



ME

AND MY BIG DATA

DEVELOPING CITIZENS DATA LITERACY: A SHORT GUIDE

Authors:

Professor Simeon J. Yates (University of Liverpool)

Dr. Elinor Carmi (City University)

1 Contents

1 Contents	2
2 How To Use This Guide?	4
Part 1: Understanding Data Literacy	6
3 What is data literacy?	6
3.1 Literacy.....	6
3.2 Media Literacy.....	7
3.3 Information Literacy.....	7
3.4 Computer Literacy	7
3.5 Digital Literacy	8
3.6 Combining these definitions.....	8
3.7 Digital Literacy	9
3.8 Data Literacy	9
3.9 Data literacy reading	10
4 Why support, develop, or teach citizens' data literacy?	11
4.1 Data literacy and citizenship.....	12
4.2 Democratic education and Data Citizenship	13
4.3 Comprehensive list of everyday data literacy activities	14
4.3.1 <i>What this guide does not cover</i>	15
Part 2: Planning your approach	16
5 Seven principles for developing citizens Data Literacy	16
6 Audiences and resources	17
6.1 Audience	17
6.2 Example personas.....	18
7 Where to find resources.....	21
7.1 Journalistic, documentary, and public media	21
7.1.1 <i>Films and documentaries</i>	21
7.1.2 <i>Media content</i>	21
7.2 Activists and key organisations web sites.....	22
7.3 Academic and investigative journalism books	23
Part 3: Suggested Activities	24
8 What is data?	24
8.1 Suggested activity 1: What is data?.....	24
8.2 Suggested activity 2: A data day	25
8.2.1 <i>Resources</i>	26
9 Understanding data collection.....	26
9.1 Suggested activity 3: How does my favourite platform collect data on me?	26
9.1.1 <i>Resources</i>	27

10 Keeping safe online	27
10.1 Suggested activity 4: Data safety	27
11 Understanding online privacy	28
11.1 Suggested activity 5: Planning for online privacy	28
11.1.1 Resources	28
11.2 Suggested activity 6: Privacy in a digital society	29
11.2.1 Resources	29
12 Data participation and data citizenship	29
12.1 Suggested activity 5: Planning for online privacy	30
12.2 Suggested activity 8: Exploring your data rights	30
12.2.1 Resources	31
12.3 Suggested activity 9: Community data civic action	31
12.3.1 Resources	31
13 Acknowledgments	32

This guide has been developed to support the development of training or community interventions designed to develop what we call **democratic education for data citizenship**. These can range from the most basic introduction to “what is data” and its basic use – for example how Google search and web ‘cookies’ work – through to the complexities of GDPR legislation or unpicking misinformation online. The goal is not to provide a prescriptive list of “skills” nor a “one size fits all” solution. Our research has demonstrated that concern about issues of data use and abuse in contemporary society is very high among citizens – but awareness of the details and how to address these concerns are quite low. We also found that understanding, skills, and awareness varied greatly between different groups. As a result, the starting point for a journey towards democratic data citizenship varies by community and audience. Meeting citizens “where they are” is key here and therefore an intervention needs to start with an understanding of this. The findings from our Me and My Big Data research project provide a starting point:

- <https://www.liverpool.ac.uk/media/livacuk/humanitiesampsocialsciences/meandmy-biddata/Understanding,Citizens,Data,Literacies,Research,,Report,Final.pdf>

We would strongly recommend reading this report alongside this guide – though be aware it is intended for academics, or expert stakeholders and policy makers. It contains additional literature data and examples that you might find useful. But our research also made clear that the specifics of citizens circumstances, the community they are part of, and their lived experience must be considered. This is best understood by citizens and communities themselves and those working with them.

This guide is not meant to be exhaustive nor as a direct teaching guide for citizens. It is designed for **educators, stakeholders, and social action groups**. It is designed to provide some initial starting points for activities, provide background reading and advice on interventions. It is not designed to be used as a ‘training package’ by students or citizens. Why have we taken this approach? Why not just provide a top ten things to do guide? This was

in fact our intention at the start of the project. Designed shortly after before the Cambridge Analytica Scandal broke the project envisioned a simple guide to “data literacy” as a key output. As our research developed and identified the variation in awareness, as issues like Cambridge Analytica came to the fore alongside misinformation online, and as we talked to different communities and constituencies, it became clear the idea of a single “one-size fits all” solution or simple ‘top tips’ was a naive approach. This guide is therefore focused on **supporting educators, stakeholders, and social action groups to understand the issues and to plan for interventions**. We have therefore tried to balance informing you about the research and academic evidence, linking you to available resources and examples, and practical advice.

But to be clear **this guide is not designed to support basic digital skills**. Existing courses and interventions provide many routes to the foundational skills and knowledge citizens need. This guide is specifically designed to address the gaps in citizens’ knowledge identified in our research that are not covered by these foundational courses and interventions. Though of course our research points out that many citizens still lack these foundational components. For support in providing foundational digital skills and awareness, as well as context specific support (e.g. around health and finances), we would recommend the materials from the Good Things Foundation:

- <https://www.goodthingsfoundation.org/learn/learn-my-way/>
- <https://www.goodthingsfoundation.org/learn/make-it-click/>

We are acutely aware that digital technologies are also constantly changing and so examples and research findings can go out of date – therefore we plan to update the web-based version of this guide each year.

The guide has therefore been put together for people looking to design interventions and educational activities to support **democratic education for data citizenship**. The guide does not assume that you are already familiar with these issues. Throughout we provide links to recent academic publications or key websites that explore the issues under discussion.

Part 1 of this guide sets the scene by explaining what we mean by Data Literacy and how this relates to ideas of Media Literacy, Information/Computer Literacy, and Digital Literacy. It also sets out the case for Data Literacy education – especially in terms of Digital Citizenship.

Part 2 considers some of the key factors you may need to take into account before you design or implement an intervention – especially your audience and appropriate learning resources.

Part 3 then focuses on the areas of significant weaknesses in Data Literacy and concern citizens had about data and platforms.

These were:

- Understanding what data is and how digital media and systems use it to deliver services to citizens.
- Concerns about and a need to understand data citizens are sharing and ‘giving’ away as they use digital media and systems.
- Concerns about and a need to understand privacy in a digital and ‘datafied’ society.
- Understanding digital and data rights as citizens of a digital and ‘datafied’ society

Part 3 provides examples of activities and resources that you might employ with different groups or in different intervention contexts. As we note below, we see Digital and Data Literacies as two sides of the same coin. Very often, specific digital skills or digital society issues include aspects of both Digital and Data literacy – as a result, development and training will often include combinations of both. For example, teaching basic digital skills such as using a social media app could also include teaching about how the platform uses data about the person in order to offer up content. Few resources available on the internet are therefore solely focused on Data Literacy. We hope that the set of links and resources provided here will help you put together support or interventions relevant to the people and communities you are working with.

We want to keep this guide ‘live’. We would welcome suggestions for further links, readings, examples or case studies that you think would be helpful for others. We would of course love to hear about any materials or interventions that you might develop.

Please send any suggestions to simeon.yates@liverpool.ac.uk

PART 1: UNDERSTANDING DATA LITERACY

3 What is data literacy?

This section is designed to provide you with a very quick overview of how we have developed the ideas of Digital and Data Literacy in the Me and My Big Data project. It provides an overview of how these ideas link to related work around citizens understandings of media and information.

There are in fact a lot of terms used to describe citizen's ability to use digital media and systems. These are often overlapping and there are few if any 'absolute' definitions. Key ones include:

- Literacy
- Media literacy
- Information literacy
- Computer literacy
- Digital literacy
- Data literacy

We will briefly look at each of these in turn. These are provided to help guide you through these issues. However, the resources provided could also be used to teach appropriate audiences about the specific issues of Media, Information, Digital and Data Literacies. In the sections that follow we walk through each of these concepts. Which is also a short history of ideas about 'literacy'. As we argue in a moment what is defined as 'literacy' is always about the use of the communication technologies available at a point in time. We conclude the section arguing that Digital and Data Literacy are two sides of the same coin. Taken in their simplest sense Digital Literacy addresses the nature of contemporary media systems and the skills to use them – akin to the idea of Media Literacy. Data Literacy addresses the data ecology and economy that underpins these contemporary systems – akin to the idea of Information Literacy. Hopefully the section provides you with a brief overview of how these ideas have developed and link together.

3.1 Literacy

The idea of literacy is of course ancient – with evidence of writing going back many thousands of years. The modern conceptions of literacy stem from the post Victorian era of mass literacy. Ideas of literacy range from the basic skills of reading and writing through to the ability to work with complex texts, be they Tolstoy or technical academic science papers. Over the years, very complex social, political, and cultural understandings of literacy have developed. These are rooted in the idea of literacy practices - the "uses" of literacy by citizens and communities. It is important to note that the ideas of Digital and Data literacy are not simply one of making an analogy between a skillset needed for 'written' texts and one for 'computer systems'. Writing is itself a technology and written literacy and digital literacy fundamentally intersect today as the majority of text consumed by citizens is provided via digital media and systems:

"In perhaps 50 years' time, our understanding of the nature of literacy and of the social functions of texts will have so radically changed that few will be alive to attest to 'how things were' at the close of the 20th century"¹.

