



City Research Online

City St George's, University of London

Citation: Yates, S., Carmi, E., Lockley, E., Wessels, B. & Pawluczuk, A. (2021). Me and My Big Data: Understanding Citizens Data Literacies - Final report. Liverpool, UK: Nuffield Foundation.

This is the published version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

Permanent repository link: <https://openaccess.city.ac.uk/id/eprint/26756/>

Copyright and Reuse: Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to [City Research Online policy](#).



Understanding Citizens Data Literacies

Research Report

Authors:

Professor Simeon J. Yates (University of Liverpool)

Dr. Elinor Carmi (City University)

Dr. Eleanor Lockley (Sheffield Hallam University)

Professor Bridgette Wessels (University of Glasgow)

Dr Alicja Pawluczuk (United Nations University Institute in Macau)

Contents

CONTENTS.....	II
ACKNOWLEDGEMENTS	V
EXECUTIVE SUMMARY	VI
1 INTRODUCTION	1
1.1 OUR GOALS	1
1.2 THE CONTEXT	2
1.2.1 <i>Pre-pandemic</i>	3
1.2.2 <i>COVID-19 pandemic</i>	3
1.3 STRUCTURE OF THE REPORT	4
2 BUILDING A MODEL OF DATA CITIZENSHIP.....	5
2.1 INTRODUCTION	5
2.2 CONTEXT	5
2.3 DIGITAL VS DATA CITIZENSHIP	6
2.4 DIGITAL CITIZENSHIP IN THE ERA OF BIG DATA DIVIDES	7
2.5 POWER, DIGITAL AND DATA CITIZENS	8
2.6 DEMOCRATIC EDUCATION AND DATA CITIZENSHIP	11
2.7 THE ELEMENTS OF DATA CITIZENSHIP	12
2.8 DATA LITERACY	14
2.9 FRAMEWORK	15
2.9.1 <i>Data Doing</i>	17
2.9.2 <i>Data Thinking</i>	19
2.9.3 <i>Data Participation</i>	21
2.10 CONCLUSION	24
3 A NATIONAL SURVEY VIEW OF DATA CITIZENSHIP	26
3.1 INTRODUCTION	26
3.1.1 <i>Survey design and administration</i>	26
3.2 DEFINING USERS OF DIGITAL SYSTEMS	27
3.2.1 <i>Demographics of different “user types”</i>	29
3.3 HOW DO OUR USER TYPES DIFFER IN TERMS OF DATA LITERACIES?	31
3.3.1 <i>Checking data and content found online</i>	31
3.3.2 <i>Knowledge of data overtly shared or passively “given off” and the reasons behind it</i>	32

3.3.3	<i>Acceptability of data collection and third-party sharing of data by platforms</i>	33
3.3.4	<i>Attitudes to companies and platforms use of and protection of data</i>	35
3.3.5	<i>Citizens confidence with their own data protection actions</i>	36
3.3.6	<i>Trust in digital content and social networks</i>	Error! Bookmark not defined.
3.3.7	<i>Citizens trust in digital platforms and organisations</i>	40
3.3.8	<i>Supporting others and using data</i>	40
3.3.9	<i>Measuring Data Citizenship</i>	42
3.4	CONCLUSIONS: PERSONAS	45
3.4.1	<i>Extensive political users</i>	45
3.4.2	<i>Non-political extensive political users</i>	46
3.4.3	<i>Social and entertainment media users</i>	47
3.4.4	<i>General users (limited social media)</i>	47
3.4.5	<i>Limited users</i>	48
4	UNDERSTANDING DATA CITIZENS	49
4.1	INTRODUCTION	49
4.2	FOCUS GROUPS	49
4.3	THEMES	50
4.4	LIMITED OR LITTLE UNDERSTANDING OF HOW DIGITAL SYSTEMS AND MEDIA WORK	51
4.5	FEELING DISEMPOWERED	55
4.5.1	<i>Feelings of disempowerment and worry</i>	56
4.5.2	<i>Lack of personal digital efficacy</i>	57
4.5.3	<i>The problem of practical action</i>	58
4.5.4	<i>Hacking</i>	60
4.5.5	<i>Rights</i>	60
4.6	SEARCHING FOR INFORMATION AND HELP ONLINE	61
4.6.1	<i>Networks of literacy</i>	61
4.6.2	<i>Verifying news and information</i>	63
4.6.3	<i>Trust in content, media, platforms, and others</i>	66
4.6.4	<i>Learning new skills</i>	68
4.7	CONCLUSION	69
4.7.1	<i>Conclusion: limited or little understanding of how digital systems and media work</i>	69
4.7.2	<i>Conclusion: feeling disempowered</i>	70
4.7.3	<i>Conclusion: “networks of literacy” and verifying information</i>	71
4.7.4	<i>Conclusion: seeking skills</i>	72
5	THE CHALLENGE OF CRITICAL EDUCATIONAL INTERVENTION	73

5.1	FEAR AND WORRY	75
5.2	THE CHALLENGE OF PRACTICAL ACTION.....	75
5.3	RESPONSIBILITIES.....	76
5.3.1	<i>Citizens</i>	76
5.3.2	<i>Industry</i>	77
5.3.3	<i>Governments, policy makers and regulators</i>	78
5.4	AUDIENCES	79
5.5	ROUTES TO ENGAGEMENT	80
5.6	SKILLS VS UNDERSTANDING	81
5.7	CRITICAL EDUCATION.....	82
6	CONCLUSIONS	83
6.1	OVERALL DATA LITERACIES	83
6.2	CITIZENS RESPONSE TO THE POWER IMBALANCE BETWEEN PLATFORMS AND THEMSELVES.....	86
6.3	POLICY RECOMMENDATIONS	87
7	THE TEAM	90
8	REFERENCES.....	91
9	METHODS AND RESULTS ANNEX	96
9.1	SURVEY SAMPLE DESIGN AND ADMINISTRATION	96
9.2	DATA ANALYSIS	96
9.3	DATA TABLES	96
9.4	FIGURES	102
9.5	FOCUS GROUP DETAILS	107

Acknowledgements

The team would like to thank the Nuffield Foundation (www.nuffieldfoundation.org) for funding this project as well as the support in input from many colleagues along the way. We would like to thank Catherine Dennison at the Nuffield Foundation for managing the project 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 thier contributions early in the project, and Dr. Al Mathers and Dr. Emma Stone at Good Things for their support and input throughout. 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

Executive Summary

Introduction

We all leave data traces as we move through our digital lives. Some of this is intentional as we sign up to services or make purchases, but much of it is not. In some if not many cases, our actions online are tracked coercively and without our knowledge. This may be cookies and apps tracking us across our various internet interactions or data we post into social media about ourselves. Much of it may be collected, amalgamated, and processed by the platforms we knowingly use, or passed on to third parties for such purposes as advertising, marketing, or political campaigning. How much do we as citizens know about where our data goes, the uses to which it is put and what impacts this might then have upon us? These questions form the heart of the 'Me and my big data – developing citizens' data literacies' project. The project was conceived before many recent events and incidents that highlighted the issue of how data about us, as citizens, is collected, processed, and used. In the UK and the USA the Cambridge Analytica scandal, issues of 'fake news' and mis-/dis-/mal-information in social media, and concerns about COVID-19 tracking apps are examples of developments that have brought questions of data use to the fore.

Yet as we finish writing this report citizens in Afghanistan are actively deleting social media profiles for fear of the abuses and surveillance that the new government can do with their data. There are also fears about the potential uses the Taliban regime may make of the e-Tazkira system a biometric identity card used by Afghanistan's National Statistics and Information Authority. This includes fingerprints, iris scans and a photograph, as well as voter registration databases. Suddenly citizens become aware of their data traces, as they are of stark concern and potentially deadly consequence. The political and civil rights issues of such systems also remain important in the UK, Europe, and the USA. Here we see contrasting models of minimal, decentralised and citizen-controlled data sharing (such as the EU vaccine passport) compared to centralised amalgamated data systems such as the original plans for NHS contact tracing system. One addresses user privacy rights while the other may create a data set of use to medical research. The data traces we leave can therefore have material consequences, both good and ill, for individuals, communities and societies and are becoming bound up in our civil, consume and personal rights.

Much prior research has focused on technical solutions to data sharing, its regulation or on broader questions of data and digital rights. In this project we have focused on the exploring the extent of citizens understanding of how their data may be collected and utilised by platforms, companies, organisations or the state. The intention being to develop potential policy intervention recommendations or educational materials to improve citizens 'data literacies'. As both the project data collection and external events unfolded, the complexities of the issues, the size of the gaps in knowledge and significant differences in awareness between different demographic groups became very clear. Overall, it seems we lack key awareness of what is happening to our data, nor are we the engaged data citizens we might hope to be and that may be needed in a highly 'datafied' society. This report focuses on the main research findings along with our basic policy and education recommendations and principles. A following educational and intervention guidance report will provide a more practical set of ideas for activities that organisations, groups and policy makers might wish to undertake.

