

Data Citizenship: Data Literacies to Challenge Power Imbalance Between Society and “Big Tech”

ELINOR CARMİ
City, University of London, UK

SIMEON J. YATES¹
University of Liverpool, UK

We argue here that data literacies and capabilities are an integral part of data justice. Based on focus group data collected as part of a 3-year empirical project research project, we find that citizens remain unaware of key aspects of the digital ecosystem, which exacerbate the power imbalance between big technology (data processors) companies and citizens (data subjects). Citizens feel concerned about the way it is operating, they do not feel confident enough to be able to address that. We find that “networks of literacy” among friends, colleagues, and trusted organizations are crucial for citizens’ capabilities. These networks influence citizens’ ability to convert their available means into capabilities to support civic engagement and their communities.

Keywords: data literacies, data justice, digital inequalities, citizens’ rights, citizens’ capabilities

Over the last decade, and accelerated by the COVID-19 pandemic, we have seen a transformation where every aspect of society from health, through culture, news, employment, the economy, relationships, and civic activity has been “datafied” (van Dijck, 2014). Though “big-tech” companies such as Facebook, Amazon, and Google have a leading role in this transformation, so do many smaller companies, quangos, and local and national governments. All are engaged in the collection, aggregation, processing, and (ab)uses of data derived from “us” as citizens. As part of this inseparable relationship that we now have with “big tech” there has been an increase in the capacity to track our behaviors over time and the trading of our data among multiple data brokers. Dependence on these platforms exposes us to a range of potential harms,

Elinor Carmi: Elinor.Carmi@city.ac.uk

Simeon Yates: scssy@liverpool.ac.uk

Date submitted: 2021-10-18

¹ Funding acknowledgment: This “Me and My Big Data—Developing Citizens’ Data Literacies” project was funded by the Nuffield Foundation under Grant No. FR-000021473. The Nuffield Foundation is an independent charitable trust with a mission to advance social well-being. The foundation has funded this project, but the views expressed are those of the authors and not necessarily of the foundation.

Copyright © 2022 (Elinor Carmi and Simeon Yates). Licensed under the Creative Commons Attribution Non-commercial No Derivatives (by-nc-nd). Available at <http://ijoc.org>.

from dark patterns, mis-/dis-/mal-information, filter bubbles, opaque terms of use, and lack of appropriate regulation and enforcement (Carmi, 2021), as well as the risks of data breaches and criminality online.

We therefore argue that there is a widening of the power imbalance between *data subjects* (*citizens*) and *data processors* (*big tech, government*). Data processors being those who come to own and manage “data subject’s” data (Hintz, Dencik, & Wahl-Jorgensen, 2017, p. 732). The differentiation between citizen data (data from interaction with the state and civic action) and consumer data (our patterns of consumption) is also becoming blurred. In this context, citizens’ ability to make rights claims (Isin & Ruppert, 2015) to change how their data and their communities’ data are being (ab)used is key. This includes the ability to change regulations and influence what is seen as the “new normal” of “networked privacy” (Hargittai & Marwick, 2016, p. 3737). “Networked privacy” refers to the situation where others (data processors) can and do transgress our privacy through extracting and trading our data: both data that citizens knowingly share with platforms or services (e.g., by signing up to a service) and data they indirectly “give off” through their digital activity (e.g., via web cookies). Being able to make digital rights claims about how data are used by “big tech,” or the state, is key to ensuring data justice for citizens.

However, to be able to make such rights claims citizens need the capabilities that allow them to both understand and assess the (ab)uses of their data. Importantly, they need to navigate the digital ecosystem of the social, political, economic, and technological context of data processing. Citizens therefore need the appropriate data literacies to be able to take appropriate civic action. To engage in this asymmetric ecosystem, citizens need to understand and be able to challenge the power imbalance between themselves (data subjects) and data processors (big tech, government). They also need to understand governments’ role in democratically regulating this ecosystem. *Therefore, data literacies are an integral and necessary part of data justice.*

In this article, we link ideas of data justice to our empirical work assessing our Data Citizenship framework (Yates, Carmi, Lockley, Wessels, & Pawluczuk, 2021). This framework views “data literacies” as the capacity of citizens to “Do” things with, “Think” critically about, and “Participate” socially and politically around their data. Data literacies underpin citizens using data they create and share for their own purposes, their understanding of what is done with it by platforms and governments, and their ability to make rights claims about data collected and used by others. The focus group work reported here formed the third phase of our project, “Me and My Big Data: Developing Citizens’ Data Literacies,” undertaken between 2018 and 2021. The Data Citizenship framework draws on an analysis of recent literature on data and digital literacies combined with ideas from critical democratic education (Dewey, 1930; Freire, 1996). We argue that because of its collective, socially contextualized, and people-centered qualities, democratic education provides a useful foundation for future data literacies education and research interventions (Carmi & Yates, 2020; Carmi, Yates, Lockley, & Pawluczuk, 2020).

We first consider data justice in terms of fundamental rights, digital rights claims, capabilities, and citizens’ engagement with data. We then compare these ideas with the findings from our focus group work. We conclude the article with a reflection on the challenges presented by our results, such as limited understanding of digital ecosystems, and the need for truly critical digital literacies education. In setting up the link between data justice, data citizenship, and data literacies we explore four key issues. First, we discuss the need to

problematize the role of individual agency, especially the extent to which platform design, regulation, and legal frameworks assume a very knowing user with extensive agency over both the technologies they use and the data they share. Second, we discuss the need to understand citizens' engagement with data and platforms in terms of socially, educationally, and economically constrained circumstances such that citizens' capability to act with agency are also constrained. Third, we discuss citizens' levels of data literacies as products of their lived experiences. We consider the likelihood that many people have low levels of data literacies. Fourth, we discuss data literacies and the ability to be active citizens making rights claims, especially about their data and the activities of platforms and governments.