Literacy is therefore always about the use of the communication technologies available at the time, though it is of course a highly social and culturally differentiated set of practices. Importantly, certain literacy practices are deemed more worthy or useful – in other words there are notable **value judgements** made around what types of behaviours and knowledge should citizens have. These points all hold for use of digital media and systems including the normative assumptions about what is 'good' Digital and Data Literacy.

¹ Danet, B. (1997). Books, letters, documents: The changing aesthetics of texts in late print culture. *Journal of Material Culture*, 2 (1), 5-38.

3.2 Media Literacy

The idea of media literacy grew from applying the idea of literacy to non-print media, especially broadcast media. As digital communications technologies have come along media literacy definitions have tried to include the breadth of broadcast and personal communications media.

The UK media regulator Ofcom defines Media Literacy as **“the ability to use, understand and create media and communications in a variety of contexts”**

Resources link: <https://www.ofcom.org.uk/research-and-data/media-literacy-research>

The Centre for Media Literacy uses the following three-part definition:

- **Media Literacy is a 21st century approach to education.**
- **It provides a framework to access, analyze, evaluate, and create messages in a variety of forms - from print to video to the Internet.**
- **Media literacy builds an understanding of the role of media in society as well as essential skills of inquiry and self-expression necessary for citizens of a democracy.**

Resources link: <https://www.medialit.org/reading-room/what-media-literacy-definitionand-more>

3.3 Information Literacy

Ideas of Information Literacy often appear in relation to the use of organised information sources – such as libraries and databases. Very often, as with the definition below, they include an element of being a ‘citizen’ or of ‘citizenship’.

The UK Chartered Institute of Library and Information Professionals (CILIP) define Information Literacy as

“the ability to think critically and make balanced judgements about any information we find and use. It empowers us as citizens to develop informed views and to engage fully with society.”

Resources link: <https://www.cilip.org.uk/news/421972/What-is-information-literacy.htm>

3.4 Computer Literacy

Computer literacy is less clearly defined and is often broadly viewed as the ability to use ‘computers’. Different definitions can stress different aspects of computer use including:

- **“understanding the basic processes of computers and technology and being able to use those processes”**To explore the extent of citizens’ understanding of the use of their data, (and its aggregate as ‘big data’) by industry, government and third sector.
- **“understanding of computer characteristics capabilities and applications, as well as an ability to implement this knowledge”**

- **“the comfort level someone has with using computer programs and other applications that are associated with computers”**
- **“the level of knowledge and skills about information and communication technologies and how to use it in your work and everyday life”**

Some definitions of computer literacy are much closer to the CLIP definition of Information Literacy. Overall, definitions range from having practical skills of keyboard and mouse use, to the ability to code, or use of all forms of Information and Communication Technologies.

3.5 Digital Literacy

In 1997 Paul Gilster² defined Digital Literacy as “**the ability to both understand and use digitised information**”. Since then, there have been a wide range of definitions that overtly build on media, computer, and information literacy as well as civic or social engagement.

For example, the American Library Association (ALA) defines digital literacy as “**the ability to use information and communication technologies to find, evaluate, create, and communicate information, requiring both cognitive and technical skills**”.

Resources link: <https://literacy.ala.org/digital-literacy>

In the UK AdvanceHE define Digital Literacies(s) – plural – as a set of “**capabilities required to thrive – that is to be an effective and responsible participant – in a digital society**”.

Resources link: <https://www.advance-he.ac.uk/knowledge-hub/digital-literacies>

Table 1: UNESCO Media and Information Literacy (note alignment of columns is not significant)

Information Literacy						
Define and articulate information needs	Locate and access information	Assess information	Organize information	Make ethical use of information	Communicate information	Use ICT skills for information processing
Media Literacy						
	Understand the role and functions of media, and Internet communications companies in democratic societies	Understand the conditions under which media can fulfil their function	Critically evaluate media content in the light of media functions	Engage with media for self-expression and democratic participation	Review skills (including ICTs) needed to produce user-generated content	
Digital Literacy						
Use of digital tools	Understand digital identity	Recognize digital rights	Assess AI issues	Improve how to communicate digitally	Manage digital health	Practice digital security and safety

3.6 Combining these definitions

UNESCO has proposed a combined Media and Information Literacy approach, to which they have added a set of digital skills (see Table 1). This framework underpins UNESCO’s approach to supporting organisations and countries to provide Media and Information Literacy education. Again, this definition and the supporting documentation focus on citizens “thinking critically and clicking wisely”.

Resources link: <https://en.unesco.org/news/media-and-information-literate-citizens-think-critically-click-wisely>

² Gilster, P. (1997). *Digital literacy*. John Wiley & Sons, Inc.

3.7 Digital Literacy

We take Digital Literacy to include the majority of the components listed by UNESCO as Media Literacy but refined to incorporate digital media and systems. By digital media and systems we mean the full range of digital devices, platforms and content citizens encounter. This includes but is not limited to:

- Broadcast TV and Radio – mostly now produced and delivered digitally
- Streaming platforms – for music, film TV and other content
- Web content and platforms of all forms – from static web pages to shopping sites, from newspapers to government services
- Apps – that provide services and content via devices

- Games – from mobile game apps to multi-platform blockbuster games

Why such a wide definition? A contemporary smartphone user might engage with all such systems within the space of a day on that one device. They might slip from news content via Twitter, to a film clip on Facebook, to booking a health appointment, to playing Wordle in the space of a few minutes.

To have the **“capabilities required to thrive – that is to be an effective and responsible participant – in a digital society”** requires the ability to work with all these media and systems.

Table 2: Digital Literacy

Digital Literacy						
Ability to effectively use digital media and systems to generate content, communicate and achieve a life citizens value.	Understand the role and functions in democratic societies of digital media and systems and the organisations that own or control them.	Understand the conditions under which digital media and systems can fulfil their social and personal functions.	Ability to critically evaluate the content and behaviour of digital media and systems in the light of their social and personal functions.	Ability to engage with digital media and systems for self-expression, democratic participation and to achieve a life citizens value.	Understand their rights over digital media and systems as citizens of a digital society.	Understand how to use digital media and systems safely, securely, legally and ethically.

3.8 Data Literacy

We take Data Literacy to include the majority of the components listed by UNESCO as Information Literacy but expanded to incorporate the nature of digital media and systems. Information Literacy was defined when most information was well managed and curated by key gatekeepers – from librarians to publishers. Even in the early years of the internet, information remained managed by a more limited list of organisations. The Process for creating information – as opposed to news or entertainment – was quite slow and deliberate.

This is no longer the case. Not only is content produced constantly on digital platforms, but the very acts of using platforms generate data and information that is then used. Though many human and institutional gatekeepers remain, this mass of data and information

is often created, collected, processed, and then affects citizens via automation and algorithms. This data economy underpins our contemporary digital society. Unlike other technologies such as print or TV where the content could be varied, but was not itself interactive and dynamic. Nor do the ‘platforms’ of print and TV in and of themselves monitor or collect data on users. Nor do they make decisions in and of themselves about content to be delivered. Digital media and systems do work in this dynamic way and data underpins all of that activity. Understanding the role of data and information – having Data Literacy – therefore needs to sit alongside understanding how to use digital media and systems. In the same manner that UNESCO placed Media Literacy alongside Information Literacy.

Table 3: Data Literacy

Data Literacy						
Define and articulate role of data and information in a digital society and articulate own data and information needs.	Be able to locate and access data and information they need as citizens of a digital society.	Be able to critically assess the provenance, veracity and reliability data and information.	Understand what data and information digital media and systems collect and the uses they and organisations make of it.	Understand how to manage and use data and information safely, securely, legally and ethically.	Understand their rights over data and information as citizens of a digital society.	Use digital media and system skills for data and information processing, manipulation and communication to achieve a life citizens value.

3.9 Data literacy reading

For those wanting to explore Data Literacy in more depth, here is a list of recent academic research and reports. They do not offer one coherent definition and some focus very heavily on analytical data skills rather than citizenship.

