).

The interactive web-series 'Do Not Track' was particularly popular with the participants. They spent the most time with this tool and highlighted its appealing visualisations and, above all, praised its *interactivity*. They believed the series was "the sort of thing that everybody – no matter what age you are would be able to grasp and be intrigued with" (P05) as "you have to be a part of it" (P04). Interactivity in general was repeatedly emphasised as a core strength of the

tools, and lacking interactivity criticised as a major weakness of the short video. This highlights that in terms of user engagement, popularity and learning effect, the importance of the interactivity of critical big data literacy tools cannot be underestimated, especially when addressing younger demographics.

Also the website 'Me and My Shadow' was popular and participants liked the "unique view on data stored about people" it provided (P04) and how it gave "straightforward advice" (P08). They particularly liked a visualisation of data traces and constructive advice in form of a five-step-list. The short video 'Reclaim our Privacy', however, was not as popular. While some participants liked it, the video and its contents were primarily forgotten after eight months, except for one participant who criticised its "scare-tactics" (P05; see below for more). Also other criticism of the tools was voiced. For example, participants found 'Do Not Track' "a bit hard to *navigate*" (P10), disliked that 'Me and My Shadow' used "lots of text" (P05) or criticised the video's lacking interactivity as "all just information" (P07).

Overall however, the participants' reflections emphasised how well-suited they found these tools – particularly 'Do Not Track' and 'Me and My Shadow' – to inform and educate people about big data practices, especially stressing the importance of interactive design, appealing visualisations and accessibility – the ability to convey complex information in a concise and easily-understandable manner. Thus, these findings also imply that some design approaches may be more or less suitable for certain goals of teaching data literacy – at least in the context of this study's sample of university students – and that it might make sense to combine some of the identified tools when educating about big data. For example, while a short video in itself may lack interactivity and vigour, it could be a great first introduction into the topic, whereas toolkits often lack an introduction to the topic, but provide great constructive advice about a more empowered internet usage. Such combination of resources could also be a workaround to issues of high production costs. Particularly suitable resources such as 'Do Not Track' that include various formats, are engaging and interactive, and include high quality design and content, come with very high production costs and are therefore rare. Combining several tools with different formats and adding interactive elements could thus constitute an alternative way to advance critical big data literacy in practice.

Calls for more education and resources

Finally, one key finding of this study was that the participants repeatedly and distinctly called for more education on big data and more data literacy resources like the ones applied in this study. Many stressed that "people are not educated enough" (P04) and that everybody should be aware of issues around big data (P05, P10). Some also explained that they "want it to be second hand nature for me to think about that issue" (P04).

As one solution, participants called for "more disclaimers" (P06) to explain data usage and for a reliable source of information on big data practices, such as a "government or independent organisation" (P04). Moreover, many called for teaching "digital awareness" (P04) from a young age in schools in order for children to learn "the ins and outs of technology" and develop a "thirst for knowledge" (P05), but also to question and critically reflect on these issues (P04). Participant P04 even considered teaching data literacy in his free time in the future, which again emphasises the participants' enthusiasm and their serious calls for more education and resources in this field. As will be further discussed below, these findings, albeit from a small and specific sample, constitute an important counterargument to discourses around people's alleged indifference about the use of their data and their supposed unwillingness to understand complex issues around big data.

3.3 IDEAS AND SUGGESTIONS FOR FUTURE TOOLS

Finally, the participants of this study, particularly those who took part in the in-depth follow up interviews after eight months, developed manifold fascinating and creative ideas on how to improve existing data literacy tools and design future resources. When enquiring about one of the original goals for this stage of the study – to learn more about participants' perceptions of and their reflections on the three tools – many individually expressed ideas for future tools that often resemble each other. These were complemented by further research on people's nuanced and complex attitudes towards privacy and big data, so that they can be summarised into something of a tentative 'guideline' for designing future critical big data literacy tools.

Format of the tool

To begin, the participants argued, such tools should include an *attention catcher*, a video or "kind of poster that attracts you to the website" (P04). Ideally, this could easily be shared on social media and thus has the potential to 'go viral'. The actual tool – a website or an app – should have a "catchy name or slogan […] that would get people's attention" (P04). The content of the tool should be *interactive and personalised* to each user and should include *multimedia usage* such as good visualisations and appealing short videos.