Data Justice and Individual Agency

As Taylor (2017) argues, when it comes to data technologies in the Global North, data justice has often been approached through the idea of fundamental rights and has mainly considered privacy and freedom of speech. Such a framing is problematic. Citizens will need to clearly identify data (ab)uses to respond to them, and often redress will be at the individual level, rather than at the level of the community or society. The major gap we find in these approaches is their assumption that citizens can both recognize the (ab)uses being conducted and know what to do about them. In other words, the assumption is that citizens have the appropriate data literacies to identify and challenge the (ab)uses of their data that they are concerned about. These approaches therefore require high-level literacies that very few citizens hold.

Similarly, Lehtiniemi and Ruckenstein (2019) show in their work on the Finnish data activism initiative "MyData" that citizen-centric initiatives already assume that people know what data are and how to use it with agency:

Even if MyData activists promote monetization of personal data as only one of many possible technical solutions, the proposition is symptomatic of a belief that individuals can control the market. Promoting the personal data market assumes that people are competent to make informed choices concerning their data. (Lehtiniemi & Ruckenstein, 2019, p. 7)

It is precisely these seemingly "informed choices" and the practices they involve that we want to examine and unpack.

Capabilities and Social Justice

It is important to emphasize that the fundamental rights approach is not the sole method to assess the imbalance between data processors (big tech) and data subjects (citizens). The capability approach (Nussbaum, 2002; Sen, 2009) provides an alternative framework. This approach includes two key claims about the achievement of social justice that also hold for data justice. First, primary importance is placed on the achievement of freedom and well-being. Second, understanding what counts as well-being and the barriers to this should be understood in terms of citizens' capabilities and their achievement of these. Capabilities can be understood as the "doings and beings" that people can achieve if they so choose. These might be very basic such as being well-nourished. They might also include varied things such as getting educated, being able to express sexuality, or being able to use the Internet.

For citizens to be able to convert their available means into a functioning capability crucially depends on their social, political, economic, and environmental conditions. In the context of digital technologies, simply having the skills to use devices is not enough. Access to devices, electricity, network connections, even a quiet place to work may also be needed to create the necessary capabilities that allow a citizen to engage in digital activity. Capabilities can therefore be viewed as “real” or “substantive” freedoms as they describe opportunities for “doing and being” that have been cleared of potential obstacles. This contrasts with mere formal rights and freedoms.

For Nussbaum (2002) social justice comes into being as we provide spaces for human agency through the development of capabilities, awareness, knowledge, and skills. Not only do these need to be provided, but they must also be proactively put into action by individuals or communities. Importantly, as Nussbaum argues, “It is necessary, as well, to prepare the material and institutional environment so that people are actually able to function” (Nussbaum, 2002, p. 132). Citizens’ choices are highly dependent on the material and social context. For data justice, this model would argue that data literacies in and of themselves are not a silver bullet. Delivering data justice is inevitably bound up with addressing aspects of the overarching problems we experience in society. Systemic inequality, austerity, racism, sexism, or technological solutionism all present potential barriers to data justice. Yet data literacies remain key in helping citizens understand their social and digital context and the extent to which that context is a limiting factor in attaining data justice. Therefore, we acknowledge the limits of data literacies in enacting data justice and in changing the “new normal.” At the same time, we would also advocate for the critical data literacies that can support the social and digital rights claims that citizens make to challenge the circumstances that limit data justice.

People and Data

An overview of the literature underpinning the Data Citizenship framework can be found in the work by Yates and colleagues (2021). However, we want to highlight some key works that have informed our project and the gaps we aim to fill in this article. Pangrazio and Sefton-Green (2019) argue that “data literacy also involves both critical understandings of the technological infrastructure and the political economy of digital platforms, as well as strategies and tactics to manage and protect privacy and resist being profiled and tracked” (p. 214).

These two components of understanding the digital and data ecosystem and having the capabilities to act touch on the systemic and the personal aspects of data citizenship. As Payal Arora (2019) argues, we should *not*, as researchers or policy makers, *assume* what people value in terms of their digital and data activities. Such assumptions tend to become highly normative and skew debates away from citizens’ self-perceived needs to the activities deemed “worthy” or “most important” by policy makers or researchers. As Arora notes, for some citizens, the most important activities may be streaming films, interacting with family, or even watching pornography. Therefore, as Kennedy (2018) argues, “[I]t is important to take account of what non-expert citizens themselves say would enable them to live better with data, based on their everyday experiences of datafication” (p. 18). These experiences are important both for research to explore and for intervention projects to consider when developing approaches to data literacies.

In terms of data justice, the key questions are: Do citizens have the capabilities to undertake the activities they deem important, and can they overcome contextual barriers they face in doing so? In the context of data citizenship, do citizens have the capabilities to make the digital rights claims they consider important? Can they articulate the demands of government or platforms around the (ab)use of their data? From these flows the question of how to create an environment that supports these capabilities.

Pangrazio and Sefton-Green (2019) argue that there have been three responses to datafication: legislative, such as the General Data Protection Regulation (GDPR); tactical (technical solutions like encryption); and educational. However, "the actual mechanisms by which that knowledge might be acquired, what it might consist of, and how indeed the possession of it would affect a whole range of behaviors, is often vague" (Pangrazio & Sefton-Green, 2019, p. 211).