- Carlson, J., & Johnston, L. (2015). *Data information literacy: Librarians, data, and the education of a new generation of researchers*. Purdue University Press.
- Civic Switchboard Guide, (2019) *Developing civic data literacy*. Retrieved from: <https://civic-switchboard.gitbook.io/guide/library-roles/developing-civic-data-literacy>
- Crusoe, D. (2016). Data Literacy defined pro populo: To read this article, please provide a little information. *The Journal of Community Informatics*, 12(3), p.27-46
- D'Ignazio, C., (2017). Creative data literacy: Bridging the gap between the data-haves and data-have nots. *Information Design Journal*, 23(1), p.6-18.
- Deahl, E. (2014.) *Better the data you know: Developing youth data literacy in schools and informal learning environments*. Available at SSRN 2445621
- Grillenberger, A. Romeike, R. (2018). Developing a theoretically founded data literacy competency model. In: Mühling, A. Cutts, Q. *Proceedings of the 13th Workshop in Primary and Secondary Computing Education*, pp. 1-10.
- Mandinach, E.B., Gummer, E.S. (2013). A systemic view of implementing data literacy in educator preparation. *Educational Researcher*. 42(1), p.30-37.
- Pangrazio, L., Selwyn, N. (2019). 'Personal data literacies': A critical literacies approach to enhancing understandings of personal digital data. *New Media & Society*. 21(2), p.419-437.
- Philip, T.M., Olivares-Pasillas, M.C., Rocha, J. (2016). Becoming racially literate about data and data-literate about race: Data visualizations in the classroom as a site of racial-ideological micro-contestations. *Cognition and Instruction*. 34(4), p.361-388.
- Vahey, P., Yarnall, L., Patton, C., Zalles, D., Swan, K. (2006). Mathematizing middle school: Results from a cross-disciplinary study of data literacy. In *Annual Meeting of the American Educational Research Association, San Francisco, CA*. p.1-15
- Williams, S. Deahl, E., Rubel, L., Lim, V. (2014). City Digits: Local Lotto: Developing Youth Data Literacy by Investigating the Lottery. *Journal of Digital Media Literacy* 1, p.1-15
- Wolff, A., Gooch, D., Montaner, J.C., Rashid, U., Kortuem, G. (2016). Creating an understanding of data literacy for a data-driven society. *The Journal of Community Informatics*. 12(3), p.9-26.

4 Why support, develop, or teach citizens' data literacy?

If you are working with this guide, we assume you are interested in supporting citizens to develop data literacy! A simple reason for doing so could be that it is a good thing in and of itself and therefore worth doing! This section explores our reasoning as to why it is important to develop democratic education or interventions to support data literacy and citizenship. The arguments made here reflect our take on this question. We hope they provide a case that others might use if they need to argue for supporting citizens to develop data literacy – be that with policy makers, with their organisation, or with communities.

A cynical position might ask why we need to teach Data Literacy? If people are happy using digital media and systems and not worried out their data – why do they need to know more? As the Americans say, “no harm – no foul”. Sadly, it is clear there is harm. Even if we leave aside criminality in the form of hacking and online scams, there have been many cases of citizens data being used in ways that may be harmful to them. The Cambridge Analytica scandal is just one such case that many people – both researchers and the people we interviewed – refer to.

Resources link: https://en.wikipedia.org/wiki/Facebook-Cambridge_Analytica_data_scandal

Our research found that Data Literacy knowledge and skills are low in the UK. We also found that many people were uncomfortable with digital media and systems using their data, tracking their activities, and sharing their data with ‘3rd parties’. Such that words such as ‘creepy’ and ‘horrid’ were used.

Of course, not everyone is in the same position, we will shortly talk through our analysis of different ‘user types’ or ‘audiences’ for Data Literacy education. This analysis indicates that those with post-18 educations and higher levels of interaction, with digital media and systems, have higher overall Data Literacy. However, even these groups can lack key knowledge or skills. So why do we need to address this gap?

We argue that citizens need a level of Data Literacy to function as active citizens on our contemporary digital society. However, our research shows that there is clearly a gap in citizens knowledge and skills that needs to be filled to address both basic awareness and to develop responses.

What happens to data we generate about ourselves, others, family, and our community – how it is used and our rights over it – are fundamental issues for contemporary society. At no point in prior history has the state, institutions or businesses had such easy access to so much information about citizens, clients, or customers. This has developed in a very short space of time – only a few decades – and we are only just coming to grips with some of the implications. These developments raise important questions for democratic societies about how to respond to this change. Especially as much of the power lies with major technology companies. They both design the technologies and conduct much of the data gathering and processing.

The GDPR legislation is often presented as an example of international cooperation to address this power imbalance and to give rights to citizens over their data. However, having data rights, understanding them, being able to exercise them and knowing how and where they apply, all depend upon understanding how the data is used in contemporary society. It requires a level of Data Literacy.

As our research has shown, even the most digitally active citizens may not have extensive or comprehensive levels of Data Literacy. Importantly, we have identified education as one of the key variables that determines data literacy. We therefore argue that developing citizens data literacy is key to supporting their active participation in healthy contemporary democratic societies.

4.1 Data literacy and citizenship

- What does it mean to be a citizen in today's 'datafied' society?
- Does it mean citizens only need to be able to send email, purchase things online and know basic management skills for their work?
- Is it the broader capacity to assess and understand how digital media and systems work, are owned, managed, and regulated?

The academic writers Isin and Ruppert argue that being a digital citizen is about the ability to make "digital rights claims". This means having enough understanding to assert political and consumer rights claims over how digital media and systems work, how they use data, as well as how they are owned, managed, and regulated.

... what makes a subject a citizen is the capacity for making rights claims [and] the citizen as subject of power comes into being through acts of making rights claims. *Conventions* are about instituting rights to govern relations between subjects and between subjects and *conventions*. By making rights claims, citizen subjects govern their relations with themselves, with others and with *conventions* (Isin and Ruppert, 2015, p.44 emphasis added)³.

Whether we view data and digital citizenship more broadly, as an ability to engage in digital life, or more specifically, as an ability to make rights claims, citizens need to have both Data and Digital Literacy as well as broader critical skills. Basic Data and Digital literacy skills might include accessing online information, contributing to online debates, or managing privacy settings. The critical skills include digital citizens' abilities to analyse, understand, and respond to the socio-economic dynamics of the digital world and their possible impact on society (e.g. digital inequalities, surveillance). Therefore, digital citizenship might be viewed

as an evolving, proactive process of social engagement, negotiations and challenges to the way civic action is enacted in a digital society.

Data is central to the digital society and, as we have argued above, Data Literacy is a converse or complimentary concept to Digital Literacy, that needs to be a key point of analytic focus. It is also central to the performance of citizenship, or the methods through which citizens are subject to the power of platforms. We would argue that there is a widening of the divide and power imbalance between **data subjects (citizens)** and **data processors (big tech)**.

Data processors, being those who own and manage data subject's data, can, as a result, capitalise on this widening divide⁴. These are largely private 'big tech', Internet service providers and the state. Technology companies hiding their data extraction behind complex, often opaque, and potential deceptively designed interfaces⁵. It is also the case that the differentiation between citizen data (data from interaction with the state and civic action) and consumer data (our patterns of consumption) is becoming blurred^{6,7,8,9}.

As we found out in our research, the sense of disempowerment around data, leads to citizens feeling confused and unable to identify, understand or respond to those who are in charge of their data^{4,10}. Our review of the literature reveals a number of studies examining the notion of the 'privacy paradox'¹⁰ whereby citizens are worried about their data online, but continue to create it nevertheless. It has been found that young people feel they have no choice but to oscillate between their desires for digital participation and online information creation, and fears related to their online privacy. Moreover, there is evidence that through implementation of methods of forced 'digital compliance' (e.g. signing terms and conditions)^{5,7}, some citizens decide to embrace the big-data divide as 'the new normal'¹¹.

³ Isin, E., Ruppert, E. (2020). *Being digital citizens*. Rowman & Littlefield Publishers.

⁴ Hintz, A., Dencik, L., Wahl-Jorgensen, K. (2017). Digital citizenship and surveillance: digital citizenship and surveillance society—introduction. *International Journal of Communication*, 11, p.731–739.

⁵ Carmi, E. (2020). *Media Distortions: Understanding the Power Behind Spam, Noise and Other Deviant Media*. New York: Peter Lang

⁶ Cheney-Lippold, J. (2017). *We Are Data: Algorithms and the Making of Our Digital Selves*. New York: New York University Press.

⁷ Barassi, V. (2019) Datafied Citizens in the Age of Coerced Digital Participation. *Sociological Research Online* 24(3), p.414-429.

⁸ Andrejevic, M. (2014) Big data, big questions: the big data divide. *International Journal of Communication* 8(17), p.1673-1689.

⁹ McCarthy, M.T. (2016). The big data divide and its consequences. *Sociology Compass*, 10(12), p.1131-1140.

¹⁰ Hargittai, E., Marwick, A. (2016). "What can I really do?" Explaining the privacy paradox with online apathy. *International Journal of Communication*. 10, p.3737–3757.

¹¹ Lin, J., Yu, W., Zhang, N., Yang, X., Zhang, H., Zhao, W., (2017). A survey on internet of things: Architecture, enabling technologies, security and privacy, and applications. *IEEE Internet of Things Journal*, 4(5), p.1125-1142.