Examples of data harm

Moreover, future resources should include "*real stories from real people*" (P04) who have experienced negative consequences of big data practices. These could include not seeing job advertisements online due to discriminatory ad settings, not receiving a credit because of erroneous credit scoring or innocently becoming a police suspect based on biased predictive policing mechanisms, but also harms such as identity theft, data breaches or hacks (for examples, see Redden & Brand, 2017). Such stories would speak to the users "on an emotional level" (P04) and address the problem that the negative impacts of data disclosure are currently often too removed from individuals, as criticised by nearly every interviewee. Their *lacking awareness of potential negative consequences* of data disclosure online was expressed, among others, by stating that they found it "quite hard to think of any situations where it would be properly terrible" (P06) if their information was used.

Related to this, many expressed the feeling that *their data would not be of interest to anyone.* During the interviews, every participant expressed this attitude, arguing that their information was irrelevant as they are "just talking about my day-to-day life" (P04) and: "they're really big companies, why would they be that bothered about this data" (P07). It seems that the participants still lack an understanding of the value of their data and potential uses. Thus, future data literacy efforts should address this gap by, for example, presenting real-life examples of data harm that demonstrate how even 'harmless' data can be used for questionable causes and could potentially have negative consequences.

Beyond data security

Moreover, such testimonials of data harm could also be used to foster an understanding of *non-security related* issues of big data practices. The participants of this study often expressed complex and fluctuating or partly contradictory attitudes. For example, while many were concerned about problems such as data breaches, data hacks and cyber security, they were less aware of longer-term impacts of big data practices such as tracking, scoring or surveillance. Also the identified data literacy tools showed a prevalence of aspects around data security. Thus, future data literacy efforts should aim to address this imbalance.

The 'shock value'

Besides, the "*shock value*" (P10) of the tools' content was controversial with the participants.

Some said they wanted to be "scare[d] into it" (P07) and that this is "always necessary" (P10), whereas another argued that it might work to scare people, yet "it's not the right approach" and tools should work toward building awareness rather than people "doing the right thing because they're scared" (P05). Thus, it is important to find the right balance on the "narrow path" (P10) between providing factual knowledge and giving constructive advice on the one hand, and emphasising the severity of the topic enough to get and keep people engaged on the other hand.

Constructive advice to avoid digital resignation

Furthermore, as already included in the concept of critical big data literacy, *constructive advice* on how to protect one's data online should play a big part in any data literacy tool. Several participants stressed its importance, highlighting that easy-to-follow advice would not only keep people engaged and immediately give them a starting point to change their behaviour, but could also help prevent *resignation* towards privacy in light of the new information received. As already identified by other scholars (Dencik & Cable, 2017; Draper, 2017; Draper & Turow, 2019; Turow et al., 2015), some internet users have 'given up' on their data as they regard any efforts to protect their data as futile. Importantly, this attitude does not express consent with data collection but rather a feeling of inability to protect their data and thus a resignation to any such efforts.

Also in this study, *resignation* towards privacy and data collection was a critical issue. In particular, one participant was identified as a 'typical case' for resignation, who was aware of the importance of privacy at all times, yet the new information they gained from the tools made them feel "depressed" in sight of the many years of data on "every part of my life" and thus "so much knowledge" these companies had about them (P10). However, despite their initially resigned reaction, this participant later took action and made various changes to their internet usage, again emphasising the importance of privacy and explaining: "I don't know if anything I do is actually helping anything, but I try". This again stresses the relevance of such easy-to-follow advice not only in preventing but also in fighting resignation. Nevertheless, it is important to clarify that this should not entail a shift of responsibility to the individual users. While it is important and necessary that citizens of datafied societies obtain a certain ability to protect their data online -providing them with an, albeit limited, sense of agency and control and circumventing resignation - these digital skills do not constitute the key goal of critical big data literacy. Instead, this literacy's main objective lies in enabling citizens to develop an informed opinion and take part in the public debate around datafied systems.