Research methods

The project report covers the seven main elements of the project:

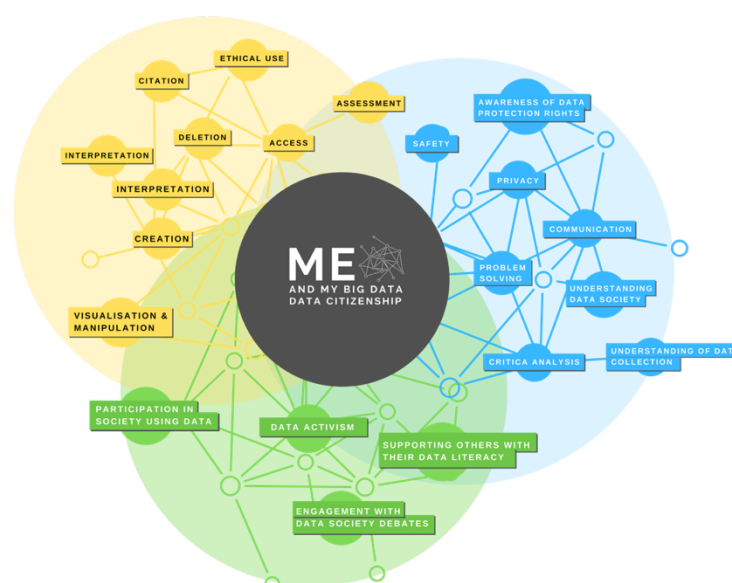
1. Review of the literature and development of a Data Citizenship framework
2. Operationalisation of the Data Citizenship model in the survey.
3. Analysis of the survey data in relation to key demographics and types of internet users
4. The development of Data Citizenship ‘Personas’.
5. Focus group interviews with respondents likely to fit the key Data Citizenship Persona groups.
6. Thematic analysis of focus group interactions
7. Development of key data literacy education principles

The project collected data through two main methods:

1. A nationally representative UK survey (n = 1,542)
2. Focus group interviews with target audience groups (groups = 14, n = 69)

Data Citizenship framework

The *Data Citizenship* Framework we have developed is intended to be open-ended in that the elements within it may change as technologies and social contexts change. Through a systematic analysis of literature on digital and data literacies we identified 21 key components of data literacy outlined by prior works. We combined these with ideas of democratic education such as those of Paolo Freire (1970). We frame Freire’s analysis of social power imbalances within the contemporary discussion on digital and divides – particularly the power imbalance between **data subjects (citizens)** and **data processors (big tech)**. We argue that, due to its collective, socially contextualised, and people-centred qualities, democratic education provides a useful foundation for any future data literacies education and research interventions. Finally, linking these elements together, we have proposed a *Data Citizenship* Framework. As discussed in the full report, we have identified three domains that define our *Data Citizenship* Framework: (1) **Data Doing**, (2) **Data Thinking**, (3) **Data Participation**. Data Citizenship views the data literate citizen as an active thinker, doer, and participant in the data society.



National Survey – Key findings

We identified six types of internet user in the survey data using Latent Class Analysis:

1. Extensive political users: Very likely to undertake all forms of online activity including political or civic action
2. Non-political extensive users: Very likely to undertake all forms of online activity but not as likely to undertake political or civic action
3. Social and entertainment media users: Low levels of overall use with a focus on social media and entertainments media
4. General users: General use with a focus on functional activities but limited social media use
5. Limited users: Limited use across all activities
6. Non-users: Do not make personal use of digital media or systems

This report focuses on the five groups that undertake some activity online. From our analysis of the survey data, we have identified some key characteristics of respondents in relation to data literacy:

1. We find that on many specific points of data literacy knowledge, practice, and awareness few respondents score highly. Except for a few more basic items, the proportions of any of our user groups showing specific knowledge or practice are rarely ever above 70%.
2. Therefore, only a small proportion of respondents have the deep and broad data literacy envisioned in our Data Citizenship framework
3. Low levels of use ('Limited' users) and narrow use ('Social and entertainment media' users) consistently correlate with lowest awareness and most limited actions.
4. Responses to specific aspects of data literacy and our overall data literacy measure strongly correspond to our internet user types.
5. Not surprising given this correspondence, data literacy scores are statistically correlated with key demographics such as education, age and socio-economic status (NRS social grade).

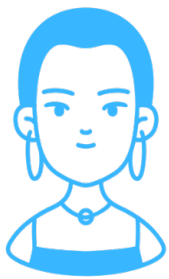
Using Multiple Correspondence Analysis on our data literacy questionnaire items we developed a 'Data Citizenship' measure. A regression model across key demographics finds that younger respondents with a university education, in NRS social grades AB will score more highly than older people with lower levels of education and in a lower socio-economic grade. Taking three possible respondents and applying our regression equation we find starkly different scores:

- A 30-year-old, higher university degree graduate in social grade AB would score in the upper quartile of our data
- A 24-year-old, with basic secondary school education in social grade DE would score in our second lower quartile
- A 65-year-old, with no formal education in social grade DE would score in the lowest quartile of our data

These scores are of course *relative* (not absolute) measures of Data Citizenship and literacy – they are scores relative to the range of results from our survey. As we have noted above, overall data literacy measures for individuals or collectively across groups are rarely high or close to the potential maxima. To understand these results better, and building on our five types of users, we have developed "Data Citizenship" personas. These 'ideal type' descriptions of users, their demographics and their specific data literacy capabilities and practices also provided a basis for our further qualitative workshops.

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

Social and entertainment media 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

Focus group findings

Our focus group interviews provided both confirmation of but also greater insight into the understandings, reactions, and experiences of our different ‘user types’. Ten focus groups mainly covered our lower level (‘Limited’) and narrow (‘Social and entertainment media’) user groups. Two groups covered our more ‘Extensive’ user groups. Key findings from the focus groups include:

- Many citizens have a very limited understanding of how both the technologies themselves work nor do they have detailed understanding of the economics and technical ecosystems that underpin platforms.
- Knowledge of the details of both overt and covert data collection, sharing and trading by platforms and other organisations may be quite limited for many users, this does not mean they are not aware it is happening. In all the focus groups respondents expressed an awareness that data around use of platforms is collected; but what, how, why and which organisations are involved with it – were often poorly understood. Similarly, nearly all respondents had not read nor understood what forms of data collection and processing they had agreed to through platform “terms and conditions”.
- ‘User types’ clearly correspond to some variations in attitudes. What they bring to the fore are the ways in which this experience plays out for different groups. This includes: the ability to understand the nuances of risks inherent in sharing and “giving off” data when online; the ability to differentiate (or not) “secure data exchange over a network” from “data being secure from unwanted use”; and ways in which citizens manage their experience of and often resignation to the realities of the ‘new normal’ of data being collected and processed.
- This resignation to the ‘new normal’ of data being collected on users is not passive and it has led to changes in behaviour even total avoidance of digital platforms. Importantly, in all the focus group sessions, respondents talked about feeling resigned but also ‘uneasy’ with data being tracked or shared in many ways. Very often this unease was expressed in quite forceful terms such as “creepy”, “scary” and “horrid”.
- Though the survey points to concerns, ambiguities, and differences in attitudes to the uses of data by platforms the survey results do not highlight the personal responses to these concerns. The focus groups make clear how citizens maintain both background and overt anxieties and concerns. Yet even the most engaged find acting on these concerns a challenge – often limited by the practical need to “get on with the job” of using platforms or though lack of knowledge. Very often these limitations are presented as personal failings as much if not more so than failings of the platforms or regulators.
- We find that respondents do not translate this unease into sustained action to protect their data nor to assert digital rights. 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. However our respondents were also just as likely to blame their own “failings”, lack of skill or lack of personal “vigilance” for the situation.

- Respondents clearly maintain 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.
- Citizens “networks of literacy” made up of friends, family colleagues and social media contacts are key to their skills and knowledge. They operate on different scales and with different levels of skill and knowledge among their members. Citizens “networks of literacy” provide the basis for citizens navigation of digital content, their acquisition of skills and knowledge, their verification of information and underpin their community engagement. They therefore underpin and cut across all three of our data-citizenship dimensions. They support citizens in “doing” things with data, they support their “thinking” about and with data, and they underpin their Data Participation.
- The focus groups provide the same mixed evidence as the survey around trust of platforms and assessment of content. With many respondents:
 - Only doing 'some things' to check content
 - Having very different levels and areas of trust in technologies, platforms and organisations
 - Only a very few having deep and broad data and digital literacies
 - All respondents face the 'challenge of practical action' in using their knowledge, understanding and skills to check content
- Limited or narrow users were the least able to articulate clearly the nature and types of data shared/“given off” and clearly did not actively think about these issues – unless a specific fear (surveillance) or a specific incident (scam/hack) had raised their awareness. Though much academic and policy work talks about ‘data’, ‘sharing of data’ or ‘data protection’ only more experienced users talked in this way and, importantly, could articulate clearly what they meant. Others were clearly more comfortable talking about ‘information’ or specific items of data/information.
- We draw the following conclusions from our focus group work.
 - The depth and breadth of knowledge about the collection, tracking and use of data by platforms is best described as “patchy”. It is especially limited for those people who have low digital skills and are limited or narrow users of digital systems.
 - All respondents feel unease and differing levels of disempowerment around how platforms collect data, the uses to which it is put and the potential of third-party trading. This unease is 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”.
 - The processes to verify information, the learning of new skills or the development of better awareness are tied to users “networks of literacy”.

The challenge of critical educational intervention

We have identified seven issues that we feel flow from our findings and that need to be addressed by any educational or policy intervention:

1. Fear and worry
 - Any educational or awareness raising interventions must ensure citizens feel more empowered and have practical and alternative routes to enact that empowerment.
2. The challenge of practical action
 - Any educational or awareness raising interventions need to 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.

3. Rights and responsibilities

- Any educational or awareness raising interventions need to make clear to citizens their rights and responsibilities in regard to their own data but also the role of industry and regulators.
- Any educational interventions or awareness raising must make clear to citizens the role of industry and the practical options provided by, but also limitations of, platforms.
- There is a clear role for government and regulators to intervene in two areas: support for and provision of data, digital and media literacy education and training; democratic regulation of digital industries.
- We would argue that responsibility of allocating appropriate resources to increase data literacy lies first with government. In particular to support those citizens outside of formal education with more accessible spaces, such as libraries, where professionals can assist and support citizens to develop greater data literacy.
- Legislators need to make sure that the mechanisms that are offered to people to make rights claims are easy to use and straightforward.
- Organisations such as the Information Commissioners Office must do more to engage with citizens from different backgrounds and make the process of making rights claims easy and fast.
- Any educational interventions or awareness raising must 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.

4. Audiences

- Any educational interventions or awareness raising must 'meet citizens where they are' in terms of their digital and social experience and context.

5. Routes to engagement

- Any educational interventions or awareness raising must address the challenge that those adults most in need of support are very likely outside formal educational settings.

6. Skills vs understanding

- We therefore argue that skills are only part of the story, and Data Citizenship requires more than just skills, it requires the development of greater critical awareness and a more holistic understanding of the data ecosystem.
- Any educational interventions or awareness raising must support skills development, but must be more than skills encompassing key elements of Data Participation.

7. Critical education

- Any educational interventions or awareness raising must 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.