The delivery of data literacies through education may therefore be presented as a way to ensure participation in a datafied society. Yet in many cases this is done without necessarily addressing the contextual and environmental context of citizens. Access and skills or knowledge do not guarantee a context in which citizens have the data and digital capabilities they need. Digital and data capabilities need to be "real" or "substantive" freedoms, real opportunities for digital and data "doing and being," not just knowledge, skills, or formal rights. However, at the same time, the problem remains that many citizens have low data literacies and do not fully understand the (ab)use of their data and the consequences of that (ab)use. When their understanding is limited, demands and practices will also be limited.

Citizenship in a Datafied Society

Capabilities to undertake everyday data and digital activities are a key part of data justice. However, if citizens lack the digital and data capabilities to express, demand, and help deliver the political and democratic change they need, they remain trapped within a constrained context. The way citizens can exercise their digital rights has been explored through various lenses. Ruppert and Isin (2015) have discussed the idea of "digital citizenship" at the core of which is the capability of digital citizens to make digital "rights claims"—these claims being fundamental to being digital citizens (p. 44). These claims might be made of others or of the digital *conventions* they encounter—the ways of "doing and being" digital within which citizens are constrained. Ruppert and Isin (2015) argue that citizens' performance of conventions is undertaken through and embedded within digital platforms. For example, communication conventions are undertaken through the acts such as posting and interacting via e-mail or Facebook. Similarly, purchasing conventions are undertaken through Amazon or eBay, and political acts might be through invocations of GDPR rights.

Citizenship can also be understood in terms of addressing datafication. For example, Milan and van der Velden (2016) define data activism as a "possible manifestation of activism in the information society—one that, however, explicitly engages with the new forms information and knowledge and their production take today, challenging dominant understandings of datafication" (p. 61).

Citizen activism, as acts of resistance and reappropriation of technology, implies and requires citizens who have very high levels of digital and data literacies. Citizen activism therefore needs a deep and critical understanding of how the data ecosystem works as well as how to reappropriate it. As Raffaghelli

(2020) argues, a lot of the counter cultures that are active against datafication—articulating clear digital and data rights claims—themselves use high-level digital methods. As she argues, “[t]hese activisms require advanced technical skills and civic engagement that go well beyond the actual opportunities in situations of marginalization and global inequities” (Raffaghelli, 2020, p. 235).

In the context of these types of “citizenship” approaches digital and data justice focuses on the potential to enact digital and data capabilities and make rights claims. There is much less focus on the supporting social and economic context that makes capabilities functional (or not).

We have two views on data justice that intersect with data citizenship. First, following Nussbaum (2002) and Sen (2009), digital and data justice flows from the creation of social and economic contexts in which citizens can make functional the capabilities they see as important. Second, data justice flows where citizens can effectively make digital rights claims and counter the conventions imposed on them by platforms and other organizations. This of course is a subset of the capabilities that citizens may see as important. In both cases a lack of digital and data literacies will severely hamper the possibility of addressing data justice for citizens. In the remainder of this article, we explore these limitations through an examination of focus group responses from our project.

Methods

Following a systematic literature review and a U.K. nationally representative survey ($n = 1,542$) we conducted 14 focus groups with U.K. citizens. From our survey work we identified six user types around which we constructed “Data Citizenship Personas.” These personas were derived quantitatively from our survey data. Respondents were grouped on Internet behaviors using latent class analysis. Corresponding demographics were modeled using logistic regression and data literacies using multiple correspondence analysis (Carmi et al., 2020; Yates et al., 2021; Yates, Kirby, & Lockley, 2015; Yates & Lockley, 2018):

1. *Extensive political users (10% of users)*. High probability of engaging in all forms of digital media use—including political action and communication.
2. *Extensive users (20% of users)*. High probability of engaging in all forms of digital media use—except political action and communication.
3. *Social and entertainment media users (17% of users)*. High likelihood of engaging with social media and entertainment media (e.g., Netflix and YouTube).
4. *General users (no social media; 31% of users)*. Lower likelihoods of engaging in most digital media forms but not social media.
5. *Limited users (22% of users)*. Limited engagement in all forms of digital media.
6. *Nonusers*. Currently non-Internet users.

Twelve of our focus group interviews consisted of respondents whose demographics were likely to make them part of groups 3, 4, and 5, and two of our focus groups consisted of respondents likely to be in groups 1 or 2. We previously identified that low digital access and skills closely correspond with other key demographics such as age, income, poverty, being in social housing, low educational attainment, long-term ill health, and ethnicity. We chose to focus on the more limited or narrow users (groups 3, 4, and 5) to unpack

one of the key themes that emerged in our national survey. Mainly, that U.K. citizens' "Data Thinking" was notably limited—especially for these groups. We also felt that these groups' views were not well represented in research on data literacies and that it is key for their voice to be heard. The focus groups were therefore undertaken on Zoom. Access was provided by local digital inclusion support groups and community centers that were part of the Good Things Foundation network in the United Kingdom. Carmi, Lockley, and Yates (forthcoming) provide full details of the running of the Zoom focus groups and reflections on the teams' experience of undertaking such sessions with digitally marginalized groups. The focus groups covered the following topics: what do we mean by data; establishing a baseline definition of data; mapping a data day; discussion of data days, platforms used, and data shared; examination of platforms, their data sharing, and data use; discussion of data rights, concerns, and networks of data literacy. Details of the composition of the focus groups can be found in Table 1.