Moreover, this participant highlighted that in order to become active and start making changes, it takes "initial investments" of time and energy (P10). This constitutes a useful insight for further constructive advice to be included in data literacy tools: the initial investment to take action should be as small as possible and it should be stressed that this is only a short, initial effort to be made. The participants suggested that one way to do this would be to develop a *simple checklist* with several easy and low-threshold first steps to take in order to protect one's data online. This would aim at building new habits, for example by suggesting alternative services people can start using, but it should also – wherever possible – provide instructions on how to remove data that have already been disclosed. One example for a similar 'to-do-list' is the "data protection toolkit" on the website 'I have something to hide'.

Options for sharing and a regular reminder

Another idea that was very popular with several participants was that of a *regular reminder*: "I think the reminder is key" (P05). They outlined that they had forgotten about changes in their internet usage they meant to make and would therefore have appreciated a reminder. One participant further explained that the common GDPR pop-ups often serve as a reminder for

them (P05). As it is "very easy to forget these things" (P05), a reminder would help make privacy-sensitive internet usage a "part of people's routine" (P04) and make them "more conscious" (P05). Finally, some participants suggested an "*aspect of sharing*" in the tool in order to "get the message around" (P04).

Overall, these manifold ideas provide a first insight into which format, design and content users of future data literacy tools might appreciate. Moreover, future initiatives and resources should be careful to find the 'narrow path' between emphasising the severity of some of the impacts big data practices can have, yet not inducing a feeling of resignation towards data collection and online privacy. Finally, the value of low-threshold constructive advice on how to curb the ubiquitous collection of personal data online should not be underestimated.

CONCLUSION

To conclude, this paper argues that data literacy today needs to go beyond the mere skills to use digital media, the internet or even the ability to use data or handle big data sets. Rather, what is required is an extended *critical big data literacy* that includes citizens' *awareness, understanding and critical reflection* of big data practices and their risks and implications, as well as the ability to *implement this knowledge* for a more empowered internet usage. Citizens need an understanding of the structural and systemic levels of changes that come with big data systems and the datafication of our societies. This paper discusses this concept, presents research findings about the kinds of critical big data literacy tools available as well as gives an insight to how student internet users view the short and also mid-term effects of these tools. The article aims to contribute to and advance debates about how data literacy efforts can be implemented and fostered.

To begin, the study provided a first, non-representative insight into the field of existing efforts to inform and educate citizens about big data practices. Nearly 40 such data literacy tools could be identified and a comparative analysis revealed a great variety in their national and production origins and in their formats. A typology revealed 15 distinct content and design approaches, including unusual and emerging formats such as a graphic novel about big data.

In a next step, three selected critical big data literacy tools were tested in a qualitative multi-methods study over three points in time. Overall, this found a positive effect of the tools on the participants' concern for privacy and their privacy-sensitive internet usage. Moreover, the study revealed fascinating findings on the users' perspective on these tools. For example, the participants highly praised the tools' interactivity, their appealing visualisations and their accessibility, and they controversially discussed the 'shock-value' of their content. Apart from this, they also clearly called for more education on the topic and more informational resources on big data, like the ones mentioned in this study. However, it is important to understand these findings, such as the popularity of interactive formats, within the context of this study's specific sample. Previous research has identified generational, socioeconomic and cultural differences in digital skills (e.g., van Djijk, 2006, p. 223). Internet users' skills but also their personal interests and their learning preferences are likely to impact their responsiveness to data literacy tools. While this study's method allowed for information-rich data on each participant's attitudes, their internet usage and their perceptions of the three data literacy tools, this method did not allow for generalising claims, and some findings may be specific for the sample's demographic.

Finally, as this study included the opportunity to follow up on some of the participants after

several months and discuss their perceptions of the applied tools in detail, the findings also offer valuable suggestions for future data literacy tools and initiatives or training programmes that aim to teach about big data practices. The participants provided detailed and creative ideas for future tools: they should have an attention catcher followed by an interactive website or app that includes personalised content, real stories about data harm, appealing visualisations, easy to follow constructive advice in form of a checklist, and options to easily recommend the tool to friends and set up a regular reminder. This was complemented by further research findings highlighting people's lacking ability to imagine negative consequences of data disclosure; the prevalence of concern about data security; and the critical issue of resignation towards privacy and data collection online. However, also here it is necessary to regard these findings in the context of the study's narrow sample and keep in mind that different target groups will likely require different design approaches when teaching about big data. Such differences between target communities would be interesting to explore in relation to responsiveness to data literacy tools in future studies.