The big data divide has an impact not only on citizens' self-awareness, but their entire web of interactions within society. Importantly, unlike common beliefs that this divide only applies to people who use the Internet, research suggests that even those who do not use the Internet or specific platforms are also subject to profiling. Data is often collected and distributed by people we know (e.g. a photo taken by a friend shared on social media) and organisations we interact with (e.g. social benefits agencies, municipality election registries). Smartphone users might have information collected when their devices or mobile applications are not in use. Citizens' data can be collected on a range of devices, which might include health apps trackers, smart-home technology or internet-of-things toys. In 2019 it was reported that 14 million users of a UK parenting platform

had their personal data collected (without their explicit consent) through sources such as websites, mobile apps, merchandise – and from the hospital bedsides of new mothers. In this way, the big technology companies who process much of our data hinder and constrain citizens' ability to exercise their rights as citizens to freedom, autonomy, agency, choice, voluntariness, privacy, and self-determination.

Yet we do not want to paint a picture of citizens of digital and 'datafied' societies as being solely defined by their data, nor absolutely constrained by the functioning of platforms or algorithms. We therefore see it as important that the conceptualisation of Data Literacy includes the capacity for citizens to question, assess, challenge, and make rights claims within a 'datafied' society.

4.2 Democratic education and Data Citizenship

We argue that digital and data literacy needing to be more than "basic digital skills". There is a need to combine these with broader critical thinking and knowledge about the digital ecosystems in which citizens are now effectively forced to operate. We have therefore brought the ideas of "democratic education"^{12,13}, into our model. Democratic education aims to equip citizens with skills and knowledge which enable them to exercise their citizenship. Democratic educators' objective is to empower citizens to critically examine their positions within the existing power structures in society, and to develop skills and understanding to take an active stand in the process of their individual and collective self-determination.

The parallels between teaching critical consciousness and data literacies education have been outlined by Tygel and Kirsch¹⁴ who proposed that data-centred democratic education should aim to enhance citizen's critical comprehension of data realities and encourage them to question the existing data-society norms. To achieve critical consciousness in the context of Data Literacy, Wolff and colleagues argue that "[data] learning experiences should be responsive to cultural differences that might affect an individual learner's view of the world"¹⁴.

This guide therefore seeks to provide you with ideas and resources focused on this issue of **democratic education for data citizenship**. The goal is to provide advice on how to support citizens to think critically and carefully about what currently happens to their data and the data produced by their community but also to think critically about what "should" happen to that data, what they think is right and just regarding the use and sharing of data in contemporary society. Therefore, it does not focus on basic digital and data skills, nor on complex data analytic skills, but rather on the key issues of data thinking and data participation that other training interventions don't address.

¹² Dewey, J. (1930). *Individualism, Old and New. II: The Lost Individual*. *New Republic*, 61, 294-296.

¹³ Freire, P. (1970/1996). *Pedagogy of the oppressed (revised)*. New York: Continuum

¹⁴ Tygel, A., & Kirsch, R. (2015). Contributions of Paulo Freire for a critical data literacy. In *Proceedings of Web Science 2015 Workshop on Data Literacy* (pp. 318-34).

4.3 Comprehensive list of everyday data literacy activities

If you explore the Data Literacy literature provided in section 3, you will see that the main issues discussed are very practical or concerned with data analysis. For example: practical data handling and management; analytic skills; and critical thinking skills. Examples of data management skills include identifying, collecting, and storing data. Examples of analytic skills include data analysis, interpretation, and visualisation.

Our definition of Data Literacy is much broader and includes a focus on critical thinking that cuts across both data management and analytic skills. Looking at our definition in Table 3 and the ideas in these studies, we have broken Data Literacy down into specific activities (see Table 4) and grouped these into three everyday activities:

- **“Data Doing”** covers many of the practical data management and analytic aspects of data literacy.

- **“Data Thinking”** relates to both practical and critical use of data to problem solve and to communicate with data.
- **“Data Participation”** covers activities that involve an active response to being a citizen in a ‘datafied’ society. In particular, the critical assessment of the use of data, skills to assess, resist, and undertake activism to change and negotiate both technologies and systems of power in a ‘datafied’ society. Importantly, we believe that this must include working with others, groups and communities and may involve the collection, repurposing and use of data in such activism.

A comprehensive approach to Data Literacy education and training would cover all of these activities. Our focus on democratic education sets up a requirement to assess these practical activities in relation to issues of politics, power, and social context. It is this critical awareness, critical thinking and active citizenship element that we will focus on in this guide.

Table 4: Elements of data literacy and citizenship

Aspects	Data Doing	Data Thinking	Data Participation	Covered in the example activities
Accessing	X			
Assessing	X			
Interpretation	X			X
Data creation	X			
Data citation	X			
Data Management	X			
Ethical use	X			X
Data Deletion	X			X
Data Visualization and Manipulation	X	X		
Understanding of data collection	X	X		X
Problem-solving using data		X		
Communicating with data		X		
Critical data analysis (e.g. data bias, cultural contexts)		X		X
Data safety (e.g. skills to manage and control ‘digital traces’)		X		X
Understanding privacy		X		X
Awareness of data protection rights		X	X	X
Understanding data society		X	X	X
Participating in society using data			X	X
Engagement with data society debates			X	X
Data Activism			X	X
Supporting others with their data literacy			X	X

4.3.1 What this guide does not cover

This guide does not cover practical data skills. This does not mean that we think these practical data skills should be ignored, rather, that we believe there are plenty of educational resources and guides out there to support these practical elements. As we noted before, for basic digital skills, we would recommend the materials from the Good Things Foundation:

- <https://www.goodthingsfoundation.org/learn/learn-my-way/>
- <https://www.goodthingsfoundation.org/learn/make-it-click/>

Table 4 also includes such things as data analysis and interpretation as well as data visualisation and communicating with data. These are skills and activities that can include the basic use of tools:

- How to use social media
- How to use communication tools such as email
- How to use office applications such as word processors and spread sheets
- How to use cloud services and back-up drives

At the other extreme it can also include the use of complex data analysis and visualisation tools:

- How to develop or manage complex databases
- How to do data wrangling and statistics with things like R or Python
- How to use complex visualisation tools like Tableau or R

The internet is replete with courses, web sites and YouTube videos that teach data analysis skills at all levels. If teaching such skills is important, you should look to select such online resources that are appropriate for the people that you are working with.

For useful courses on data analytics that start from a very basic level we would recommend the free materials provided by:

- <https://schoolofdata.org/courses/>

For an example of a web site that covers the full range of data analytics – but assumes a high level of knowledge and experience, see:

- <http://www.sthda.com/english/>

For an example of a web site that covers the full range of data visualisation – but assumes a high level of knowledge and experience, see:

- <https://r-graph-gallery.com/index.html>

PART 2: PLANNING YOUR APPROACH

5 Seven principles for developing citizens Data Literacy

From our research, we developed seven principles for developing interventions to support citizens Data Literacy.

Any educational intervention, awareness raising, or community engagement must:

- 1. Ensure citizens feel more empowered** and have practical and alternative routes to enact that empowerment.
 - o Don't just make people scared of using digital media and systems – help them feel empowered to deal with them.
- 2. Consider the design and practical challenges citizens face** in managing and controlling the data they share or “give off” whilst also being actively involved with others via the plethora of platforms in our digital society.
 - o Citizens can't just “stop using Facebook” – help them manage the risks and benefits.
- 3. Make clear to citizens their rights - as citizens not just consumers** - to make claims in regard to data use, sharing and trading and also of digital systems and platforms.
 - o Remind them that they have rights under GDPR and that they can complain, flag up social media content, disable adverts on systems, etc.
- 4. 'Meet citizens where they are'** in terms of their digital and social experience and context.
 - o Know your audience
- 5. Address the challenge that those adults most in need of support are very likely outside formal educational settings.**
 - o People in education or who have had post-18 education are the least in need. Though everyone needs to improve their Data Literacy.
- 6. Support skills development but must be more than skills,** encompassing key elements of Data Participation.
 - o Don't just focus on skills – in fact there is lots of help out there for skills – it is the critical awareness and proactive citizenship that are missing in most training and support.
- 7. Seek to provide deep critical consciousness the power relationships in our 'datafied' society** and support them to exercise their right to challenge this imbalance and demand change.
 - o The goal has to be empowered digital citizens with good Data Literacy

6 Audiences and resources

We would argue that there are two factors to consider in putting together an intervention to support Data Literacy. These are your audience and appropriate resources. There is little point providing an academic reading list to people who have low levels of digital skills. Similarly, there is little point undertaking an activity on 'what is data' with a set of computing experts. That said, we would caution not to assume things about your audience. There tends to be an assumption that "digital natives" know everything, whereas in our research, we find many young people with low Data Literacy. We also found older people, with low digital skills, applying good Media Literacy ideas about bias to social media. Social context is also important. Younger people tend to use a different mix of social media than older people. Some communities or even family groups prefer one medium over another. For example, one group might use a closed system like WhatsApp, whereas others may use more open platforms like Facebook. We also find that individuals and groups have 'go to people' for help with digital and these 'go to people' maybe key to driving change in a community's approach to digital. We will explore further how to assess and understand your audience in a moment.