Table 1. Focus Group Participants.

| Focus Group | Target Group | Group Size | Likely User Types |
|--------------------|---|-------------------|---|
| A | Younger people (<30 - preferably under 25) who may have some digital skills and/or are social media focused | 9 | Social and entertainment media users or extensive users |
| B | Younger people (<30 - preferably under 25) who may have some digital skills and/or are social media focused | 5 | Social and entertainment media users |
| C | Younger people (<30 - preferably under 25) who may have some digital skills and/or are social media focused | 4 | Social and entertainment media users |
| D | Older adults (55+) who are offline or with limited digital skills | 7 | General or limited users |
| E | Older adults (55+) who are offline or with limited digital skills | 6 | General or limited users |
| F | Older adults (55+) who are offline or with limited digital skills | 4 | General or limited users |
| G | Older adults (55+) with digital skills | 7 | General users |
| H | Older adults (55+) with digital skills | 5 | General or extensive users |
| I | Older adults (55+) who are offline or with limited digital skills | 3 | General or limited users |
| J | Older adults (55+) who are offline or with limited digital skills | 4 | General or limited users |
| K | Adults (30+) with limited digital skills | 2 | Limited users |
| L | Adults (30+) with limited digital skills | 4 | General or limited users |
| M | Students (18+) with higher digital skills | 5 | Extensive or extensive political users |
| N | Students (18 +) with higher digital skills | 5 | Extensive or extensive political users |

We conducted a thematic analysis of the focus group data following the Data Citizenship framework (Yates et al., 2021). The results from our analyses were grouped under the three dimensions of our framework. However, we particularly focused on aspects of Data Thinking. This ties in closest with citizens' initial starting points and underpins personal capabilities.

Focus Group Findings

As with our survey findings, respondents in all our groups, especially the limited and narrow users, demonstrated very limited understanding of the digital ecosystems in which they operated. This knowledge is best described as "patchy." Although this lack of knowledge did not mean respondents were unaware of the broad picture of data (ab)use by data processors, all respondents felt unease and levels of disempowerment, especially about how platforms extract data, the uses to which data are put, and the potential of this use being undertaken by "third parties." This unease was described in quite emotive terms and undermines any idea that users are happy with their data being collected and used "so long as the service is free."

Importantly, the processes to verify information, the learning of new skills, or the development of better awareness were tied to users "networks of literacy." These local, often personal but sometimes digital, connections are key to understanding the context of digital citizens. While people's levels of literacy are usually measured and taught on an individual level, the concept of "networks of literacy" shows that digital and data literacies must be considered in a community-based and contextual manner. They are both a resource for and a limitation of citizens' development of relevant capabilities. Your "networks of literacy" can determine what is available to you not only in terms of material means but also in terms of the digital literacies you can develop (Yates & Carmi, forthcoming).

Data Doing

Data Doing is the one component of our model that matches closest to many other models of data literacy (Yates et al., 2021). It predominantly consists of the practical actions that citizens take to manage, protect, and make basic use of their data. Yet we found very "patchy" and very contextual Data Doing practices among respondents. For example, in relation to keeping personal data secure, practices varied greatly between our limited and extensive users. Nearly all respondents expressed low levels of skill and confidence in this area:

Not confident at all. . . . I feel like cookies get the better of me and I need to learn better how to. When you go onto a website it says about privacy and I really don't know the best way to say I don't want cookies. . . . In the end I give up and say oh I'll agree, and then you end up with loads of information you don't want. (H5; F; 72; 18+)²

This contrasts with a respondent who had post-18 education and self-identified as being "skilled" ("a geek") who talked about data protection: "I do think because I'm a geek I tend to do things that people

² Demographic details: Focus group code (e.g., H5), see Table 1; gender; age; and education—pre-18 (<18) or post-18 (18+)

online don't tend to: Wipe out history all the time, remove the cookies weekly, start again next week if they want" (H1; M; 64; 18+). However, all groups discussed the limits of skills and behaviors:

I would say fairly confident in like changing passwords and things, but then I don't really know what else I should be doing, like other than that. So, I don't know, if that's what you would call knowledgeable or unknowledgeable, like I don't really know other steps that I should be taking. (N1; F; 21; 18+)

As visible in these examples, respondents constructed their "limited skills" as personal failure. Respondents placed the fault or responsibility *on themselves and not on platforms or policy makers*: "I don't know how to do X" or "I should do X, but I don't." In constructing the problem in this way respondents took on responsibility for things that were external to them such as platform design or personal circumstances. They did not critique the conventions that they were subject to, especially the fact that platforms make difficult the task of keeping data safe.

Respondents also described a lack of desire to post things on social media platforms. Though silent participation such as scrolling through feeds (Carmi, 2020) rather than posting is a very long-standing phenomenon on online fora, we noticed that a lack of trust in platforms was a clearly articulated reason for this. This predominantly stemmed from a fear of providing too much data to the social media platform and the possibility of the platform (ab)using the data in ways they did not agree with:

I don't post, and I rarely like because I'm not convinced that any information that you put on there, you know particularly if you're commenting on somebody else's post or something like that, is necessarily used in the best way. And I don't necessarily blame the organization for that, I just think that that's maybe a comment on society, you know. . . . We've got new starters at [the digital center], I'm introducing them to Facebook, and there is always a note of caution of how you use it, you know, and what you put on there. (I1; F; 70; <18)

Other reasons for not posting included keeping information to closed groups (though of course platforms themselves may have access): "I don't post anything. If I like anything I just sharing it about my life with my friends on WhatsApp, that's it" (C2; M; 23; 18+.). Others expressed concerns about the way data are used to generate advertising:

I tend to be a bit cautious in giving too much information because I'm not really sure what they actually do with it. Because if you buy anything online or look at something online the next day, you'll find in your email that you've got all sorts of things as a result of what you've been looking at, so they know what you've been doing, what you've been looking at. (G4; F; 84; 18+)

These and similar arguments permeated all our focus group discussions. We view these responses as a form of Data Doing to resist the (ab)uses of data by platforms. It implies some level of what we call

Data Thinking. Yet this sits alongside considerable unease and resignation to the “new normal” of “networked privacy” (Hargittai & Marwick, 2016, p. 3737). We will discuss this resignation and related unease below.