Overall, this study contributed novel findings to various fields of scholarly research. First and foremost, this article presented the concept of critical big data literacy, building on and advancing research from the fields of critical data studies, media and digital literacy, and data literacy. Moreover, the study's findings as presented in this article reaffirm people's lacking knowledge about big data practices, and confirm existing research which has shown that while they may be aware that their data is being collected, many people only have a vague idea of the longer-term implications their data disclosure might have. In contrast to the common claim that internet users do not care about their data and feel like they have 'nothing to hide' (Solove, 2007; Marwick & Hargittai, 2018), the participants of this study were eager to learn more, highly appreciated the tools and this opportunity, and many seemed keen to protect their data. Such desire and motivation to learn more about such complex issues was unexpected. While they may not always understand *what* they have to 'hide', they were clearly not indifferent towards the usage of their data and they became more concerned when learning more about big data practices. This also confirms the findings of Worledge & Bamford (2019), who demonstrated that internet users find big data practices less acceptable the more they learn about them. Furthermore, this study advances existing research about people's resignation towards privacy by highlighting an example case and providing insights, through qualitative investigation, about some of the origins of this feeling and potential ways to prevent it.

Apart from these contributions to scholarly understanding, this study's findings on how to implement critical big data literacy are, albeit based on a small and specific sample, also relevant for practitioners in different fields and they provide first insights for policymakers. Literacy is an issue that can have deep impacts on citizenship and in order for policymakers and citizens to make informed decisions, datafied societies need informed public debate about the use and implications of data science technologies. This article suggests that one way to enhance such debate is through critical big data literacy, and it proposes what such data literacy should entail, suggestions on how to put this into practice, what to keep in mind when designing future data literacy resources and what they could look like.

ACKNOWLEDGEMENTS

pleasant and productive working atmosphere.

REFERENCES

boyd, d., & Crawford, K. (2012). Critical Questions For Big Data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, *15*(5), 662–679. https://doi.org/10.1080/1369118X.2012.678878

Bucher, T. (2017). The algorithmic imaginary: Exploring the ordinary affects of Facebook algorithms. *Information, Communication & Society*, *20*(1), 30–44. https://doi.org/10.1080/1369118X.2016.1154086

Chellappa, R. K., & Sin, R. G. (2005). Personalization versus privacy: An empirical examination of the online consumer's dilemma. *Information Technology and Management*, *6*(2), 181–202. https://doi.org/10.1007/s10799-005-5879-y

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. http://ww.w.ci-journal.net/index.php/ciej/article/view/1290

Dencik, L., & Cable, J. (2017). The Advent of Surveillance Realism: Public Opinion and Activist Responses to the Snowden Leaks. *International Journal of Communication*, *11*, 763–781. https://ijoc.org/index.php/ijoc/article/view/5524/1939

D'Ignazio, C., & Bhargava, R. (2015, September 28). *Approaches to Building Big Data Literacy* [Paper presentation]. Bloomberg Data for Good Exchange Conference, New York. http://rahul-beta.connectionlab.org/wp-content/uploads/2011/11/Edu_DIgnazio_52.pdf

D'Ignazio, C., & Bhargava, R. (2018). Cultivating a Data Mindset in the Arts and Humanities. *Public*, *4*(2). https://public.imaginingamerica.org/blog/article/cultivating-a-data-mindset-in-the-arts-and-humanities/

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. https://doi.org/10.1075/idj.23.1.03dig

Do Not Track (2015). About. *Do Not Track*. https://donottrack-doc.com/en/about/

Doteveryone. (2018). *People, Power and Technology: The 2018 Digital Attitudes Report*. https://attitudes.doteveryone.org.uk

van Dijk, J. A. G. M. (2006). Digital divide research, achievements and shortcomings. *Poetics, 34*(4–5), 221–235. https://doi.org/10.1016/j.poetic.2006.05.004

Draper, N. A. (2017). From Privacy Pragmatist to Privacy Resigned: Challenging Narratives of Rational Choice in Digital Privacy Debates. *Policy & Internet*, *9*(2), 232–251. https://doi.org/10.1002/poi3.142

Draper, N. A., & Turow, J. (2019). The corporate cultivation of digital resignation. *New Media & Society*, *21*(8). https://doi.org/10.1177/1461444819833331

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

Fotopoulou, A. (In press). Conceptualising critical data literacies for civil society organisations:

agency, care, and social responsibility. *Information Communication and Society*.