Understanding Citizens Data Literacies

Research Report

1 Introduction

The Me and My Big Data project sought to understand the levels of and demographic variations in UK citizens data literacies. From this we have developed policy advice and recommendations to support improving this. The outline for this project was submitted the week before the “Cambridge Analytica” scandal broke in 2018. This made clear the extent to which we as citizens were unaware, and as we will show remain unaware, of the uses and abuses to which our data can be put. Improving digital and data literacies remains a key policy goal of governments worldwide.

A key component of citizens’ digital literacy is an understanding of the uses made of their data. Unfortunately, prior evidence from the UK and USA, as well as our results presented here, indicate that many citizens continue to have limited understanding of what is data, how their data is tracked and collected, its use by various organisations, nor do they engage in basic data and privacy protection behaviours. Citizens are also very unlikely to engage in proactive citizenship in relation to our ‘datafied’ society; nor are they aware of how they can use publicly available data to undertake both personal and civic action. This lack of “data literacies” opens citizens up to risks and limits their ability to operate as proactive citizens in a digital and ‘datafied’ society. Importantly, evidence is growing of inequalities in data literacies that mirror broader social inequality.

1.1 Our goals

Over the last five years (2017-2021) Ofcom’s annual media literacy report has pointed out limitations in citizens digital and data literacies. For example, though most internet users were aware of at least one of the ways in which companies collect information about them; the majority remain unaware of the breadth and depth of uses. Nearly 40% of respondents could not correctly identify “advertising” content from other results in web search results. At the same time many citizens lacked the knowledge and skills needed to use publicly available data as part of personal or civic action (Doteveryone, 2018). This lack of “data literacy” is viewed as leaving citizens at risk of ‘online harms’ – both personal and financial – but can also limit their ability to operate as active citizens in an increasingly digital society (Lloyds, 2018). Therefore, the initial goal of this project was:

- 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.

Importantly, in developing the proposal we were aware, from our own deeper analyses of the Ofcom data, that differences in digital literacy mirrored other indices of inequality (Yates and Lockley, 2018). We therefore sought:

- To understand the intersectional basis of variations and inequalities in data literacy across a range of demographics factors.

Any work in this area would intersect with ongoing policy debates and work at an international level (e.g., UNESCO's (Law et al, 2018) work on digital literacies) or through regulatory changes (e.g. the bringing in of the General Data Protection Regulation in 2018). Such work will also importantly link to public concerns over the effects of social media (mis-/dis-/mal-information, Cambridge Analytica), repeated data breaches, and growing inequities in the uses of digital media. At the time, the issues had also been raised in UK government reports (House of Commons June 2016, DBIS & DCMS 2016). The project therefore also sought:

- To develop policy recommendations for stakeholders on developing citizens' data literacy.

1.2 The context

The project builds on one of the key findings from a prior study by team members, a systematic research review commissioned by the Economic and Social Research Council on "Ways of Being in a Digital Age" (see Yates and Rice, 2020). This review covered seven 'domains' of digital society research and included expert panel assessments of key current and future research challenges. Across all seven domains, the work identified the need for research to better understand and then support policy to improve, citizens' digital and data literacies. In addition, the work highlighted the need to understand the intersection of these literacies with digital and social inequalities. In designing the study, we viewed data literacy as covering citizens' understanding of three things:

1. Knowledge about what data they are regularly sharing¹ and what public, private and other organisations do with that data.
2. How they can control both what data they release and how it is used.
3. The use of data (personal and public) by citizens and the skills they need for this, so as to undertake civic action and participation.

¹ When designing the study, we used the word 'sharing' to imply two things: first, proactive and knowing sharing of data, such as giving an email address to an app; and second, passive unknowing sharing of data, such as through cookies. In discussion with our advisory board, we are separating these out as the term sharing connotes more active engagement than is the case for data collected without users' active engagement. On reflection we view citizens as "giving off" data in a manner similar to that described by Goffman of behaviour in public. Some, but not all, has an intentional or known aspect, but much may not. As described by Goffman in relation to public performance, this passive "giving off" of data is far less controllable than active intentional interaction.

The idea of digital technology use being a form of literacy and vice versa has a long history, especially in the UK (e.g. the BBC Computer Literacy Project (1980-1989)). The theoretical examination of ICT use as form of literacy has a similarly long heritage (e.g. Finnegan, 1989). Digital literacy, as defined above, is therefore a much broader concept within which we are focusing on “data literacy”. Our prior work (Yates et al, 2020; Yates and Lockley, 2018) clearly indicated that the breadth and depth of digital engagement varies with social class – corresponding to markers of economic, social, and cultural capital.

1.2.1 Pre-pandemic

As noted above the project was initially designed just before the Cambridge Analytica scandal broke. Though barely four years ago much has changed since that point. The focus of concern around online harms has shifted from privacy and data exploitation to mis-/dis-/mal-information (Carmi et al., 2020). Particularly the idea of ‘fake news’ and the visible (and invisible) distribution and perpetuation via social media of un-truths and forms of propaganda have taken centre stage. Data, though, remains at the heart of these concerns as it is the processing of data on attitudes and behaviours that underpins the distribution of content across social networks. Significant data breaches have taken place globally during this period and have been reported in the global press. These are often framed as “cybercrime” and less often as threats to the wellbeing of citizens. As the project progressed, we expected this public discussion and visibility to impact respondent behaviours and attitudes. As our results will show, despite the high visibility of these issues in the press, individual respondents still show low levels of awareness and low levels of potentially protective or proactive behaviour.

1.2.2 COVID-19 pandemic

The pandemic has had several impacts on the project. First, the project had to pause for 6 months between the survey and workshops stages of the project as we could not undertake face-to-face interactions with respondents. Importantly, this changed the qualitative element from face-to-face workshops to online (Zoom) based focus-group interviews. This has led to a different approach to the educational policy elements of the project. The initial plan had been to co-create and test actual educational materials with respondents. The inability to undertake the workshops face-to-face as planned directly prevented this approach. The actual experience of the focus groups, building on the survey findings, made very clear the diversity of audiences. These findings indicated that a much more diverse and contextual model of data literacies education, training and skills was needed. The project therefore pivoted to setting out guidance for educators in developing interventions and the aggregating of materials available for educators.

A second impact has been the digital and data impacts of the pandemic. The pandemic has made clear our dependence on digital systems and made very visible the significant digital inequalities and low levels of

digital and data literacy among citizens. The political mis-/dis-/mal-information environment described above has developed and moved on to issues of the pandemic and vaccination where wrong or poor information has life and death consequences. This so called “infodemic” again highlights the need for robust critical information, digital and data literacies. Team members have been involved in research and policy responses to this².

1.3 Structure of the report

The report has six main sections:

In **section 2** we set out the background to our model of ‘Data Citizenship’ drawing on existing literature and ideas of ‘democratic education’. We then break the model down into key components and behaviours that we operationalised into a national survey.

Section 3 presents the results of this survey and builds a statistical model of the demographic factors underpinning levels of data literacy and citizenship. Section 3 also presents five “Data Citizenship’ personas to help identify audiences for policy interventions and to help define target groups for the following focus group work.

Section 4 presents our thematic analysis of the focus group discussions. These expand upon some of the statistical findings and provide insight into the experiences of digital systems and media use that underpin citizens data and digital literacies.

Section 5 lays out our provisional framework for policy or education interventions. A fuller exposition of this work will be provided in a further report.

Section 6 concludes with summarises the findings and frameworks.

² <https://fakenewsimmunity.liverpool.ac.uk/publications/>

2 Building a model of Data Citizenship

2.1 Introduction

In this section we outline the development of our *Data Citizenship* Framework for both examining and developing citizens' data literacies and Data Citizenship – broadly understood as “citizens' ability to understand and proactively use data in their everyday lives and communities”. We initially examine how data collection methods, by platforms and organisations, impact on citizens’ opportunities for meaningful civic and digital participation. We explore a range of interpretations of digital and Data Citizenship and the conceptions of digital and data literacies that underpin them.

The theoretical foundations of our *Data Citizenship* Framework can be found in the work of democratic educators, critical data studies and digital citizenship scholars. The Framework is intended to be open-ended in that the elements within it may change as technologies and social contexts change. Through a systematic analysis of literature on digital and data literacies we identify 21 key components of data literacy outlined by prior works. We combine these with ideas of democratic education such as those of Freire (1970). We frame Freire’s analysis of social power imbalances within the contemporary discussion on digital and divides. We argue that, due to its collective, socially contextualised, and people-centred qualities, democratic education provides a useful foundation for any future data literacies education and research interventions. Finally, linking these elements together, we have proposed a *Data Citizenship* Framework. As we discuss later, we have identified three domains that define our *Data Citizenship* framework: (1) Data Doing, (2) Data Thinking, (3) Data Participation. Data Citizenship views the data literate citizen as an active thinker, doer, and participant in the data society.

2.2 Context

The uses of digital media and platforms are significantly shaped by three key factors:

- First, their hidden infrastructures and the algorithms informing them (See Bucher, 2012, Carmi, 2020).
- Second, by the commercial and data-extractive strategies of ‘big tech’ corporations owning and operating them (Andrejevic, 2014; Dencik et al, 2018).
- Third, through taking place within the power structures of surveillance capitalism (Zuboff, 2019).

It is also increasingly difficult for citizens to opt out of using digital media and systems as ever more aspects of social life are becoming ‘datafied’, effectively generating coerced forms of digital participation (Barassi, 2019). In this context, people often feel powerless (Hintz et al 2018), and it is argued that many platforms are seen to cultivate ‘digital resignation’ as a strategy to neutralize critical, collective, or political action (Draper and Turow, 2019).

The 2017 Cambridge Analytica case represents a clear example of how issues of data use, data power and citizens' lack of data literacies can play out in civic and political terms (Issak & Hannah, 2018).

Unfortunately, the power imbalance at play in the Cambridge Analytica case has become the 'new normal'. Citizens' ability to make rights claims so as to change regulations and affect digital systems and media is therefore key. Equipping citizens with the necessary digital skills to engage in this also depends on exposing and challenging the stark power imbalance between 'big-tech' corporations and citizens in the context of digital media. A critical awareness and understanding of broader digital society and its political and economic landscape is, therefore, a core element of digital citizenship in the 21st century.