In terms of resources, you need to think about the types of content your audience will engage with, in regard to both depth and relevance. Complex arguments about the legal use of scraped social media data may be of interest to activists in a protest group, but not people just getting to grips with using a smart phone. A lot of the available resources – many examples of which we provide in Part 3 – are targeted at those interested in Data Literacy issues. These resources might be journalism, academic research, or the outputs of activist groups. There is also a very large array of materials available to support teaching basic digital skills, basic data security and basic privacy management. We are not going to re-invent such things here, and as noted in Part 1, we would recommend the Learn My Way materials from the Good Things Foundation as a starting point.

We believe that what is missing are developed approaches to move people from basic digital and data skills, through to the broader issues of digital and data justice, and citizenship. These are mostly discussed in academic and some media circles.

6.1 Audience

So how to assess and understand your audience? In our Me and My Big Data project we developed a set of 'personas' for different types of 'users' of digital media and systems. For details on how we developed these personas, see:

- <https://www.liverpool.ac.uk/media/livacuk/humanitiesampsocialsciences/meandmy-biddata/Understanding,Citizens,Data,Literacies,Research,,Report,Final.pdf>

We provide descriptions of each of these at the end of this Part. We would hope that these can work well as starting points for you to reflect on the circumstances of the people you are looking to work with or provide training for. These are very broad-brush descriptions and there are of course many social, cultural, and personal variations among these groups.

If you have time, we suggest working with your intended audience to collect some information and insight on them through individual or group conversations or potentially some more quantitative information through short surveys. The key questions about your audience include:

- What are their levels of digital skill and competence?
- What are the main platforms they use?
- How actively do they use different platforms?
- Do they use digital platforms in their work or study as well as everyday life?
- Do they already have a good knowledge of how data are used by platforms?
- Do they have good broadcast media literacy?
- Who do they rely on for help with platforms and data?
- What concerns do they have about data and how platforms use it?
- Are they engaged in digital or data activism or participation?

Getting answers to these kinds of questions will allow you to better select resources and plan activities for your participants.

6.2 Example personas

The following personas describe different groups **based on their use of digital media and systems**. To be clear they are not based on such things as their social background or their employment. Though in each case we comment on how key features such as age, education, social class, and levels Data Literacy map on to (correspond) with being in each group. These are not meant to be in any sense 'absolute' and definitive, rather they are a tool for exploring different groups of digital users when thinking about Digital and Data Literacy. The percentages reported are from our ME and My Big Data survey.

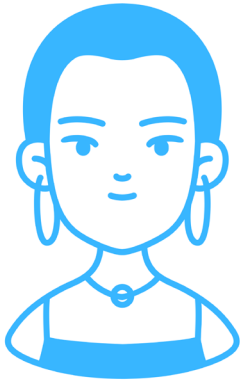
Data Citizenship personas



Extensive political users

- 64% are under 45 years old
- Most likely to be in NRS social grades AB and C1
- Likely to be in employment (not retired)
- Very likely to have post 16 education
- Highest Data Citizenship/data literacy scores alongside 'Non-political extensive users'
- Just behind 'Non-political extensive users' in levels of Data Participation

- Much higher-than-average data literacy scores
- Average trust in information from friends and found on social media
- 80% do some checks of social media content
- 84% do some checks of search engine and online content
- 94% had some awareness of data collected by platforms
 - On average they identified 7 out of a possible 11 types of data that they may be overtly sharing with or passively "giving off" to platforms
- 98% had some awareness of reasons for data collection by platforms
 - On average they identified 5 out of a possible 8 reasons why platforms may collect data
- 2nd happiest group with data collection to deliver consumer benefit (36% to 50%)
- 66% are uncomfortable with 3rd party sharing of personal data
- Despite being some of our most active users 32% feel platforms make changing privacy setting 'too much effort'
- They are split 42% vs 41% over whether there is any point changing settings on platforms
- Confident in 4 out 8 data protection activities
- Above average dependence on social networks as routes for information
- Much higher than average levels of Data Participation



Non-political extensive political users

- 715 are under 45 years old
- Most likely to be in NRS social grades AB and C1
- Likely to be in employment (not retired)
- Very likely to have post 16 education
- Most likely to have a higher university degree
- Highest Data Citizenship/data literacy scores alongside 'Extensive political users'
- Just ahead of 'Extensive political users' in levels of Data Participation

- Much higher-than-average data literacy scores
- Average trust in information from friends and found on social media
- 84% do some checks of social media content
- 80% do some checks of search engine and online content
- 97% had some awareness of data collected by platforms
 - On average they identified 8 out of a possible 11 types of data that they may be overtly sharing with or passively "giving off" too platforms
- 98% had some awareness of reasons for data collection by platforms
 - On average they identified 6 out of a possible 8 reasons why platforms may collect data
- Happiest group with data collection to deliver consumer benefit (48% to 59%)
- 63% are uncomfortable with 3rd party sharing of personal data
- Majority (60%) did not feel platforms make changing privacy setting 'too much effort'
- Majority (52%) felt it was worthwhile changing settings on platforms
- Confident in 5 out 8 data protection activities
- Above average dependence on social networks as routes for information
- Much higher than average levels of Data Participation



Non-political extensive political users

- 70% are under 45 years old
- Most likely to be in NRS social grades C1, C2 and DE
- Unlikely to be retired
- Very unlikely to have post 16 education
- 2nd lowest Data Citizenship/data literacy scores
- 2nd lowest Data Participation scores

- Lower-than-average data literacy scores
- Above average and highest overall levels of trust in information from friends and found on social media
- 62% do some checks of social media content – but average range of checks (one type) is very low
- 61% do some checks of search engine and online content – but average range of checks (one type) is very low
- 85% had some awareness of data collected by platforms
 - Though on average they only identified 4 out of a possible 11 types of data that they may be overtly sharing with or passively "giving off" too platforms
- 88% had some awareness of reasons for data collection by platforms
 - Though on average they identified 3 out of a possible 8 reasons why platforms may collect data
- 3rd happiest group with data collection to deliver consumer benefit (37% to 44%)
- 61% are uncomfortable with 3rd party sharing of personal data
- They were split 41% vs 48% on whether platforms make changing privacy setting 'too much effort'
- They were split 40% vs 42% if it was worthwhile changing settings on platforms
- Only confident in 3 out 8 data protection activities
- Above average and highest overall levels of "mostly reading social media that shares their own values"
- Above average dependence on social networks as routes for information
- Below average and second lowest levels of Data Participation



General users (limited social media)

- 70% are over 24 years old
- Most likely to be in NRS social grades AB, C1, and C2
- Unlikely to be retired
- Evenly split between having and not having a post 16 education
- Lower Data Citizenship/Data literacy scores
- Mixed set of data behaviours – often just one or two activities across the range available

- Just above-average data literacy scores
- Average trust in information from friends and found on social media
- 62% do some checks of social media content – but average range of checks (one type) is very low
- 61% do some checks of search engine and online content – but average range of checks (one type) is very low
- 94% had some awareness of data collected by platforms
 - On average they identified 7 out of a possible 11 types of data that they may be overtly sharing with or passively “giving off” too platforms
- 96% had some awareness of reasons for data collection by platforms
 - On average they identified 5 out of a possible 8 reasons why platforms may collect data
- 4th out 5 in happiness group with data collection to deliver consumer benefit (28% to 37%)
- Group most uncomfortable with 3rd party sharing of personal data (74%)
- A majority (55%) did not think platforms make changing privacy setting ‘too much effort’
- A majority (48%) thought it was worthwhile changing settings on platforms
- Only confident in 3 out 8 data protection activities
- Average dependence on social networks as routes for information
- Average levels of Data Participation



Limited users

- 69% are over 45 years old
- Most likely to be in NRS social grades C1, C2 and DE
- 36% over retirement age
- Very unlikely to have post-16 education, most likely to have no qualifications
- Lowest Data Citizenship/data literacy scores
- Lowest Data Participation scores

- Lowest data literacy scores
- Above average lack of trust in information from friends and found on social media
- Only 22% do some checks of social media content – but most do none or don’t use social media
- Only 45% do some checks of search engine and online content – but mostly only one type of check
 - On average they only identified 2 out of a possible 11 types of data that they may be overtly sharing with or passively “giving off” too platforms
- 96% had some awareness of reasons for data collection by platforms
 - On average they only identified 2 out of a possible 8 reasons why platforms may collect data
- 4th out 5 in happiness group with data collection to deliver consumer benefit (28% to 37%)
- Group second most uncomfortable with 3rd party sharing of personal data (71%)
- They area split 37% to 44% over whether platforms make changing privacy setting ‘too much effort’
- A majority (45%) do not think it was worthwhile changing settings on platforms
- Only confident in 1 out 8 data protection activities
- Lowest dependence on social networks as routes for information
- Lowest levels of Data Participation

7 Where to find resources

Digital technologies, the platforms we use, and the ways we use them, are changing all the time. As are the concerns that citizens have about their use of these technologies. Our relationships with them and through them, to friends and family, are also changing. Choosing the right resources for your target audience therefore needs to balance their level of knowledge, topics of concern and the goals of your educational or training activity. It is clear from our research that short video content, especially YouTube, is one of the key 'go to' sources of help and support for all types of users. For those with higher levels of digital and data literacy, media content such as news and documentaries and even films were drawn on to help understand platforms and the data economy.