Garcia, A., Mirra, N., Morrell, E., Martinez, A., & Scorza, D. (2015). The Council of Youth Research: Critical Literacy and Civic Agency in the Digital Age. *Reading & Writing Quarterly*, *31*(2), 151–167. https://doi.org/10.1080/10573569.2014.962203

Gray, J., Gerlitz, C., & Bounegru, L. (2018). Data infrastructure literacy. *Big Data & Society*, *5*(2). https://doi.org/10.1177/2053951718786316

Grzymek, V., & Puntschuh, M. (2019). *Was Europa über Algorithmen weiß und denkt. Ergebnisse einer repräsentativen Bevölkerungsumfrage*. Bertelsmann Stiftung. https://doi.org/10.11586/2019006

Hammer, R. (2011). Critical Media Literacy as Engaged Pedagogy. *E-Learning and Digital Media*, *8*(4), 357–363. https://doi.org/10.2304/elea.2011.8.4.357

Hansen, A., & Machin, D. (2013). *Media and communication research methods*. Palgrave Macmillan.

Hinrichsen, J., & Coombs, A. (2013). The five resources of critical digital literacy: A framework for curriculum integration. *Research in Learning Technology*, *21*. https://doi.org/10.3402/rlt.v21.21334

Kennedy, H., & Moss, G. (2015). Known or knowing publics? Social media data mining and the question of public agency. *Big Data & Society*, *2*(2). https://doi.org/10.1177/2053951715611145

Kitchin, R., & Lauriault, T. P. (2014). *Towards critical data studies: Charting and unpacking data assemblages and their work*. [The Programmable City Working Paper No. 2]. Maynooth University. http://mural.maynoothuniversity.ie/5683/

Malhotra, N. K., Kim, S. S., & Agarwal, J. (2004). Internet Users' Information Privacy Concerns (IUIPC): The Construct, the Scale, and a Causal Model. *Information Systems Research*, *15*(4), 336–355. https://doi.org/10.1287/isre.1040.0032

Marwick, A., & Hargittai, E. (2018). Nothing to hide, nothing to lose? Incentives and disincentives to sharing information with institutions online. *Information, Communication & Society*, *22*(12), 1–17. https://doi.org/10.1080/1369118X.2018.1450432

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

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

Müller-Peters, H. (2019). Big Data: Chancen und Risiken aus Sicht der Bürger [Big Data: Chances and Risks from Citizens' Perspective]. In S. Knorre, H. Müller-Peters, & F. Wagner (Eds.), *Die Big-Data-Debatte* (pp. 137–193). Springer. https://doi.org/10.1007/978-3-658-27258-6_3

Ofcom. (2019). *Adults: Media use and attitudes report* [Report]. https://www.ofcom.org.uk/__data/assets/pdf_file/0021/149124/adults-media-use-and-attitudes-report.pdf.

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Allen Lane.

Pangrazio, L. (2016). Reconceptualising critical digital literacy. *Discourse: Studies in the Cultural Politics of Education*, *37*(2), 163–174. https://doi.org/10.1080/01596306.2014.942836

Pangrazio, L., & 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

Philip, T. M., Schuler-Brown, S., & Way, W. (2013). A Framework for Learning About Big Data with Mobile Technologies for Democratic Participation: Possibilities, Limitations, and Unanticipated Obstacles. *Technology, Knowledge and Learning*, *18*(3), 103–120. https://doi.org/10.1007/s10758-013-9202-4

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

Pritchard, A. (2008). *Ways of Learning. Learning Theories and Learning Styles in the Classroom* (2nd ed.). Routledge. https://doi.org/10.4324/9780203887240

Pybus, J., Coté, M., & Blanke, T. (2015). Hacking the social life of Big Data. *Big Data & Society*, *2*(2). https://doi.org/10.1177/2053951715616649