Digital literacy builds on multiple ideas including media literacy (Mihailidis & Thevenin, 2013), data literacy (Crusoe, 2016; Grillenberger & Romeike, 2018) and information literacy (Carlson & Johnston, 2015). All of these models emphasise the importance of digital participation by pro-active and informed citizens, with roots in ideas of critical literacies (Street and Street, 1984). Active digital and Data Citizenship therefore both rest upon a foundation of basic skills and broader complex literacies. To facilitate meaningful and democratic participation, citizens must therefore be able to understand, gather and analyse information, develop informed opinions, and discuss and negotiate these perspectives with others (Mihailidis & Thevenin, 2013).

2.3 Digital vs Data Citizenship

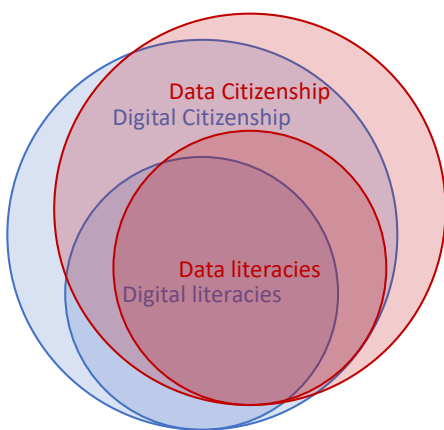


Figure 1: Overlapping but separate concepts of digital and Data Citizenship and literacies

In the project we worked with four concepts that are closely intertwined in the literature:

- Digital citizenship **and** Digital literacies
- Data citizenship **and** Data literacies

In some of the literature we discuss below, these terms are used almost synonymously. It is also our experience from our prior ESRC systematic review (Yates and Rice, 2020) that academics, industry, and government stakeholders also find it challenging to differentiate key concepts in this domain. For example,

experts found it difficult to unpick ideas around digital/data/algorithms/automation (Yates and Blejmar, 2020). In fact, as we will note again later, any actual activity in our digital and ‘datafied’ society often involves the elements of all these issues. For the purposes of the project, we would argue that the four digital and data concepts above are best understood as either overlapping issues, skills, knowledge, and actions or as two sides of the same coin – different but interdependent and closely aligned (see Figure 1).

Considering the mounting evidence of data misuses, privacy breaches and algorithmic influence, many have argued that citizens need to be equipped not only with technical but also critical skills to make sense of and manage the data they generate online. They also need to be aware of the data that corporations collect, aggregate, analyse and sell to third parties about them (Andrejevic, 2014; Selwyn, and Pangrazio, 2018; Hintz & Brown, 2019; Zuboff, 2019).

The analysis presented here reveals that while there is extensive literature on data literacies covering citizens' practical data management skills and critical thinking; there is limited emphasis on the value of critical and pro-active use of data by citizens in the data society – with the exception of data activist initiatives and scholarship (Gutierrez, 2019; Kennedy, 2018; Lehtiniemi and Haapoja, 2019; Milan and van der Velden, 2016). As we will present later (section 3), we have taken the Data Citizenship framework described here to link issues of data literacy and citizenship to key demographics and types of internet user (see also Carmi et al, 2020a; 2020b; Yates et al, 2020a Yates et al, 2015; 2019; 2020b; Yates and Lockley, 2018).

2.4 Digital citizenship in the era of big data divides

Digital citizenship follows on from a long history of scholarship and policymaking on citizenship and more recently ‘active citizenship’ often emphasising individual responsibility to participate in society (see Hintz et al 2018). Digital citizenship is discussed and defined within the literature in a variety of ways (see Yates, et al, 2020c); from the use of digital tools to specifically participate in ‘formal digital politics’ (Frame and Brachotte, 2016), through boarder conceptions of civic engagement or contribution to the “public sphere”, to a broader conception of social engagement. Hintz et al (2017) note that digital citizenship is often understood as “the (self) enactment of people’s role in society through the use of digital technologies” (p.1). This definition attributes empowering and democratic characteristics to citizenship by emphasising citizens’ abilities (and responsibilities) to understand, navigate and participate in the digital world.

However, it has been shown that the power relations and extensive data extraction processes characterising our digital landscape challenge in crucial ways citizens’ agency and ability to enact their citizenship (Hintz et al 2017).

2.5 Power, digital and data citizens

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? Enacting digital citizenship is more than just an ability to function within a digital or 'datafied' society. Isin and Ruppert (2015) argue that:

... 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).

In the context of the digital society Isin and Ruppert argue that citizens' acts, their performance of *conventions*, are undertaken through and embedded within digital platforms. For example, communication *conventions* are undertaken through blogging, emailing, liking, posting, retweeting etc. (2015, p.72). As a result, they argue that the analysis of the dynamics of citizen subjects participating, connecting, knowingly sharing or passively "giving off" data:

...enables us to understand to what extent citizen subjects are able to make digital rights claims in the form of 'I, we or they have a right to' by re-signifying conventions in which they are implicated rather than only obeying or submitting to them (Isin and Ruppert, 2015, p.98).

Whether we view data and digital citizenship more broadly as an ability to engage in civic 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.

Data is central to the digital society and, as we have argued above, Data Citizenship as a converse or complimentary concept to digital citizenship, 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. Yet research focusing on citizens' Data Citizenship in the context of the big data divide has been relatively recent (for example Andrejevic, 2014; McCarthy, 2016; Hintz et al 2017, 2018). Andrejevic, (2014)

argues that big data leaves citizens *powerless* in the light of the increasing forms of data collection and data mining and argues that the ‘big data divide’ reflects:

both the relations of ownership and control that shape access to communication and information resources, and growing awareness of just how little people know about the ways in which their data might be turned back upon them (2014, p.1675).

This implies that data and digital citizenship, either as a civic role or performance of rights claims, is often limited by the current power dynamics of our ‘datafied’ society in which ‘those who hold, manage, and control the personal data of digital citizens are offered unprecedented insights into our lives, minds, and bodies’ (Hintz et al 2017, p.732).

We would argue that there is therefore 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 and as a result can capitalise on this widening divide (Hintz et al 2017, p.732). These are largely private ‘big tech’, Internet service providers and the state. Technology companies hiding their data extraction behind complex and often opaque and potential deceptively designed interfaces (Carmi, 2020). 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. This blurring supports the creation of subjects who are governed by and through the collection and processing of their data by state and non-state organisations. Cheney-Lippold (2017) argues this is a form of governance by and around algorithms – as against governance in relation to space, place, territory or relationships. As Barasi (2019) points out, much of this governance is effectively hidden and coerced. Though it is clear that there is a divide among citizens as algorithmically driven platforms are primarily managed and understood by those who have the appropriate finances, infrastructure, and expertise (Andrejevic, 2014; McCarthy, 2016). It is therefore not surprising that citizens are not fully aware of how or when their data is collected, analysed, shared, and re-used (Andrejevic, 2014).

The sense of disempowerment around data leads to citizens feeling confused, powerless, and unable to identify, understand or respond to those who are in charge of their data (Hargittai & Marwick, 2016; Hintz et al, 2017, 2018). Our review of the literature reveals a number of studies examining the notion of the ‘privacy paradox’ (Hargittai & Marwick, 2016) whereby citizens are worried about their data online, but continue to create it nevertheless. Hargittai and Marwick (2016) have reported that in the context of the big-data divide, young people 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)

(Barassi, 2019, Carmi, 2020), some citizens decide to embrace the big-data divide as ‘the new normal’ (Lin et al., 2017).

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. Schmidt (2018) revealed that even when an Android mobile phone is stationary and idle over the course of a 24-hour period, 900 data samples are still shared with Google. 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. Thus, the analysis of citizens' digital activity as well as *inactivity* provide valuable insights to companies who aim to trade such data. As a result, McCarthy argues that:

... we are losing control in defining who we are online, or more specifically we are losing ownership over the meaning of the categories that constitute our identities. This algorithmic self-reference has the tendency to exacerbate whatever classification, trait, or division identified and contains the potential to shape user practices and self-identity (McCarthy 2016, p.1133).

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. The algorithmic governance of Cheney-Lippold is not yet absolute. Though data define many aspects of our contemporary citizenship, to assume that this is the totality of things, or that ‘resistance is futile’, is to ignore the “messiness and human dimension of data governance” (Barassi, 2019, p.426). We therefore see it as important that the conceptualisation of Data Citizenship includes the capacity for citizens to question, assess, challenge, and make rights claims within a ‘datafied’ society. This sets up key questions for our project:

- How do citizens respond to the power imbalance between these platforms, organisations and themselves?

- How do citizens feel in relation to this imbalance?
- To what extent do citizens trust digital media and systems – especially major platforms?
- To what extent to citizens trust the organisations that aggregate and use their data?

2.6 Democratic education and Data Citizenship

Data citizenship therefore takes place in the context of digital, social, cultural, economic, and democratic divides. Particularly in relation to citizens' ability to engage, participate and to feel efficacious in relation to their provision of, others use of, and their own use of data. We therefore argue that citizens must be provided with access to education, tools, and community support to foster both critical awareness of this context and new forms of agency in relation to society's data divides. In line with Hintz and Brown (2017), we argue here that citizens need to be empowered and supported to play a proactive role in the big-data divide debates (Barassi, 2019; Hintz & Brown, 2017).

This points to 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 eco-systems in which citizens are now effectively forced to operate. We have therefore brought the ideas of "democratic education" (Dewey, 1930; Freire, 1970) 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. Dewey argues that, "the communication of ideals, hopes, expectations, standards, opinions [by citizens]" (1930, p.4) are fundamental to achieve a just society. Freire argued that democratic education and collective reflection enables communities to co-create new ways to pursue "a fuller humanity" (1970, p.40); a more conscious, informed, and active way of living.