There are three types of resources our respondents, other academic colleagues and we have found during our research:

- Journalism and media coverage
- The work of Activist groups
- Academic research and findings

Very often these three groups work together.

7.1 Journalistic, documentary, and public media

These often provide quite digestible discussions of the issues of digital and data literacy. They range from YouTube videos to Ted Talks and from full blown documentaries to short media reports. Though some of these easily stand on their own as items worth watching, nearly all will need an appropriate educational or training 'wrapper' around them. This might consist of briefing notes on key issues, follow up discussions or self-assessed questions.

Such resources will of course constantly increase as new issues, new journalism and new media content develops. We should suggest that anyone working to deliver digital and data literacy development and training should maintain a list of such useful content. Some current and recent examples of currently available Journalistic, documentary, and public media include:

7.1.1 Films and documentaries

- The Social Network: https://en.wikipedia.org/wiki/The_Social_Network
- The Great Hack: https://en.wikipedia.org/wiki/The_Great_Hack
- The Power of Privacy: <https://www.youtube.com/watch?v=KGX-c5BJNFk>
- Terms and Conditions May Apply: https://en.wikipedia.org/wiki/Terms_and_Conditions_May_Apply
- Do not track: <https://donottrack-doc.com/en/intro/>

7.1.2 Media content

- A BBC film about a data day: <https://www.bbc.co.uk/news/av/technology-55907736>
- A TED talk on "What will a future without secrets look like?": https://www.ted.com/talks/alessandro_acquisti_why_privacy_matters
- A keynote talk on "Privacy and big data": <https://www.youtube.com/watch?v=oWwQfgpvlzI>
- A BBC article on "How market research reveals what you really think": <https://www.bbc.co.uk/news/business-57399780>
- A video from the 'The Economist' on "How internet advertisers read your mind": <https://www.youtube.com/watch?v=8KYugpMDXAE>

7.2 Activists and key organisations web sites

Many major 3rd sector organisations, governments and activist groups maintain websites that seek to explain issues, provide training and development, or contain their own substantial education resources.

Ofcom and DCMS are the UK media regulators and have obligations to develop citizens Digital and Media Literacies. UNESCO takes a similar role for the United Nations. They all have relevant web resources:

- Ofcom: <https://www.ofcom.org.uk/research-and-data/media-literacy-research>
- DCMS – strategy: <https://www.gov.uk/government/publications/online-media-literacy-strategy>
- DCMS – skills: <https://www.gov.uk/government/publications/essential-digital-skills-framework/essential-digital-skills-framework>
- UNESCO: <https://www.unesco.org/en/communication-information/media-information-literacy>

For an example of libraries taking a lead on Data Literacy see:

- <https://civic-switchboard.gitbook.io/guide/>

This also includes lots of links to basic data skills and other resource tool kits.

Some key 3rd sector groups seeking to support digital skills and engagement include:

- For adults and other 3rd sector groups: <https://www.goodthingsfoundation.org>
- For children and parents: <https://www.internetmatters.org>
- For rights and regulations: <https://doteveryone.org.uk>
- For inclusion and young people: <https://www.nominet.uk/social-impact/>

Some key activist organisations include:

- For secure apps and privacy activism: <https://guardianproject.info>
- For data and democracy activism: <https://privacyinternational.org/>
- For citizen and civil society data activism: <https://tacticaltech.org/>
- For advancing data literacy in civil society: <https://schoolofdata.org>

7.3 Academic and investigative journalism books

There is a great deal of academic, activist and long form journalism that covers issues of digital and data literacy. We have listed below a set of books that we think are either key text or provide an accessible route into key issues (alphabetic order):

- Arthur, C. (2021). *Social Warming: The Dangerous and Polarising Effects of Social Media*. Simon and Schuster.h
- Athique, A. (2013). *Digital media and society: An introduction*. John Wiley & Sons.
- Baym, N. K. (2015). *Personal connections in the digital age*. John Wiley & Sons.
- Carmi, E. (2020). *Media distortions: Understanding the power behind spam, noise, and other deviant media*. Peter Lang International Academic Publishers.
- Cheney-Lippold, J. (2017). *We are data*. New York University Press.
- D'ignazio, C., & Klein, L. F. (2020). *Data feminism*. MIT press.
- Doss, A. F. (2020). *Cyber Privacy: Who Has Your Data and Why You Should Care*. BenBella Books.
- Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.
- Highfield, T. (2017). *Social media and everyday politics*. John Wiley & Sons.
- Isin, E., & Ruppert, E. (2020). *Being digital citizens*. Rowman & Littlefield Publishers.
- Kearns, M., & Roth, A. (2019). *The ethical algorithm: The science of socially aware algorithm design*. Oxford University Press.
- Kennedy, H. (2016). *Post, mine, repeat: Social media data mining becomes ordinary*. London: Palgrave Macmillan.
- Lupton, D. (2016). *The quantified self*. John Wiley & Sons.
- Lupton, D. (2020). *Data Selves*. Polity.
- Lyon, D. (2018). *The culture of surveillance: Watching as a way of life*. John Wiley & Sons.
- Rohlinger, D. A. (2019). *New media and society*. New York University Press.
- Véliz, C. (2021). *Privacy is power*. Melville House.
- Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. Profile books.

PART 3: SUGGESTED ACTIVITIES

8 What is data?

One of the first challenges we faced when working with citizens was that of articulating what data is. We found that limited or narrow users (see our Personas in Part 2: Section 6) were the least able to talk about the nature and types of data they were sharing or giving away when online. They also clearly did not actively think about these issues – unless a specific fear (surveillance) or a specific incident (scam/hack) had raised their awareness. One of the first activities we did in our focus groups was therefore to explore “What is data”.

This is a classic example of knowing about and not making assumptions about your audiences. We as researchers who had engaged with much academic and policy work were very happy to talk about ‘data’, ‘sharing of data’ or ‘data protection’ and so forth. In our focus groups only the more experienced users talked in this way and, importantly, could articulate clearly what they meant. Other respondents were more comfortable talking about ‘information’ or specific items of data/information – such as “my bank details”.

We would therefore argue that supporting those with lower digital and data literacy to understand the nature of data – and how they end up sharing it has to be a key part of basic Data Literacy

8.1 Suggested activity 1: What is data?

Level: mainly for participants with low digital skills or awareness – but a good activity to do for all levels as a starting point.

Learning outcome: An understanding of what data is and what information platforms collect from citizens. This is core to developing an ‘Understanding of data collection’ (Table 4).

As a starting activity – especially for people with lower digital skills – we suggest having an open discussion about the nature of data and information in contemporary everyday life. Exploring this can help educators explore citizens levels of understanding. Through the discussion a shared understanding can be developed that can underpin later activities. Questions that can guide this discussion include:

- If I use the term ‘data’ what does this mean for you?
- If I use the term ‘information’ what does this mean for you?
- What data/information do you think Facebook holds about you? Though you might want to select a platform that is appropriate to your audience (TikTok, Twitter, Google, Amazon etc.).

Alternatively, you can ask several versions of this question to see if people perceive the data held by platforms differently.

- What data/information do you access or use at work?
- What data/information would you not want people or platforms to share about you?

In our research we found that some people struggled to understand what ‘data’ meant. In exploring these issues, you may find that you need to explain how platforms such as Facebook or Google collect and use data. It can be helpful to move directly into Activity 2 below on order to make ideas about data concrete and linked to everyday activity.

‘Data’ is a term that is used by both experts and lay people alike but often they are not talking about the same thing. Similarly, the term ‘information’ can have very technical to very broad meanings. Some ‘definitions’ of ‘data’ and ‘information’ we found in our focus groups included:

- Data as your “smart phone data allowance” – a commodity like water or electricity that can be limited in volume.

- Data as “lots of numerical information” – like data in science, abstract stuff to be analysed, stuff in spreadsheets.
 - Data as “data on me” or Information as “information on me” – data and information collected from or surveillance of citizens.
 - Information as “curated data” – information is somehow more formal and organised than data.
- Very often both academics and lay people use the terms information and data interchangeably.

8.2 Suggested activity 2: A data day

Level: this activity is designed for people with any level of digital skills.

Learning outcomes: An understanding of what data we “give off” as we go through our daily lives. An understanding of how and why platforms and companies collect and use this data. This basic understanding is needed to help with ‘Data interpretation’, ‘Ethical use’, ‘Data deletion’, and ‘Data safety’, it is core to developing an ‘Understanding of data collection’ (Table 4).

Table 5: A data day

Time of day	Device and platform	Data given to or collected by device or platform / Date taken by user.
Breakfast		
Morning		
Lunch		
Afternoon		
Dinner		
Evening		

Ask people to complete the matrix for a ‘typical day’. They can fill this in individually or in small groups. For a more comprehensive activity you could do two sheets, one for a weekday/workday and one of a weekend/holiday day.