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

Sander, I. (2019). *A Critically Commented Guide to Data Literacy Tools*. https://doi.org/10.5281/zenodo.3241422

Sander, I. (2020). *Critical Big Data Literacy Tools – Engaging Citizens and Promoting Empowered Internet Usage*. ORCA. http://orca.cf.ac.uk/id/eprint/131843

Silent Circle. (n.d.). Silent Circle | Secure Enterprise Communication Solutions. https://www.silentcircle.com/

Smith, H. J., Milberg, S. J., & Burke, S. J. (1996). Information Privacy: Measuring Individuals' Concerns about Organizational Practices. *MIS Quarterly*, *20*(2), 167–196. https://doi.org/10.2307/249477

Solove, D. J. (2007). "I've Got Nothing to Hide" and Other Misunderstandings of Privacy. *San Diego Law Review*, *44*(4), 745–765. https://digital.sandiego.edu/sdlr/vol44/iss4/5/

Turow, J., Hennessy, M., & Draper, N. (2018). Persistent Misperceptions: Americans' Misplaced Confidence in Privacy Policies, 2003–2015. *Journal of Broadcasting & Electronic Media*, *62*(3), 461–478. https://doi.org/10.1080/08838151.2018.1451867

Turow, J., Hennessy, M., & Draper, N. A. (2015). *The tradeoff fallacy: How marketers are misrepresenting American consumers and opening them up to exploitation* [Report]. Annenberg School for Communication. https://www.asc.upenn.edu/sites/default/files/TradeoffFallacy_1.pdf

Vasileiou, K., Barnett, J., Thorpe, S., & Young, T. (2018). Characterising and Justifying Sample Size Sufficiency in Interview-Based Studies: Systematic Analysis of Qualitative Health Research over a 15-Year Period. *BMC Medical Research Methodology*, *18*. https://doi.org/10.1186/s12874-018-0594-7

Worledge, M., & Bamford, M. (2019). *Adtech: Market Research Report*. Information Commissioner's Office; Ofcom. https://www.ofcom.org.uk/__data/assets/pdf_file/0023/141683/ico-adtech-research.pdf

Yates, S., Carmi, E., Pawluczuk, A., Wessels, B., Lockley, E., & Gangneux, J. (2020). *Understanding citizens data literacy: thinking, doing & participating with our data* (Me & My Big Data Report 2020). Me and My Big Data project, University of Liverpool. https://www.liverpool.ac.uk/humanities-and-social-sciences/research/research-themes/centre-for-digital-humanities/projects/big-data/publications/

Zuboff, S. (2015). Big other: Surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology*, *30*(1), 75–89. https://doi.org/10.1057/jit.2015.5

APPENDIX: LIST OF ALL IDENTIFIED DATA LITERACY TOOLS

Acquisti, A. (2013, June). *What will a future without secrets look like?* [Video]. TED Conferences. https://www.ted.com/talks/alessandro_acquisti_why_privacy_matters

Asin, A. (2014, November 20). *Big Data and the Hypocrisy of Privacy* [Keynote, video]. Strata + Hadoop World Europe 2014, Barcelona. https://www.youtube.com/watch?v=oWwQfgpvlzI

Gaylor, B. (Director). (2015). *Do Not Track*. Upian; Arte; Office national du film du Canada; Bayerischer Rundfunk. https://donottrack-doc.com/en/

Disconnect. (2013, June 30). *Unwanted tracking is not cool* [Video]. YouTube. https://www.youtube.com/watch?v=UU2_0G1nnHY

The Economist. (2014, September 11). *How internet advertisers read your mind* [Video]. YouTube. https://www.youtube.com/watch?v=8KYugpMDXAE

Electronic Frontier Foundation. (n.d.). *Surveillance Self-Defense*. https://ssd.eff.org

Electronic Frontier Foundation. (n.d.). *Surveillance Self-Defense Basics*. https://ssd.eff.org/module-categories/basics

Electronic Frontier Foundation. (n.d.). *Surveillance Self-Defense Tool-Guides*. https://ssd.eff.org/module-categories/tool-guides

Fight for the Future. (n.d.). *I feel naked*. https://www.ifeelnaked.org

FreeNet Film (2012, December 15). *Do you care about your privacy in the web?* [Video]. YouTube. https://www.youtube.com/watch?v=jtGtIxgS7io