We would argue that Freire's conceptualisation of the oppressed-oppressors relationship is applicable in the context of the contemporary forms of data driven technology power divides. Instances of data mining, algorithmic selection and online surveillance have been defined as modern-day forms of social control (Zuboff, 2019). We argue that the big-data divide – between large organisations, institutions, and platforms with extensive capacity to collect, aggregate and analyse data and citizens with limited ability to control the use of their data – is analogous to Freire's conceptualisation of the "oppressed-oppressors". We will describe these as **data-processors (big tech)** and **data-subjects (citizens)**.

What is important for understanding Data Citizenship through the lens of Freire's conception of the "oppressed-oppressors" relationship is the manner in which the choices of one person or organisation become prescribed behaviour for others. In this case, the 'big tech' and major digital platforms have been

able to systematically prescribe how citizens, as users of platforms, must offer up their data and be subject to the outcomes of platforms decision making processes. Such processes are applied in, and have consequences for, many social spheres including politics, employment, health, insurance, travel and many more. This is analogous, if not equivalent to Isin and Ruppert's description of the “conventions” noted above, that digital citizens are subjected to and follow. Therefore, following Freire, we would argue that developing a deep critical consciousness of this power relationship is needed by citizens to allow them to exercise their right to challenge this imbalance and demand change. As Freire notes, such relationships become entrenched and normalised and this can be seen in how platforms’ use of data about and generated by citizens is ‘the new normal’.

The parallels between teaching critical consciousness and data literacies education have been outlined by Tygel and Kirsch (2015), who proposed that data-cantered 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” (2016, p.14).

The fundamental quality of critical consciousness is that it is a proactive mind-set, and an intellectual and practical engagement with a given social reality and social structure. Thus, democratic education is not a static and top-down knowledge process, nor is it primarily concerned with getting a specific set of skills of expertise. Instead, it aims to tap into citizens' social experiences with the [data] society and inspire their interest and action in their positions in the social and data power structures. For our project the questions become:

- To what extent have the conventions around data extraction and use by data-processors become ‘the new normal’?
- To what extent are citizens comfortable with this ‘new normal’?
- What actions do citizens take in response to these conventions – to protect their data or assess content that is shared?

2.7 The elements of Data Citizenship

Our Data Citizenship Framework therefore builds on two elements. First, an assessment of data literacies(s) from an analysis and review of current literature. Second, an assessment of data literacies within the broad framework of data and digital Citizenship and democratic education described above. Our assessment of current ideas of data literacy is built from an analysis of publications gathered through an extensive literature survey. The search resulted in 251 items of literature including (83) reports, (76) academic

articles, (21) books, and (71) news articles. We systematically reviewed the content of these works, focusing on the relationship between data literacy and citizenship. We focused on the how the models or findings could support:

- The development of a framework that specifically fills the existing gaps in this topic
- Link with ideas of democratic education and digital citizenship
- Provide a basis for the empirical assessment of data literacy

From this broad set of materials, we identified 13 items that explicitly provided definitions of data literacy, published between 2006 and 2019 (see Table 1). These form the core of our review and model.

Table 1: Key sources

Definition	Source	Type	Focus
“Data Literacy defined pro populo”	Crusoe, D. (2016)	Theoretical	Power relationships and control of data
Data Literacy	Wolff, A., Gooch, D., Montaner, J. J. C., Rashid, U., & Kortuem, G. (2016)	Literature review and analysis	Use and analysis of data
Data Literacy	Deahl, E. (2014)	Literature and case studies	Everyday awareness, use and analysis of data
Data Information Literacy	Carlson, J., & Johnston, L. (2015)	Empirical qualitative research and case studies	Data literacy for Higher Education
Data Literacy	Mandinach, E. B., & Gummer, E. S. (2013)	Literature review and analysis	Data literacy for Higher Education Managers
Data Literacy	Grillenberger, A., & Romeike, R. (2018)	Literature and curriculum analysis	Data literacy for computer science in Higher Education
Data Literacy	Vahey, P., Yarnall, L., Patton, C., Zalles, D., & Swan, K. (2006)	Quasi-experimental study of educational intervention	Data analysis skills development in school classroom setting
Youth Data Literacy	Williams, S., Deahl, E., Rubel, L., & Lim, V. (2014)	Participatory science project	Data analysis skills development in school classroom setting
Critical Data Literacy	Tygel, A., & Kirsch, R. (2015)	Theoretical	Critical data literacy for citizens
Racial Data Literacy	Philip, T. M., Olivares-Pasillas, M. C., & Rocha, J. (2016)	Qualitative case studies	Critical assessment of data analytics and data visualisation school teaching in relation to race
Personal Data Literacy	Pangrazio, L., & Selwyn, N. (2019)	Theoretical	Personal critical data literacies for citizens
Civic Data Literacy	(Civic Switchboard Guide, 2019)	Guidelines for using civic data	Libraries and support for citizens using civic data
Creative Data Literacy	D’Ignazio (2017)	Review of case studies	Using creative practice to build data skills

2.8 Data literacy

Many of the papers in Table 1 are summaries or overviews of literature but maintain a focus on one element of the broader spectrum represented by the other works. The main foci being: practical data handling and management; analytic skills; and critical thinking skills (Grillenberger & Romeike, 2018; Mandinach & Gummer, 2013; Wolf et al., 2016; Vahey et al., 2006). Examples of data management skills include identifying, collecting, and storing data. Examples of analytic skills include data analysis, interpretation, and visualisation. Critical thinking abilities cut across both data management and analytic skills as they require an ability to assess these in relation to power and social context.

In three of the studies (Grillenberger & Romeike, 2018; Wolf et al., 2016; Vahey et al., 2006), citizens are positioned as compliant and semi-autonomous participants of the data collection processes. Though all the papers emphasise the importance of data literacies they provide limited discussion of how to ground these competencies within a wider context of conscious and active citizenship within a 'datafied' society. For example, Crusoe's (2016) definition examines the disempowering nature of current data culture, but it mainly emphasises the importance of individuals' data protection behaviours. Grillenberger and Romeike (2018) propose a data literacy model that is primarily focused on students' and academics' competences to handle, store, and interpret data.

While all the papers in Table 1 have identified data literacies as a cornerstone in addressing the big data divide, there is limited guidance in terms of how to empower citizens to take on pro-active roles in wider political debates nor civic action around data governance. Overall, data literacies have been primarily understood as a set of practical and critical skills related to citizens' abilities to operate *within* the existing data power-structures (Crusoe, 2016; Grillenberger & Romeike, 2018; Vahey et al., 2006). Thus, it might be argued, that some of the existing interpretations of data literacies might, in some cases, reinforce the wider problem of the big-data divide. In Freire's terms, citizens are working withing the prescribed structures and practices of the major data processing organisations, institutions, and platforms. Though data literacies, as data management and analysis, could support and empower citizens, it also makes citizens more effective data contributors to those platforms. Therefore, reinforcing the unequal power dynamic by further individualising the responsibility to manage their data within existing *structures, conventions, and prescriptions*.

While we acknowledge the importance of data management and critical analysis skills in the formation of our data literacy framework, we would argue that a broader citizen focused definition is needed. Pangrazio & Selwyn (2019) provide the most general 'citizen' centred model of data literacy. Their focus is on personal 'data literacy' and they split data literacy into five elements (Pangrazio & Selwyn, 2019, p. 428-9):

1. Data Identification
2. Data Understandings
3. Data Reflexivity
4. Data Uses
5. Data Tactics

These are very useful foci, and we would note that this approach covers many but not all elements present in the other listed works. We would also note that many digital activities cut across and combine these elements as indicated in Table 2.

Table 2: Elements of data literacy and citizenship

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

This creates a challenge in identifying these five elements as separate factors in empirical work. Looking across the literature we have identified 21 activities that appear in the definitions of data literacies –

described in one manner or another. We have grouped these into three overlapping aspects of in Table 2. **“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 has to include working with others, groups and communities and may involve the collection, repurposing and use of data in such activism. “Data Participation” is implied but not fully developed in the literature we assessed. Our next step is to take these three aspects and discuss these in the context of a broad interpretation of Data Citizenship.

2.9 Framework

The Data Citizenship Framework we put forward is intended to be flexible and we would expect it be modified in the light of technological, environmental, and social change. Our analysis provides key points for reference for data literacies researchers and educators, but we acknowledge that any conceptualisation of Data Citizenship requires an ongoing review process and updates. Figure 2 illustrates a visual representation of Data Citizenship’s three overlapping domains: Data Thinking, Data Doing, and Data Participation. Below we provide an overview of Data Citizenship’s theoretical foundations and examine its importance for future research.

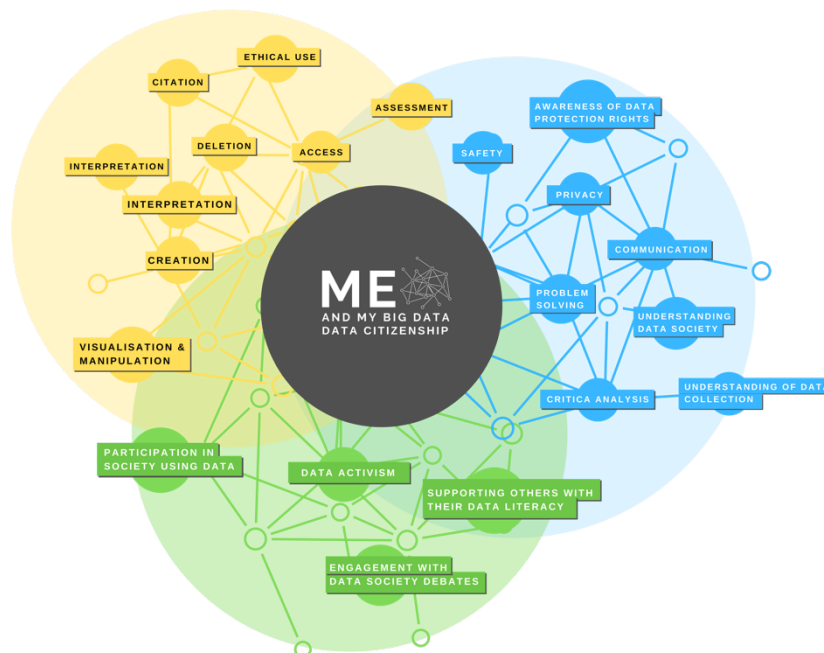


Figure 2: Three theoretical domains of the Data Citizenship framework: (1) Data Doing (yellow); (2) Data Thinking (blue); (3) Data Participation (green).