Data Thinking

For people to become active data citizens they need to first understand the data ecosystems in which they exist and operate. More than that, they need to be able to make rights claims for themselves and their communities. We explored a range of questions in this area, in particular, the nature of data: What are the “data” you share when using platforms? How are they being produced, extracted, and traded? How are algorithms influencing the way you engage with things? We also explored digital activism and participation: Which digital rights can citizens claim in this context?

Most participants in our focus groups had limited or no understanding of these issues. On a fundamental level, if citizens do not understand how the elements of our datafied society work in their everyday digital “doing and being,” it is difficult for them to make rights claims. It is also difficult for them to use data for their own advantage. This situation turns citizens into mere consumers, efficient enough workers, and “data providers” for platforms but without offering them the appropriate critical literacies. These literacies would be those that assist them to use data in a way that benefits them and their communities, rather than provide value and resources to platforms.

In the context of Data Thinking we identified five main limitations in the focus group discussions:

1. No understanding of basic definitions such as “data”
2. No understanding of how digital systems and media actually work
3. No understanding of what data platforms do
4. The potential consequences of data tracking and trading
5. The organizations involved in the digital ecosystem

In many cases, these issues were described in combination. This lack of Data Thinking has a direct impact on personal, community, and civic behavior.

What Is Data?

Scholars, policy makers, and practitioners constantly use the word “data” to denote the information we share and “give off” online. As a result, there is an inherent assumption among scholars and policy makers that citizens understand the term. As our findings show, this is not the case. When we explored this term, respondents clearly had quite limited awareness of this definition, or even what things constitute such data. For example, several participants focused on the volume of data they get in Internet packages: “I think data is good to help me out with my Internet work or homework, any tasks, and to make my life easier if I use data” (C4; M; 18; <18). Others see it more like a commodity resource that gets used:

Data, for me, I think, data is just like the thing we use to upload our information on the media or back up information, like using Facebook. You need data for Facebook, you need data for Instagram. I mean just to get access onto the Internet is data. (C3; M; 24; <18)

For many other participants the focus was on quite “administrative” data: “All the information about yourself, about as much information about you as you are prepared to give. . . . Where you live, name, date of birth, address—information like that. Email address.” (G4; F; 84; 18+)

The only groups to immediately focus on the academic and policy definition were our younger “Extensive” users:

[I]t’s anything they can use, particularly when it comes to social media, like boost your interaction and the amount of time spent on the platform because more time spent on the platform means more ad revenue for them. And the same with shopping, just like trying to find your preferences. And then when it comes to things like Google it’s kind of tailoring the ads that you do get to things that you’re more likely to buy although, saying that, they do obviously certain people pay them a lot of money to put ads up there that aren’t necessarily things you’ll want most, but things that are giving let’s say Google, the most money. (M4; M; 20; 18+)

Our focus group discussions and survey findings make clear that detailed knowledge of both overt and covert data collection and sharing is quite limited for many users. Yet, this does not mean they are not aware it happens. In all the focus groups respondents expressed an awareness that data from platform use are collected; but what, how, and why were often poorly understood.

How Digital Systems and Media Actually Work

Several of our respondents used relevant terms such as “cloud” or “encryption” while also clearly demonstrating a very limited understanding of what these involved:

I’ve often wondered where it goes, but I know it’s stored somewhere like Facebook Google they allegedly say it’s anonymized but with WhatsApp it’s encrypted allegedly, but I just assumed it’s all stored in a cloud somewhere. I’ve often wondered where it goes. (F1; F; 45+; <18)

Similarly, there is awareness of tracking and third-party data sharing but a lack of deeper understanding of the processes by which data are extracted, linked, used, and sold:

If I’d be honest with you, I really don’t know, I just know that they give it out to people and they sell it to people. I often sit there and think how the hell have they got my information, you know, but somebody will say something and put 2 and 2 together come up with 7 and realize that they’ve sold my information to them. It’s just too easy to give away your information, that’s all I can tell you. (I3; F; 60; <18)

The lack of specific knowledge about what happens with data leads to a complex position to be comfortable with this "new normal." For example, A6 describes this as being "cynical" while "liking to think" collected data are put to a "good purpose":

I think I'm a little cynical with it because no one really knows. I think there are so many options as to where it could be going. I think there's always, there's so many documentaries going around sort of saying how it influences ads and your behavior online, and how certain organizations and certain companies can see what you're browsing for, see what you're clicking on, and use that to influence what you click on next. So, what advertising companies I click on, I have a cynical view on it, I'd like to think it goes to a good purpose but I'm not too sure, to be honest. (A6; F; 25; 18+)

What Data Platforms Do

Discussions about the details of the data ecosystems of platforms found mixed awareness in general but with little detailed understanding. For example, K1 could explain the overall structure but could not provide detail:

Obviously, they sort of harvest all of your data to target advertising for you. I mean if you go look at a pair of shoes online, 5 minutes later, you're getting adverts for shoes on your web browser and that sort of thing. Like you say, it's shared with a dozen other people so that their companies can target you as well. (K1; M; 46; <18)

Many descriptions of such processes included similar phrases such as "I don't know," "I assume that," "I wondered where" (see F1 and I3 quoted above). Again, there is "awareness," but lack of specific knowledge. This makes it much harder for citizens to express and put forward digital rights claims. For example, when it comes to citizens' GDPR rights, how can citizens demand their data to be erased or changed when they do not know what data are and who is involved in the data ecosystem?