Greenwald, G. (2014, October 10). *Why Privacy Matters* [Video]. YouTube. https://www.youtube.com/watch?v=pcSlowAhvUk

The Guardian Project. (n.d.). *The Guardian Project*. https://guardianproject.info

Hill, A. (2014, September 15). A Day in the Life of a Data Mined Kid. *Marketplace*. https://www.marketplace.org/2014/09/15/education/learning-curve/day-life-data-mined-kid

Internet Society (2015, May 14). *Four Reasons to Care About Your Digital Footprint* [Video]. YouTube. https://www.youtube.com/watch?v=OA6aiFeMQZ0

Julie, Ozoux, P., Daniel, & Bouda, P. (n.d.). *I have something to hide*. https://ihavesomethingtohi.de

Keller, M. & Neufeld, J. (2014, October 30). *Terms of Service. Understanding our role in the world of Big Data*. http://projects.aljazeera.com/2014/terms-of-service/#1

La Quadrature du Net. (2014, February 11). *Reclaim Our Privacy* [Video]. YouTube. https://www.youtube.com/watch?v=TnDd5JmNFXE

La Quadrature du Net. (n.d.). *Website – Section Privacy*. https://www.laquadrature.net/en/Privacy

Mozilla. (n.d.). *Learning.mozilla.org*. https://web.archive.org/web/20200216134228/https://learning.mozilla.org/en-US/

Mozilla. (n.d.). *SmartOn Privacy and Security*. https://web.archive.org/web/20180510112403/https://www.mozilla.org/en-US/teach/smarton/

PBS Nova Labs – Cybersecurity Lab. (n.d.). *A Cyber Privacy Parable* [Video]. Public Broadcasting Service. http://www.pbs.org/wgbh/nova/labs/lab/cyber/2/1/

PBS Nova Labs – Cybersecurity Lab. (2014). *Cyber Lab* [Game]. Public Broadcasting Service. http://www.pbs.org/wgbh/nova/labs/lab/cyber/research#/newuser

PBS Nova Labs – Cybersecurity Lab (n.d.). *Cybersecurity videos*. Public Broadcasting Service. http://www.pbs.org/wgbh/nova/labs/videos/#cybersecurity

Privacy International. (n.d.). *Privacy International Website*. https://privacyinternational.org/

Privacy International. (n.d.). *Invisible Manipulation: 10 ways our data is being used against us*. https://www.privacyinternational.org/long-read/1064/invisible-manipulation-10-ways-our-data-being-used-against-us

Privacy International (n.d.). *What Is Data Protection?* https://privacyinternational.org/explainer/41/101-data-protection

Silent Circle (2016, January 27). *The Power of Privacy*. https://www.youtube.com/watch?v=BvQ6I9xrEu0

Silent Circle (2015, January 27). *#Privacy Project*. https://www.youtube.com/watch?v=ZcjtEKNP05c

Smith, T. (n.d.). *Big Data* [Video]. TED-Ed. https://ed.ted.com/lessons/exploration-on-the-big-data-frontier-tim-smith

Tactical Technology Collective. (2017). *A Data Day in London*. https://tacticaltech.org/#/news/a-data-day

Tactical Technology Collective. (n.d.). *Data Detox Kit*. https://datadetoxkit.org/de/home

Tactical Technology Collective. (n.d.). *Exposing the Invisible*. https://exposingtheinvisible.org

Tactical Technology Collective. (n.d.). *The Glass Room*. https://theglassroom.org

Tactical Technology Collective. (n.d.). *Me and My Shadow*. https://myshadow.org

Tactical Technology Collective. (n.d.). *Ononymous*. https://ononymous.org

Tactical Technology Collective. (n.d.). *Our Data Our Selves*. https://ourdataourselves.tacticaltech.org

Tactical Technology Collective. (n.d.). *School of Data*. https://schoolofdata.org

Tactical Technology Collective. (n.d.). *Security-in-a-box*. https://securityinabox.org/en/

FOOTNOTES

1. For a list of all identified tools and their URLs, see appendix.

2. More detailed findings on the effect the tools had on my participants' internet usage and their privacy attitudes are illustrated in a previous publication (Sander, 2020).