2.9.1 Data Doing

Data Doing incorporates practical skills that involve data handling and data management. Data literacy has become increasingly considered as a life skill (Wolf et al., 2016). Prior scholars have conceptualised this as core to data literacy (Crusoe, 2016; Grillenberger, & Romeike, 2018; Philip, Olivares-Pasillas & Rocha, 2016; Wolf et al., 2016) emphasising skills such as data access, data credibility assessment (e.g. fake news, misinformation), data interpretation, ethical use, and data manipulation. For example, Grillenberger and Romeike (2018) view data literacy as a way to collect, manage, evaluate, and apply data, in a critical manner. Wolf et al.'s (2016) Foundational Competencies for Data Literacy Framework argues that citizens should know how to select, clean, analyse and visualise data (2016, p.23) and how to transform data into actionable knowledge.

Our literature review indicates that ideas about the importance of Data Doing skills are primarily derived from within the educational domains of Computer and Data Science (Grillenberger, & Romeike, 2018; Mandinach & Gummer, 2016) and Information Literacy (Carlson & Johnston, 2015). In this strand of literature, data skills are often linked to Information Literacy, defined as:

... the ability to recognise when information is needed and have the ability to locate, evaluate and effectively use the needed information, with the ultimate goal of enabling lifelong learning (Carlson & Johnston, 2015).

Similarly, in their examination of data literacy education, Mandinach and Gummer (2016) describe the use of data as the largest component of their Data Literacy for Teachers conceptual framework (DLFT). DLFT provides an in-depth analysis of data literacy skills which include finding, locating, accessing, and retrieving data; using technologies to support data use; organising data into a meaningful and manageable representation of the information. In the Data Citizenship framework, such practical skills related to data management are considered as 'Data Doing'.

Ethical data use might involve citing the original data source or anonymising sensitive data. Data Doing advocates that, for example, social media users should be provided with opportunities to learn how to identify and highlight the source of the information they share in their timeline. For example, a data literate social media user would ensure that a photo or content they share or reshare online is appropriately cited. This helps in both crediting the people who created the photo/content as well as contextualise it and potentially help preventing misinformation that is often caused by 'cherry picking' (Carmi et al., 2021). The key elements of Data Doing are presented in Table 3.

Table 3: Data Doing domain of the Data Citizenship framework Participation

Data Doing		
Domain	Description	Literature References
Accessing	The ability to search for, identify and access services, websites, and data.	Crusoe, 2016; Deahl, 2014; D'Ignazio, 2017; Mandinach & Gummer, 2013; Grillenberger & Romeike, 2018; Montes & Slater, 2019; Pangrazio & Selwyn, 2019; Sweeper, 2014; Tygel & Kirsch, 2015; Vahey et al., 2005; Williams et al., 2014; Wolff et al., 2016;
Assessing	The ability to evaluate data quality and credibility (e.g. fact-checking, checking sources of social media posts)	Crusoe, 2016; Deahl, 2014; D'Ignazio, 2017; Mandinach & Gummer, 2013; Grillenberger & Romeike, 2018; Montes & Slater, 2019; Pangrazio & Selwyn, 2019; Philip et al., 2016; Sweeper, 2014; Tygel & Kirsch, 2015; Vahey et al., 2005; Williams et al., 2014; Wolff et al., 2016;
Interpretation	The ability to interpret different data formats (e.g. graphs, infographics, interface features)	Crusoe, 2016; Deahl, 2014; D'Ignazio, 2017; Mandinach & Gummer, 2013; Grillenberger & Romeike, 2018; Montes & Slater, 2019; Pangrazio & Selwyn, 2019; Philip et al., 2016; Sweeper, 2014; Tygel & Kirsch, 2015; Vahey et al., 2005; Williams et al., 2014; Wolff et al., 2016;
Data creation	The ability to create data in different formats (e.g. creation of a blog post, social media post/hashtag, presentation)	Crusoe, 2016; Deahl, 2014; D'Ignazio, 2017; Mandinach & Gummer, 2013; Grillenberger & Romeike, 2018; Montes & Slater, 2019; Pangrazio & Selwyn, 2019; Sweeper, 2014; Tygel & Kirsch, 2015; Vahey et al., 2005; Williams et al., 2014; Wolff et al., 2016;
Data citation	The ability to cite data sources (e.g. text references, images sources)	D'Ignazio, 2017; Mandinach & Gummer, 2013; Pangrazio & Selwyn, 2019; Sweeper, 2014; Vahey et al., 2005; Wolff et al., 2016;
Data Management	The ability to store, encrypt and manage data in a safe and secure way	Crusoe, 2016; Deahl, 2014; D'Ignazio, 2017; Mandinach & Gummer, 2013; Grillenberger & Romeike, 2018; Pangrazio & Selwyn, 2019; Sweeper, 2014; Tygel & Kirsch, 2015; Vahey et al., 2005; Williams et al., 2014; Wolff et al., 2016;
Data Visualization and Manipulation	The ability to represent data in a visual way (e.g. using infographics)	Crusoe, 2016; D'Ignazio, 2017; Mandinach & Gummer, 2013; Pangrazio & Selwyn, 2019; Philip et al., 2016; Sweeper, 2014; Tygel & Kirsch, 2015; Williams et al., 2014; Wolff et al., 2016;
Data Deletion	The ability to delete data (e.g. deletion of cookies, browsing history)	
Ethical use	The ability to use data ethically (e.g. not sharing someone's else personal data, not manipulating or mis-quoting data, anonymising people's identity)	D'Ignazio, 2017; Mandinach & Gummer, 2013; Grillenberger & Romeike, 2018; Pangrazio & Selwyn, 2019; Philip et al., 2016; Tygel & Kirsch, 2015; Williams et al., 2014; Wolff et al., 2016;

Data Doing can be defined as a set of core data literacy skills essential for thoughtful and informed engagement with data within a ‘datafied’ society. They are more than just “rote learned skills” and are core to a definition of data literacy. We would argue that Data Doing involves citizens’ abilities to not only appropriately manage data but to do so in an ethical and critical manner.

2.9.2 Data Thinking

Data Citizenship draws on Paulo Freire’s (1970) concept of *conscientização*– or critical consciousness. Such critical understanding aims to empower citizens to reflect on their social and data reality “not as a closed world from which there is no exit, but as a limiting situation which they can transform” (1970, p.49). Thus, Data Citizenship as either civic engagement or the making of digital rights claims is dependent on more than knowledge of the data processing infrastructure, but requires an engaged, critical, and active process of knowledge creation. Critical engagement with data services, requires one’s abilities to use, understand and create media and communication in a variety of contexts (Ofcom, 2018, p.2), including positioning data in different dimensions such as the political, cultural, and societal. As argued by Dencik et al. (2019):

The processing of data from across our lives can fundamentally shape social relations, the kinds of information valued and what is ‘knowable’ and therefore acted upon (p.873).

Data Thinking is particularly important to algorithmic discrimination, data breaches, and disinformation (Dencik et al, 2019; Zuboff, 2019). While algorithms and surveillance technology have become embedded into the structures of society (Dencik et al., 2019; Zuboff, 2019), it is essential for citizens to manage and critically think *about* and *through* data. Thinking *about* data involves citizens’ awareness of how and why data is being collected, analysed, and traded in multiple societal spheres. For example, a social media user might critically approach the way their data is being collected and what implications this might have in their life. This can include the way that social media prioritise sensational content to increase engagement and hence promote more disinformation, conspiracy and hate speech on peoples’ newsfeeds. In other words, understanding that algorithms are not neutral or objective and that they are influenced by economic and political rationale.

Table 4: Data Thinking domain of the Data Citizenship framework

Data Thinking		
Domain	Description	Literature References
Awareness of data protection rights	Being aware of local (UK) or regional (The European Union General Data Protection Regulation) data protection laws.	Crusoe, 2016; Pangrazio & Selwyn, 2019
Communicating with data	The ability to refer to, contextualise and use data for communication. Including providing evidence to validate an argument, on social media, in research or everyday interaction.	Crusoe, 2016; Deahl, 2014; D'Ignazio, 2017; Mandinach & Gummer, 2013; Grillenberger & Romeike, 2018; Montes & Slater, 2019; Pangrazio & Selwyn, 2019; Philip et al., 2016; Tygel & Kirsch, 2015; Vahey et al., 2005; Williams et al., 2014; Wolff et al., 2016;
Critical data analysis	The ability to consider, examine and discuss data bias, methodological errors, inaccurate data visualisation including data bias and cultural contexts.	Crusoe, 2016; D'Ignazio, 2017; Mandinach & Gummer, 2013; Grillenberger & Romeike, 2018; Montes & Slater, 2019; Pangrazio & Selwyn, 2019; Philip et al., 2016; Tygel & Kirsch, 2015; Williams et al., 2014;
Data safety	The ability to consider and implement a data protective step when using data. Such as using private browsing features or more secure browsers, using more secure search engines strong passwords, or skills to manage and control 'digital traces'.	Crusoe, 2016; Pangrazio & Selwyn, 2019; Philip et al., 2016; Tygel & Kirsch, 2015
Privacy	The ability to consider and implement privacy-protective behaviour when using data; for example using avatars, deleting tweets every couple of weeks.	Crusoe, 2016; Grillenberger & Romeike, 2018; Mandinach & Gummer, 2013; Pangrazio & Selwyn, 2019; Wolff et al., 2016;
Problem-solving using data	The ability to search for, identify and use data for solving problems.	Crusoe, 2016; Deahl, 2014; D'Ignazio, 2017; Grillenberger & Romeike, 2018; Mandinach & Gummer, 2013; Montes & Slater, 2019; Pangrazio & Selwyn, 2019; Philip et al., 2016; Tygel & Kirsch, 2015; Vahey et al., 2005; Williams et al., 2014; Wolff et al., 2016;
Understanding Data Society	The ability to understand the way data economy works. Including specifics of how platforms are funded, what cookies are, broadly what algorithms do, as well as the broader impact, procedures, and power-dynamics of platforms.	Crusoe, 2016; Montes & Slater, 2019; Pangrazio & Selwyn, 2019; Philip et al., 2016; Tygel & Kirsch, 2015;
Understanding of data collection	The ability to understand different data collection practices of different institutions. Including governments, advertising organizations, and data brokers. As well as the different databases such as platforms, National Health Service (NHS), local government voters registers and so forth.	Crusoe, 2016; Grillenberger & Romeike, 2018; Pangrazio & Selwyn, 2019; Sweeper, 2016

The importance of ‘critical abilities’ in making informed decisions about the ways in which data is used is highlighted by the European General Data Protection Regulation (GDPR). GDPR states that individuals have the right to be informed about the collection and use of their personal data (ICO, 2018). Data literate citizens should, therefore, be informed about their digital rights. While the GDPR emphasises the personal data realm, it is useful to consider citizens' understanding of collective data practices. This collective understanding is needed as citizens live in “networked publics” (boyd, 2014) - consisting of citizens’ active webs of information exchange (e.g. social media). In networked publics, data are aggregated, analysed, re-analysed and guide societal responses (Pangrazio & Selwyn, 2019, p.420).