The Potential Consequences of Data Tracking and Trading

Similarly, respondents were aware of and very concerned about the potential consequences of data tracking and trading, especially with third parties. The key occasional reference point for such concerns were Cambridge Analytica and the Netflix film *The Social Dilemma*. For example, H3 stated:

I mean the Facebook; some organizations are more scrupulous than others in that way. I mean Facebook-Cambridge Analytica stuff was terrifying for me, what they did in America with voter suppression, it's quite powerful isn't it? If it gets into the hands of the wrong people. (M; 41; 18+)

However, we also found quite contradictory and confused responses demonstrating conflicted experience and understanding of these issues. Despite the lack of understanding about the technical

processes of data tracking, extracting, and trading respondents expressed concern about the consequences of this across a range of contexts such as the workplace:

It feels like people know where you are all the time. I try not to use my phone all the time, I use the company sat nav and I think even if you have your telephone on you, they can track where you are anyway. It is becoming a bit like a big brother state anyway where everybody knows where you are. (D4; M; 62; 18+)

For others it was (ab)use by state and security services or criminals:

Sometimes maybe if we are not careful with security or policy, we will share our data with hackers or something that will affect our lives especially with bank account or specific information that no other know about that may affect our life or something like that. (C1; F; 29; <18)

For some it was the pressure of relentless advertising:

Say I'm on Instagram and it'll pop up with something like Asos has a sale on, and that leads to you think OK I'll just have a browse and then you end up going onto there and spending like £100 on nothing really just out of the blue. I do feel like there is an indirect pressure just because it's so much in your face. (A7; M; 22; 18+)

Across the spectrum of Data Thinking issues, we find the same pattern in our focus groups: awareness or experience of the issues yet a lack of detailed understanding that would allow citizens to have confidence in their ability to address the issues they identify. This awareness with lack of knowledge combined with concerns and a lack of confidence regarding skills and practical actions (Data Doing) create a very disempowered context, one where citizens have limited resources to develop the functioning capabilities to both act and make rights claims as digital citizens. This leads, we would argue, to one of the key findings from the focus groups and survey: the lack of Data Participation actions and the considerable sense of disempowerment and worry felt by respondents.

Data Participation: Feeling Disempowered

Citizens' feelings of disempowerment are mostly associated with three issues already exemplified in the quotes and comments above:

1. Lack of confidence in their digital "efficacy" with regard to managing and protecting data
2. Concerns over the potential consequences of data (ab)use, tracking, and trading, including being manipulated
3. Feeling disempowered in relation to the complexity of platforms and technologies

These lead to ongoing feelings of worry and anxiety over uses of their data by platforms. However, as both our survey and our focus group findings demonstrate, this does not lead, for many citizens, to practical steps to address concerns.

Feelings of Disempowerment and Worry

All respondents, even those with greater confidence, expressed some aspect of resignation and disempowerment in relation to managing, protecting, and controlling their data, and as noted above, very often constructed these issues as a personal failing or lack of skills. This links to the argument that the "new normal" of "networked privacy" conventions is such that major digital platforms systematically prescribe how citizens, as users of platforms, must offer up their data and be subject to the outcomes of platforms' decision-making processes. It reflects the internalization of these prescriptions that digital citizens are subjected to. As a result our respondents often talked about it being hard to "fight back" or "resist" the practices and processes used by platforms:

I think the whole point of these apps and stuff there's so much to them that it's very hard to fight back. It's like there's so much that you feel inert almost, so as I say, I try to avoid apps and I've uninstalled apps and deleted accounts and such just purely because T&Cs hides a lot behind it legally and miles and miles of content. It's just incredibly difficult to pass through it all but also, I would say I'd like to think I'm confident, but I don't know how far the iceberg goes, if that makes sense. (N4; M; 25; 18+)

This resignation is not passive. It leads to changes in behavior, such as not posting on social media or avoidance of certain apps. It is also not without its social and personal cost. Importantly, in all the focus groups, respondents talked about feeling resigned but also "uneasy" with data being tracked or shared. Very often this unease was expressed in quite emotive terms such as "creepy": "Basically, there's a digital version of you that's stored somewhere that you can't really access or know about . . . but you don't really know who has access to that and who doesn't which is a bit creepy" (M4; M; 20; 18+). Respondents were therefore not happy about providing data for a service "just because it is free." In fact, the complex trade-offs they make are more about managing worry, fear, and unease against the pressure to use, need to use, or usefulness of platforms: "It is useful, and it is creepy at the same time because it's kind of like they're spying on you" (C3; M; 24; < 18). This includes elements of fear: "I think it is scary about how much information the companies and governments have about me. It is scary" (B1; M; 26; <18). Also, revulsion:

I don't like the idea of them being able to see private conversations on a messenger app, for example, or something like that. Things like location, which I'm sure Google maps send, I don't like that idea they can know where you are and what your routine is and everything. So yes, data like that is horrid. (M2; M; 20; 18+)

Words such as "worry," "creepy," "disturbing," "invasion," and "horrid" can be found throughout our discussions. Respondents clearly expressed anxiety and unease, what we might call "wrongness," about their "datafied" experiences, particularly about:

- how platforms extract, use, and trade data about citizens, and
- their responses to being placed in a situation of being resigned to these activities.