First, ask your participants to list the digital media and systems they likely interact with at different times of the day. Remind them that all devices (smart phones, game devices, Alexa like devices, smart watches, tablets, laptops etc.) and all systems (social media, email, websites, apps, etc.) count.

Second, ask your participants to list:

- The data they overtly give to these systems – e.g. what data they input for Amazon to make a purchase.
- The data that these systems might collect about them? – e.g. what they searched for on Amazon or Google.
- What data they took from the systems – e.g. web search results, time of the next train, rating for a product they want to buy.

The goal is to get people to explore the data economy that they live in. In particular it looks to explore the extent to which they ‘give off’ data through everyday interaction with digital media and systems.

We suggest you provide people with a simple A4 or A3 matrix split into day segments, systems used, and data given off or data interacted with (see Table 5 as an example).

Be mindful of the fact that many people may not know or will not bring to mind data that may be collected ‘passively’. We found that some people forgot or did not realise that mobile devices track location or that web sites may share data via cookies.

The table that participants have created can be used in and of itself to support discussion and reflection. It can also be used to underpin further activities including all those discussed in the following sections. Some suggestions include:

- Creating a visualisation of a ‘a data day’ by getting participants to present their day in a graphical format or poster.
- Participants could overlay their ‘data day’ on a map or even on a diagram of their social network of friends and family.
- Identifying the key platforms participants use and exploring in greater depth the technologies, the companies that own them and their history.

8.2.1 Resources

Three good examples – with links to further resources – on ‘a data day’:

- Tactical Tech ‘A data day’: <https://cdn.ttc.io/s/tacticaltech.org/Data-Day-updated-poster.pdf>
- Tactical Tech ‘A data day in London’: <https://cdn.ttc.io/s/tacticaltech.org/a-data-day-london.pdf>
- A BBC film about a data day: <https://www.bbc.co.uk/news/av/technology-55907736>

A similar guide is provided by Apple – though clearly designed to foreground privacy features of Apple products:

- Apple data day: https://www.apple.com/uk/privacy/docs/GBEN_Privacy_ADITL.pdf

Simple outlines of what data different platforms collect about people:

- ‘Visualcapitalist’ provides an neat graphic covering data collected: <https://www.visualcapitalist.com/heres-what-the-big-tech-companies-know-about-you/?fbclid=IwAR3OO2Qlo40NZB7uhTvlrriiOpc-lxa7rNMNzzGBte5H7XappG-NnmFw1M>
- This is based on research by security.org: <https://www.security.org/resources/data-tech-companies-have/>

9 Understanding data collection

An understanding of data collection is key to citizens basic comprehension of the data economy and society in which they live. It is often very difficult to begin this conversation with citizens who have limited experience of digital media and systems or are starting from a very limited understanding of how they work. For such participants, we would suggest that our “What is data” activities are the best starting point before moving to this topic.

Our research found that most respondents had an awareness of, but lacked a detailed understand of, how their data were being extracted, tracked, how the underlying technologies work nor how the economics of digital platforms and systems.

We would therefore argue that supporting all citizens to have a good grasp of what platforms do with their data is key to basic Data Literacy.

9.1 Suggested activity 3: How does my favourite platform collect data on me?

Level: this activity is designed for people with any level of digital skills.

Learning outcomes: An understanding of data collection and data protection. This is core to developing an ‘Understanding of data collection’ and ‘Data safety’, and practical aspects of ‘Understanding privacy’ and ‘Understanding data society’ (Table 4).

Following on from Activity 2 above, we suggest picking a major platform that your participants use and exploring its data collection activity in more depth. Facebook is of course one of the most used platforms but also has also been heavily criticised in the media for its data extraction and data use activities. Two activities that could be undertaken include exploring Facebook Privacy settings and advertising settings.

We would recommend that participants do this individually or privately – especially if they have not done this before – as they may not want to share with others what they find. You should also be prepared for this to cause concern for some people and have at hand advice and support on improving data privacy.

1. Exploring Facebook Privacy settings – get your participants to look at their privacy settings on Facebook. Explore what each of these does and what it implies about the data they are sharing with Facebook. Facebook itself provides advice on how to use its privacy settings:
 - https://www.facebook.com/help/238318146535333/?helpref=uf_share

2. Look at Facebook advertising settings to explore what Facebook thinks participants are interested in from processing the data they have shared. Explore with participants how the data Facebook has collected about them by monitoring likes and posts might have led to these advertising preferences

1. https://www.facebook.com/adpreferences/ad_settings

9.1.1 Resources

- A BBC article on “How market research reveals what you really think”: <https://www.bbc.co.uk/news/business-57399780>
- A video from the ‘The Economist’ on “How internet advertisers read your mind”: <https://www.youtube.com/watch?v=8KYugpMDXAE>

10 Keeping safe online

Keeping both themselves and their data safe on-line is important for all citizens. Though we are concerned here mainly with the ‘legal’ use of data worries about loss of data and hacking can be stop people using digital media and systems.

Our research found that respondents maintained a distinction between “legal” activity – even if they are uncomfortable with it or see it as “creepy” – and overtly illegal behaviour in the form of “hacking”. Even though the personal or material consequences could be comparable – such as public disclosure of personal data or financial loss, political manipulation or exposure to mis-/dis-/mal-information.

Personal and data security and safety are a key part of both digital and data literacy.

This is a different type of activity and overlaps with some of the basic digital skills training referred to in section 1.

10.1 Suggested activity 4: Data safety

Level: this is appropriate to all people with some digital skills

Learning outcomes: a good understanding of how to keep data safe. This understanding is needed to help with ‘Ethical use’, ‘Data deletion’, and ‘Data safety’, it is core to developing an ‘Understanding of data collection’, it underpins ‘Awareness of data protection rights’ (Table 4).

Data safety is a core part of safe and secure internet use. This has been covered by many organisations in the UK and globally. We would suggest tapping into some of the high-quality free training available. If you are working in a large organisation, you likely have access to good data and digital security training. For others we would recommend using:

- Good Things Foundation – Learn My Way courses provide a great basic grounding in digital safety: <https://www.learnmyway.com/courses/keeping-your-personal-data-safe/>
- Future Learn – A gentle introduction to key data privacy topics: <https://www.futurelearn.com/info/blog/how-to-protect-your-data>

If working with young people and children, Internet Matters provide links to a wealth of materials and lesson plans:

- <https://www.internetmatters.org/resources/>

11 Understanding online privacy

An understanding of online privacy is a key component of data thinking. This is more than keeping data safe, privacy settings or cookies – this is covered by Data Safety above. This is a broader understanding of privacy in our digital age. What we found notable about our results was the language used to describe how platforms use of data affected people’s feelings of privacy.

In our research citizens used terms such as “creepy”, “horrible”, and “scary” to describe data being collected by platforms. All of these are words with connotations of fear and particularly of unwanted surveillance and a pervasive feeling of ‘wrongness’. These are very emotive responses. They make clear that citizens are not happy with what researchers have called the “privacy paradox” – where we give up privacy for free access to digital services like Facebook. This ‘new normal’ or having to share information about yourself to get a services leaves citizens feeling unsafe and at risk.

It is not possible to address such concerns just through training or education. This reflects the divide and power imbalance we noted earlier between **data subjects (citizens)** and **data processors (big tech)**.

We would also note that exploring these issues further could generate greater anxiety for some people. Knowing your audience will help assess how to approach this issue appropriately. We see two approaches to this issue for educators:

- Providing tools, techniques, and skills to improve online privacy.
- Exploring the issue in depth and how social groups, activists and policy makers are seeking to improve privacy online.

11.1 Suggested activity 5: Planning for online privacy

Level: this activity is designed for people with good digital skills and who use a variety of digital platforms – likely our general to extensive users (see personas in Part 2: Section 6).

Learning outcomes: An understanding of data privacy and how to practically achieve it. This is core to developing an ‘Understanding of data collection’ and ‘Data safety’, ‘Understanding privacy’ and ‘Understanding data society’ (Table 4).

Ask your participants to work individually or in a group through the materials on the Electronic Frontier Foundation “Surveillance Self-Defense” website:

- <https://ssd.eff.org>

Once they have read through the site suggest that they select one of the security scenarios that is closest to them or of interest to them:

- <https://ssd.eff.org/en/module-categories/security-scenarios>

Ask your participants to work, individually or in a group, through each of the tasks for their selected security scenario. Ensure that they develop a ‘Security Plan’. Once they have completed these tasks get the participants to discuss:

- What practical things did they learn about data security?
- Would implementing the security plan increase their feelings of safety and security online? Would they feel their privacy was better protected?
- Did they feel the balance of work needed was worthwhile when set against the improvements in their online privacy?

11.1.1 Resources

- Electronic Frontier Foundation – “Surveillance Self-Defense” is a broad as well as in-depth assessment of both online privacy and how to use tools to manage this: <https://ssd.eff.org>

11.2 Suggested activity 6: Privacy in a digital society

Level: this activity is designed for people undertaking A-level or University study or those who might be part of a social action group

Learning outcomes: A broad social, cultural, and political understanding of data privacy and data protection. This is core to developing an 'Understanding of data collection' and 'Data safety', 'Understanding privacy' and 'Understanding data society' (Table 4).