The process of data decoding (an analogue of ‘media decoding’ as proposed by Hall (1980)) requires critical data literacy abilities such as solving problems with data (Wolf et al., 2016), communicating using data, and the development and evaluation of data-based explanations (Philip et al., 2016). An example of problem-solving with data, that also includes making digital rights claims, might be the collection and analysis of geo-located data on community facilities as part of a local political campaign. A ‘digital rights’ claim here might be a demand for access to this data (e.g. through Freedom of Information Act). Thus, Data Thinking can be compared to what Mandinach and Gummer define as “drilling into data” - the process of critical examination of data collection and the context in which data is collected and (re)used (2016, p.372). In our proposed Data Citizenship framework, seven domains can be distinguished as Data Thinking (Table 4).

*Ideally mindful engagement and understanding of data – **Data Thinking** - should be at the centre of citizens' decision-making processes. Whether it is buying a product online or posting an image on a social networking platform, citizens need to be able to understand that data is part of various social and economic processes presented in specially designed interfaces. Data literate citizens should use their critical skills as they view and analyse the world through data.*

2.9.3 Data Participation

While Data Thinking and Data Doing primarily focus on citizens’ abilities to interact with data, Data Participation examines the collective and interconnected nature of data society (boyd et al., 2014). It seeks to address the problem of citizens’ feelings of disempowerment in data-driven environments and the unequal power dynamics between the data processors and the data subjects. This is the key third element of our Data Citizenship model and aims to emphasise the importance of citizens’ proactive engagement with data and data society structures. This engagement has the potential to foster and sustain collective

civic action to challenge and reclaim these structures, conventions and prescriptions – to make ‘digital rights claims’ as Isin and Ruppert would argue.

Data Participation is crucial in increasingly ‘datafied’ and algorithms-reliant societies in which citizens are assessed as ‘data points’ made of continuously changing networked data flows resulting in discriminatory, unequal and overly-simplified views of specific groups in society (boyd et al., 2014, Noble, 2018). Barocas and Selbst argue that data mining – based on aggregated data, algorithms, and predictive analysis “is always a form of statistical (and therefore seemingly rational) discrimination” (Barocas & Selbst, 2016, p.674). Data mining means that citizens are algorithmically classified according to offline identifiers (e.g. age, ethnicity, geographical location) that have been extracted from their online behaviour, which subsequently are aggregated in datasets sold to third parties and public institutions to inform policy decisions (see Dencik et al. 2018 for examples of citizen scoring in UK public services), profile and target people, and predict future behaviours (Zuboff, 2019). The quality and reliability of aggregated data processing and analysis, as well as its impact on society have been widely put into question in the literature (boyd et al., 2014; Barocas & Selbst, 2016; Mittelstadt, 2017; Zuboff, 2019). In this context, it is crucial to put this collective and societal dimension at the forefront of discussions and educational initiatives – it represents a key critical aspect of data literacy and core practices for Data Citizenship.

Through Data Participation citizens are able to seek opportunities to exercise their digital / data human rights as well as contribute and shape their collective data experiences. Data Participation understood as a civic and/or collective action has been discussed and put into practice by data activists and scholars (Gutierrez, 2018; Kennedy, 2018; Lehtiniemi and Haapoja, 2019; Milan and van der Velden, 2016). Indeed, data activism often aspires to explore and draw on new forms of civic engagement to respond to datafication and enhance social justice in the present but also foster and envision alternative and more responsible futures (Beraldo and Milan, 2019; Kennedy, 2018, Lehtiniemi and Ruckenstein, 2018; Milan and van der Velden, 2016). However, this work has tended to remain and be perceived as the remit of specific communities of technical experts and activists without necessarily engaging non-expert citizens (Kennedy, 2018). Our conception of Data Participation encompasses the aspirations of data activism and aims to embed them as part of the Data Citizenship Framework. To do so, our concept of Data Participation combines our analysis of data literacy literature (Pangrazio & Selwyn, 2019; Wolf et al., 2019) with ideas from digital citizenship and democratic education (Freire, 1970/1996). This strand of literature positions citizens as active actors and co-constructors of their realities. In line with Freire’s vision of an active and empowered citizen.

Our key elements of Data Participation are laid out in Table 5. We argue that citizens’ pro-active participation with the structures of data society is essential to protect civic rights and liberties and to

enable active digital citizenship. Data literacies and Data Citizenship should aim to support citizens' agency and encourage them to question and resist autocratic data structures, so as to:

... enable communities to decrease dependencies on specialists to operationalise on data, and to increase informed actions and general agency with data (Wolf et al., 2019, p.3)

Table 5: Data Participation domain of the Data Citizenship framework

Data Participation		
Domain	Description	Literature
Participating in society using data	The ability to utilise data for societal participation and civic action (e.g. citizen-led campaigns, using online government services such as the NHS, Her Majesty's Revenue and Customs (HMRC)).	Crusoe, 2016; Montes & Slater, 2019; Pangrazio & Selwyn, 2019; Williams et al., 2014
Engagement with data society debates	The ability to engage in debates on data protection rights or/and Internet Governance (e.g. engagement in privacy or/and misinformation debates)	Crusoe, 2016; Pangrazio & Selwyn, 2019
Data Activism (pro-active engagement with data structures, including data hacking)	The ability to take pro-active steps to protect ones personal and collective privacy and wellbeing in the data society (e.g. reporting inappropriate or 'fake' content online, blocking or mitigating data collection using apps such as Add Blocker). The ability to collectively promote and exercise digital rights (e.g., using obfuscation or collective group uses of social media accounts). The ability to object/resist/modify the hegemonic way of using data services. The use of open-data to improve citizens' environments and to hold public institutions and private corporations accountable.	Pangrazio & Selwyn, 2019; Montes & Slater, 2019; Williams et al., 2014
Supporting others with their data literacy	The ability to help others with their data literacy (e.g. helping others with their privacy settings, explaining to people what clicking 'consent' means)	

Through **Data Participation** we stress that the ways to tackle the big data divide are not to "integrate" citizens into the 'datafied' structure, but to enable them to investigate, challenge, negotiate, protest, act upon and transform the structure so that they can become conscious data citizens. Both collective civic action and engagement in the governance of the data society is crucial to address the problem of the normalisation of data processors-subjects reality.

2.10 Conclusion

In our Data Citizenship model we have outlined three core domains of citizens' data literacy needed to support Data Citizenship:

1. **Data Doing** (citizens' critical data handling and management)
2. **Data Thinking** (citizens' critical understanding of data)
3. **Data Participation** (citizens' proactive engagement with data in their everyday lives)

Our analysis was framed within the wider discourse on the power imbalances in relation to data in contemporary society between 'data processors' and 'data subjects'. Especially the subsequent powerlessness citizens may experience in the context of digital and big-data divides. We have brought to the analysis key ideas from the theory of democratic education. This has allowed us to link ideas of skills, literacies and participation with the critical awareness needed to engage with and challenge current data structures/conventions/prescriptions. We see this as necessary as there is limited emphasis in the literature on the value of critical and Data Participation by citizens, at both an individual and a collective level.

As we noted at the start of section 2 the concepts of "Digital citizenship/Digital literacy" and "Data citizenship/Data literacy" are highly intertwined. Clearly data are central to our digital society and are the 'fuel' for digital systems. We do not offer here a sharp distinction between data and digital literacies nor citizenship. As we noted before, we consider these as complimentary overlapping sets of ideas or better still, two sides of a coin. In this sense, reflecting more the perspective of the researcher and educator and the aspects of our digital/'datafied' society they are examining. Importantly, in both cases we see that core skills and knowledge are needed for robust literacies but that the ability to use these critically in both thinking about and participating in our digital/'datafied' society forms the basis of digital/Data Citizenship. Nor are we suggesting that specific skills define data literacies. In light of the continually shifting norms, practices and beliefs related to data society, it would be counterproductive to argue for a rigid set of skills for data literacies. Rather we would argue that the three domains of 'doing', 'thinking' and 'participating' can be understood relative to the digital, data and social context under examination. What counted as core digital or data skills in the mid-1980s, or mid-2000s would look very different from those that count now or may be important in future. But citizens will still need to 'do', 'think' and 'participate' in order to undertake social action, civic role or make digital and data rights claims. Thus, Data Citizenship should be seen as an open-ended and flexible framework, that should be revised and updated according to contemporary technology and legal changes, and it can be co-developed by scholars, data literacy educators, and data citizens.

Our Data Citizenship Framework is therefore intended to serve four goals:

- First, it provides a framework within which researchers can define data literacies and Data Citizenship relative to socio-legal context and technologies.
- Second, it provides a tool for researchers to then assess 'levels' and forms of data literacies.
- Third, it provides a route to explore data literacy with citizens either within qualitative research, or in action research contexts.
- Fourth, it provides a framework for thinking about data literacies education that is about more than just skills but considers critical assessment of and participation in data society as a key element. Importantly, with a focus on citizens making rights claims and undertaking civic actions that address the power relationships defined by data structures/conventions/prescriptions.