Disempowerment therefore stems from people's feeling that they do not have a choice but to use these platforms. Predominantly because these platforms are so big and effectively ubiquitous for everyday life practices, that "everyone and everything is there." It also stems from a lack of Data Thinking, which involves knowledge about privacy settings, as well as alternative and more privacy-friendly services (e.g., DuckDuckGo search engine) that those with much greater awareness may use (Milan & van der Velden, 2016). Importantly, disempowerment stems from feeling like these ecosystems are the reality, what Dencik (2018) calls "Surveillance Realism." But this understanding or acceptance of "reality" derives from not knowing what rights claims are available to them as citizens.

The Problem of Practical Action

Why does this worry and concern not translate into action? Two reasons seem to come through in the data. First, as described above a lack of knowledge and awareness (limited Data Thinking) to support practical action (limited Data Doing) is a clear impediment. Second, even among our more extensive users, there is the problem of "practical action." For example, few engaged with platforms' terms and conditions. As a result, they had no knowledge as to what rights over their data they handed over to these platforms: "Again, I think it's very difficult, you sign onto something, and you get the privacy policy thing that you never read because it is so ridiculously long, so everyone just puts a little tick in the box" (E1; M; 45+; <18). Even when confident in such tasks the practicalities of "getting things done" leads to a resigned agreement to default settings:

Before I was very confident, I was thinking I will take care of myself and read the small print and all that. But now it's like even on YouTube they nag you, they say cookies and all that and it's a big list, and sometimes I want to switch the video for research, and I just go like agree, because it's time consuming, and sometimes I'm tired and I want to finish my work. (F2; F; 45+; <18)

Others were simply resigned to the fact that it was too much work to manage settings and that there was in fact no real choice but to go along with the demands of the platform or system:

If you don't agree to their T&C you can't use the service then you won't be able to do anything online. All the apps and programs they have boxes to tick, you have to agree to them to use them, without agreeing to them you won't be able to use them, or you don't have a choice really. (D3; M; 54; <18)

Not knowing how the data ecosystem works and which practical actions to take also stems in part from these manipulative interfaces and the long and jargon-laden "terms and conditions" that are designed to push people to "consent" to the exploitative data ecosystem (Carmi, 2021). Such manipulative designs exacerbate the power asymmetries between the data processors and data subjects, especially where the latter have lower data literacies.

Citizen Resources: Networks of Literacy

There are many examples in the focus groups discussions of respondents drawing on and overtly establishing social networks to support their digital activities. Very often this was based around close family or social ties:

I've got a 6-year-old who's a right whizz, so you know, but also I've got a boyfriend who works in IT and he's very helpful. In fact, I've got two friends both work in IT, so I ring them up and say so how do I do this, and they explain over the phone how to do it. (E5; F; 45+; <18)

For many respondents, reaching out to their social media contacts was another means of support. Very often such interactions were about "reassurance" or "confirmation" with regard to respondents' concerns or activities. Alternatively, others actively used digital technologies to create social networks to support their personal or community activity:

I use WhatsApp for the community benefits, like if there is information from the library or from [the digital center] they want to know, so I take the leaflet and take a picture and use WhatsApp group to inform the community that this thing is happening. And I use NHS website to help my family who can't access it or don't know about it so they can look at their condition and what to ask the doctor and things like that. (F2; F; 45+; <18)

We would argue that these networks of literacy, some quite extensive and informed and some very small, are in fact key to citizens' data and digital literacies. They provide the basis for their navigation of digital content. They are the "go to sites" for the acquisition of skills and knowledge and for the verification of information. Very often they underpin community engagement. They support citizens in "doing" things with data, they support their "thinking" about and with data, and they underpin any Data Participation they undertake.

Conclusion

Developing data literacies for citizens is crucial but it is not the sole solution. The broader social, economic, technological, and personal contexts of citizens need to be both part of the analysis and also part of any supporting interventions. Following the capability approach (Nussbaum, 2002; Sen, 2009), we argue that access to devices, network connections, as well as skills and literacies are all needed to create the necessary capabilities that allow a citizen to engage in digital activity. Digital capabilities must be "real" or "substantive" opportunities for "doing and being" and only come into being when cleared of any potential obstacles. At the same time digital and data citizenship require real opportunities to affect the digital world, how it is run, regulated, and governed. In the case of data, we as citizens need to be able to understand what happens to our data, to be able to use data for our personal or collective goals, and to be active participants in our deified society. Developing citizens' digital and data literacies is therefore clearly *necessary* but *not sufficient* to deliver this.

Unfortunately, the analysis of our focus groups and our survey data (Yates et al., 2021) demonstrate that there is a considerable gap between many citizens' awareness and skills and those they likely need to deliver the types of Data Participation, digital citizenship (Isin & Ruppert, 2015), and activism (Raffaghelli, 2020) that we and others have described or envisaged. It is also clear that citizens have enough awareness of the issues around the (ab)use of their data to express their unease and concern in quite emotive terms. Yet their lack of skills and literacies limit the extent to which they can act on and think through these concerns. Even those with deeper knowledge and skills suffer enforced resignation because of the realities of "practical engagement" with digital technologies and platforms; this may be through "dark patterns" (Carmi, 2021) that are designed to constrain or nudge users toward platform-beneficial behaviors or simply because of the constraints of time and life pressures. Your child needs a video on volcanoes for homework, so you click "yes" to the cookies on the website to get the task done.