Ask your participants to work, individually or in a group, through the materials on the Privacy international "Invisible Manipulation" website:

- <https://www.privacyinternational.org/long-read/1064/invisible-manipulation-10-ways-our-data-being-used-against-us>

Ask your participants, individually or in groups, to do the following four tasks

- Select one of the 10 subject areas Privacy International identify and work through the material on the site.
- Gather further information on the topic from both academic publications and research and from media and journalism sources.
- Use these materials to review and assess the Privacy International web content.
- Prepare a poster or pamphlet that explains these issues for a 'lay' or 'non-technical' audience.

11.2.1 Resources

- A TED talk on "What will a future without secrets look like?": https://www.ted.com/talks/alessandro_acquisti_why_privacy_matters
- A keynote talk on "Privacy and big data": <https://www.youtube.com/watch?v=oWwQfgpvlzI>

12 Data participation and data citizenship

In our research we found that:

Respondents did not translate their unease and disquiet about platforms use of their data into sustained action to protect their data nor to assert digital rights. UK respondents also attributed some of their feelings of disempowerment or lack of action to the difficulties of navigating digital systems and media and the practices of the companies that run them. But our respondents were also just as likely to blame their own "failings", lack of skill or lack of personal "vigilance" for the situation.

We would argue that people viewing these things as 'personal failings' prevents them from seeing them as things to be socially or politically challenged. We also found that many people were not aware of their data rights under GDPR or more broadly. Though many people mentioned Cambridge Analytica as an example of the misuse of data, only a few had

a reasonable knowledge of the details. Finally, we found that few survey respondents scored highly in terms of Data Participation and few if any focus groups respondents had used data to help their communities. The suggested activities below are designed to address these three issues.

12.1 Suggested activity 5: Planning for online privacy

Level: this activity is designed for people undertaking A-level or University study or those who might be part of a social action group.

Learning outcomes: An understanding of the Cambridge Analytica scandal, the role of algorithms and critical data analysis, as well as the social, cultural and political aspects of data collection and data protection. This is core to developing an 'Understanding of data collection', 'Data safety', 'Data ethics', 'Understanding privacy', 'Engaging with data society debates' and 'Understanding data society' (Table 4).

This activity is designed to get you participants to explore the wider issues of living in a 'datafied' society. This is a longer activity that might take several sessions or work between sessions.

Get your participants working in a group to explore the detail of the Cambridge Analytica scandal. The Wikipedia page is a good starting point:

- https://en.wikipedia.org/wiki/Facebook-Cambridge_Analytica_data_scandal

Ask them to undertake the following:

- Gather as much media coverage of the scandal as they can from the web or library archives if they have access.
- Explore the topics and issues discussed in the press coverage. How was the scandal framed? What were the key concerns? Loss of privacy? Political manipulation? Ethics?
- Gather as much academic discussions and research on the scandal as they can from the web or library archives if they have access.
- Explore the topics and issues discussed in the press coverage. How was the scandal framed? What were the key concerns? Loss of privacy? Political manipulation? Ethics?
- Watch "The Great Hack": https://en.wikipedia.org/wiki/The_Great_Hack
- Explore the topics and issues discussed in the documentary. How was the scandal framed? What were the key concerns? Loss of privacy? Political manipulation? Ethics?

Finally, ask your participants to create a group presentation or activity that would explain the scandal to a non-technical public audience.

12.2 Suggested activity 8: Exploring your data rights

Level: this is appropriate to all people with some digital skills

Learning outcomes: An understanding of citizens data rights in the UK. This is core to developing 'Data safety', 'Understanding privacy', 'Awareness of data protection rights' and 'Understanding data society' (Table 4).

This activity can stand alone but also would work following on from Activity 7. It is designed to educate people about their rights as UK digital citizens in regard to how their data are used. You will need to consider carefully your participants levels of knowledge and potentially provide some additional support materials and guidance.

This may take more than one session and is designed to be an interactive question and answer activity in which you and the participants explore data rights as set out by the Information Commissioners office:

- <https://ico.org.uk/your-data-matters/>

Participants should work through each of the sections under "Your Rights". You may want to set some questions for them to answer in relation to each section, or just allow them to ask questions as they work through. The goal is to ensure that participants have a good understanding of their data rights in the UK.

12.2.1 Resources

Other ICO resources that might be of use include:

- Explanation of the GDPR: <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/>
- Examples of enforcement activity the ICO has taken: <https://ico.org.uk/action-weve-taken/enforcement/>

If this activity is done in conjunction with Activity 7 you may want participants to explore the investigation undertaken by ICO on use of data for political purposes:

- Web site covering the investigation and enforcement activities: <https://ico.org.uk/action-weve-taken/investigation-into-data-analytics-for-political-purposes/>

12.3 Suggested activity 9: Community data civic action

Level: this activity is designed for people with good digital skills and who use a variety of digital platforms – likely our general to extensive users (see personas in Part 2: Section 6).

Learning outcomes: An experience of active Data Participation. This will engage all elements of Data Literacy (Table 4) including the foundational and data analytic skills not covered directly in this guide. It will ensure experience of ‘Understanding data society’, ‘Participating in society using data’, ‘Engagement with data society debates’, ‘Data Activism, and ‘Supporting others with their data literacy’.

This activity is designed to engage your participants in concrete Data Participation activity. **It will require much greater preparation and there is no single approach.** To undertake this, you will need to be working with or engaged with a community or civic organisation. The activity will have four components:

1. Working with the community or civic organisation to define a problem they need solving about their data, through analysing or visualising their data, or helping people to use their data. For example:
 - Using data to help a local charity better understand the people who donate or the people they support.
 - Developing a social media campaign for a local community centre.
 - Helping a local community improve its Digital or Data Literacy.
2. Support the participants to design with the community or civic organisation an intervention to address the identified challenge.
3. Implement and evaluate the intervention.
4. Get the participants to write up the evaluation reflecting on:
 - The Digital and Data skills they used
 - The barriers they and the community or civic organisation encountered – exploring the underlying social, economic, cultural and personal aspects of these barriers
 - The limitations placed on the intervention by the available data and the technology platforms used
 - The opportunities provided to the intervention by the available data and the technology platforms used

12.3.1 Resources

- A data civics guide for libraries – also includes basic data civics resources: <https://civic-switchboard.gitbook.io/guide/>
- A digital civics tool kit developed by Harvard University: <https://www.digitalcivics toolkit.org>
- A digital civic engagement tool kit developed through Erasmus+ funding: <https://www.studentcivicingagers.eu/toolkit-en/>
- A general engagement toolkit from Imperial college: <https://www.imperial.ac.uk/be-inspired/societal-engagement/resources-and-case-studies/engagement-toolkit/>

13 Acknowledgments

The team would like to thank the Nuffield Foundation (www.nuffieldfoundation.org) for funding this project as well as the support and input from many colleagues along the way. We would particularly like to thank Catherine Dennison at the Nuffield Foundation for being our project liaison and all members of the Advisory Group for their helpful insights throughout the project.

We would like to thank our colleagues at University of Liverpool, Good Things Foundation, and Critical Research for their support, comments and encouragement. In particular we would thank Dr. Tamara West and Dr Justine Gangneux, for their contributions early on in the project, and Dr. Al Mathers and Dr. Emma Stone at Good Things for their support and input. Huge thanks to Niki McQuaid for ceaseless administrative and PA support throughout. We extend special thanks to the centre staff in the Good Things network who supported the recruitment of respondents and the management of focus groups during the constant challenges of COVID-19 lockdowns. Last, and by no means least, huge thanks to the survey respondents and focus group participants whose responses and comments underpin these results.

As Principal Investigator I (Prof. Yates) would like to thank the post-doctoral research team – Dr. Elinor Carmi and Dr. Alicja Pawluczuk for their hard work and support.

This ‘Me and my big data – developing citizens’ data literacies’ project was funded by The Nuffield Foundation under grant number FR-000021473. The Nuffield Foundation is an independent charitable trust with a mission to advance social well-being. It funds research that informs social policy, primarily in Education, Welfare, and Justice. It also funds student programmes that provide opportunities for young people to develop skills in quantitative and scientific methods. The Nuffield Foundation is the founder and co-funder of the Nuffield Council on Bioethics and the Ada Lovelace Institute. The Foundation has funded this project, but the views expressed are those of the authors and not necessarily the Foundation.



Advisory Board

Prof. Payal Arora. Erasmus University Rotterdam
Dr. Priya Kumer. Global Affairs Canada / Social Media Lab, Ryerson University
Dr. Josie Bernard. De Montfort University
Mr Douglas White. Carnegie Trust UK
Dr. Grant Blank. Oxford Internet Institute. University of Oxford
Prof. Sheila Cotton. Michigan State University
Prof. Catherin Brooks. University of Arizona
Mr Julian Williamson. Department for Work and Pensions