Data justice and rebalancing of power between data subjects (citizens) and data processors (big tech, government) therefore requires three things. First, the material social context to support "real" or "substantive" opportunities for "doing and being" digital. Second, the depth and breadth of digital and data literacies to understand and use that material context to act as digital and data citizens. Third, meaningful opportunities for digital citizenship rights claims and activism so that citizens can seek to change their digital context. Without these things it will be very difficult for citizens to make their data and digital capabilities functional and achieve the "real" or "substantive" opportunities for "doing and being" digital as they wish.

It is a matter of concern that citizens in our data understood their limitations as a personal failure to understand systems or protect their data rather than seeing these as limitations placed on them by the current conventions of digital systems and platforms. As such they internalize the problem as one of personal improvement rather than one of demanding better systems and better behavior from platforms. We would therefore argue that data and digital literacy educational interventions need to include a strong critical democratic education component (Dewey, 1930; Freire, 1996) that will underpin citizens' active Data Doing, Thinking, and Participating. Otherwise, digital skills training will simply reinforce the sense of resignation and compliance to conventions that we currently find in our research project.

References

- Arora, P. (2019). *The next billion users*. Cambridge, MA: Harvard University Press.
- Carmi, E. (2020). *Media distortions: Understanding the power behind spam, noise, and other deviant media*. New York, NY: Peter Lang.
- Carmi, E. (2021). A feminist critique to digital consent. *Seminar.net*, 17(2), 1–21. doi:10.7577/seminar.4291
- Carmi, E., & Yates, S. J. (2020). What do digital inclusion and data literacy mean today? *Internet Policy Review*, 9(2), 4–14. doi:10.14763/2020.2.1474

- Carmi, E., Lockley, E., & Yates, S. (forthcoming). Inclusive digital focus groups—lessons from working with citizens with limited digital literacies. In J. Skopek (Ed.), *Research handbook of digital sociology*. Cheltenham, UK: Edward Elgar.
- Carmi, E., Yates, S. J., Lockley, E., & Pawluczuk, A. (2020). Data citizenship: Rethinking data literacy in the age of disinformation, misinformation, and malinformation. *Internet Policy Review*, 9(2), 1–22. doi:10.14763/2020.2.1481
- Dencik, L. (2018). Surveillance realism and the politics of imagination: Is there no alternative. *Krisis: Journal for Contemporary Philosophy*, 2019(1), 31–43.
- Dewey, J. (1930). Individualism, old and new. II: The lost individual. *New Republic*, 61, 294–296.
- Freire, P. (1996). *Pedagogy of the oppressed* (Rev. ed.). New York, NY: Continuum. (Originally published in 1970).
- Hargittai, E., & Marwick, A. (2016). “What can I really do?” Explaining the privacy paradox with online apathy. *International Journal of Communication*, 10, 3737–3757.
- Hintz, A., Dencik, L., & Wahl-Jorgensen, K. (2017). Digital citizenship and surveillance| society. *International Journal of Communication*, 11, 731–739.
- Insin, E., & Ruppert, E. (2015). *Being digital citizens*. London, UK: Rowman & Littlefield Publishers.
- Kennedy, H. (2018). Living with data: Aligning data studies and data activism through a focus on everyday experiences of datafication. *Krisis: Journal for Contemporary Philosophy*, 2018(1), 18–30.
- Lehtiniemi, T., & Ruckenstein, M. (2019). The social imaginaries of data activism. *Big Data & Society*, 6(1), 1–12. doi:10.1177/2053951718821146
- Milan, S., & van der Velden, L. (2016). The alternative epistemologies of data activism. *Digital Culture & Society*, 2(2), 57–74. doi:10.14361/dcs-2016-0205
- Nussbaum, M. (2002). Capabilities and social justice. *International Studies Review*, 4(2), 123–135. doi:10.1111/1521-9488.00258
- Pangrazio, L., & Sefton-Green, J. (2019). The social utility of “data literacy.” *Learning, Media and Technology*, 45(2), 208–220. doi:10.1080/17439884.2020.1707223
- Raffaghelli, J. E. (2020). Is data literacy a catalyst of social justice? A response from nine data literacy initiatives in higher education. *Education Sciences*, 10(9), 233. doi:10.3390/educsci10090233
- Sen, A. K. (2009). *The idea of justice*. Cambridge, MA: Harvard University Press.

- Taylor, L. (2017). What is data justice? The case for connecting digital rights and freedoms globally. *Big Data & Society*, 4(2), 1–14. doi:10.1177/2053951717736335
- van Dijck, J. (2014). Datafication, dataism and dataveillance: Big data between scientific paradigm and ideology. *Surveillance & Society*, 12(2), 197–208. doi:10.24908/ss.v12i2.4776
- Yates, S., & Carmi, E. (forthcoming). Citizens networks of digital and data literacy. In J. McDougall & K. Fowler-Watt (Eds.), *Palgrave handbook of media misinformation*. London, UK: Palgrave.
- Yates, S. J., Carmi, E., Lockley, E., Wessels, B. E., & Pawluczuk, A., (2021). *Me and my big data final report: Understanding citizens' data literacies*. Liverpool, UK: University of Liverpool.
- Yates, S. J., Kirby, J., & Lockley, E. (2015). Digital media use: Differences and inequalities in relation to class and age. *Sociological Research Online*, 20(4). doi:10.5153/sro.3751
- Yates, S. J., & Lockley, E. (2018). Social media and social class. *American Behavioral Scientist*, 62(9), 1291–1316. doi:10.1177/0002764218773821