We would argue that this fourth goal, strongly linked to critical Data Participation, is a significant gap in the existing literature dedicated to digital and Data Citizenship whereby only a limited scholarship has focused on this. We would argue that further work is needed to understand, address, and develop educational responses to support Data Citizenship and data literacy education in the context of the big data divide.

We have argued above that a developed form of Data Citizenship might be a fundamental quality of critical consciousness in contemporary society. It is a proactive mind-set, and an intellectual and practical engagement with a key feature of current social reality and social structure. Thus, democratic education to support the development of data literacies is not a static and top-down process, nor is it primarily concerned with getting a specific set of skills or expertise. Rather it is about the personal and intellectual tools to both make digital rights claims and to engage in digital civil society and action. Taking our lead from the work of democratic educators, Data Citizenship emphasises the importance of individual and collective knowledge creation, collaboration, and action to support pro-active citizenship. We believe that each citizen should be provided with access to relevant Data Citizenship education.

3 A national survey view of Data Citizenship

3.1 Introduction

This section describes the results of our national survey. We first use the data to identify different types of users of digital systems and media and their demographics. We then use these ‘user types’ to explore the responses to survey questions and topics. We bring these results together through a multiple correspondence analysis and regression model to explore the demographic factors underpinning Data Citizenship and literacy. From these results we build ‘personas’ for each of our user types to help support both policy interventions and the design of our focus groups. Data tables to go with graphs and statistical models and additional analytic figures not within the body of this section can be found in the Methods and Results Annex (section 9).

3.1.1 Survey design and administration

We designed our survey to operationalise our three domains of Data Doing, Thinking, and Participating into concrete research tools for both quantitative survey work and qualitative exploration with individuals and groups. We developed and conducted a survey of UK population (Yates et al, 2020a) focused on linking issues of data literacy and citizenship to key demographics and types of internet user (Carmi et al, 2020a; 2020b; Yates et al, 2015; 2015; 2020b; Yates and Lockley, 2018). Critical Research undertook the survey field work in summer and autumn of 2019. The sampling frame (see section 9.1) included 125 sampling points to achieve n = 1,542 completed interviews. These points were selected to be a representative cross section of UK addresses. Quotas were set to be reflective of the UK population by age, gender, and household socio-economic group, and urbanity.

The survey was designed to allow demographic comparison of key national data sets – especially the Ofcom Media Literacy survey. We also identified a range of prior survey research, both academic and third sector, that had addressed similar issues. These surveys provided a starting point and potential points of comparison for our survey design. We clustered prior survey questions, key variables and key behaviours under the headings detailed in Table 6. Potential questions were then reworked and revised with a focus on consistency of language and coverage of key elements in the Data Citizenship model. Questions went through several rounds of pilot testing with teams, colleagues and finally members of public. The draft questionnaire was then tested and critiqued by our survey partner Critical Research. This included rewording and supporting text for clarification. A final round of revision to address timing and administration of survey required rationalisation of the total number of measures. Critical Research then prepared the questionnaire for computer-assisted personal interviewing. The items within the survey are detailed in Table 6.

Table 6: Topics and measures for Data Citizenship survey

Variables	Measures
Core demographics	Location, Age, Gender, NRS Social Grade, Work status (S1 to S7).
Digital media and systems access	Devices owned by ten types (including none), mobile phone ownership, smartphone ownership (Q1 to Q3).
Digital media and systems use	Ever go online, devices used to go online, time spent online, confidence online (Q4 to Q7).
Type of online activity	Twelve activities: ever done online, ever done in last week.
Trusted features of digital media and systems	Eight methods to evaluate trustworthiness of online content (Q10).
Social media use	Three items: Ever use social media, number of social media accounts, frequency of posting on social media (Q11 to Q13).
Verifying social media content	Thirteen items: Trust in content and twelve methods to check content (including not to check) (Q14 and Q15).
Search engine use	Types of search engine used (Q16).
Verifying search engine content	Thirteen measures: Trust in content and twelve methods to check content (including not to check) (Q17 and Q18).
Types of information shared/collected when people are online	Thirteen items: including no collection at all (Q19).
Reasons platforms collect data	Thirteen items including no reasons (Q20).
Acceptance of data uses by platforms	Seven items: 5-point Likert acceptability scale and 'Don't know' (Q21).
Trust in platforms and organisations	Eleven items: 5-point Likert trust scale and 'Don't know' (Q22).
Comfort with third party data sharing by platforms	Nine types of data including none (Q23).
Control over data sharing	Five items: 5-point Likert agreement scale and 'Don't know' (Q24).
Data protection confidence	Seven items: 3-point confidence scale and 'Don't know' (Q25).
Data Participation behaviours	Eight behaviours (three personal, three with others) including none (Q26); ten data collection behaviours (Q29); five data analysis and editing behaviours including none (Q30).
Trust in news content online and offline	Four items: 5-point Likert agreement scale and 'Don't know' (Q27).
Trust in social media and content from friends	Five items: 5-point Likert agreement scale and 'Don't know' (Q28).

3.2 Defining users of digital systems

Following a closely similar methodology to that employed in our recent studies of digital inequalities (Yates et al, 2020a, Yates et al, 2015; 2019; 2020b; Yates and Lockley, 2018) we used Latent Class Analysis to

group our survey respondents according to their use of digital systems and media. Figure 3 shows the likelihood of each group undertaking specific digital activities.

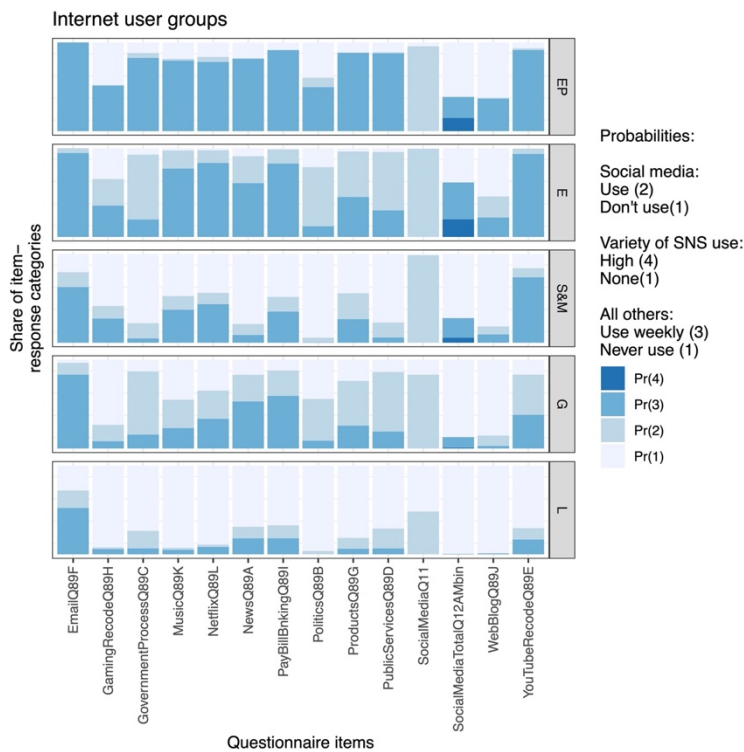


Figure 3: Latent class groups and probabilities of undertaking digital media and systems use

Our survey identified six groups as described in Table 7. These results are comparable to the findings in (Yates et al, 2015; 2019; Yates and Lockley, 2018). Due to space constraints our survey contained slightly fewer digital ‘use’ items than the Ofcom data used in prior studies. As a result, this analysis conflates the two ‘limited users’ groups previously identified into a single group and does not as strongly differentiate respondents out of the ‘General users’ category (Table XX).

Table 7: User types

User group	Description
Extensive political users	Very likely to undertake all forms of online activity including political or civic action
Non-political extensive users	very likely to undertake all forms of online activity but not as likely to undertake political or civic action
Social and entertainment media users	Low levels of overall use with a focus on social media and entertainments media
General users	General use with a focus on functional activities but limited social media use
Limited users	Limited use across all activities
Non-users	Do not make personal use of digital media or systems

3.2.1 Demographics of different “user types”

We have explored the demographics of our five digital user types. Using a multinomial regression. The demographic predictors of group membership are very similar to the prior research. We find that ‘Limited users’ are older, less likely to have a post-18 education and to be from lower socio-economic groups (NRS Grades D&E). For reference see Table 8 for an explanation of NRS grades. Similarly, our ‘Social and entertainment media’ users are also likely to lack a post-18 education and be from similar lower socio-economic groups (C2, D&E). ‘Extensive’ users are most likely to have a post-16 education and come from NRS socio-economic grades AB. Table 12 to Table 15, Figure 4 to Figure 6 and Figure 51 to Figure 54 provide detail on these demographic results. These results are highly comparable to our prior analyses based on Ofcom data that provides more detail on digital behaviours and a broader set of demographics (Yates et al, 2020b).

Table 8: National Readership Survey (NRS) social grades

NRS social grade	Description
A	Higher managerial, administrative, or professional
B	Intermediate managerial, administrative, or professional
C1	Supervisory or clerical and junior managerial, administrative, or professional
C2	Skilled manual workers
D	Semi and unskilled manual workers
E	Casual or lowest grade workers, pensioners, and others who depend on the welfare state for their income

Drawing on our more detailed prior analyses of these groups (Yates et al, 2020b) we can provide the following characteristics of the groups as compared to ‘*Extensive political*’ users as a baseline:

- *Non-political extensive users* are 2.9 times more likely to be 55+ than under 34, and are between 1.2, 2.4 and 3.6 times more likely to be in NRS social grades C1, C2 and DE respectively. They are more likely (4.0 times) to be retired. They also have a small likelihood (1.6 times) of being in deprivation and are slightly more likely to have left education at or before age 16 (1.6 times). They are slightly more likely to live in rural than urban areas (1.6 times).
- *General (no social media) users* are 4.3 times more likely to be 55+ than under 34 but show no significant statistical variance by NRS social grade. They are more likely (5.4 times) to be retired. They also have a small likelihood (1.3 times) of having left education at or before the age of 16.
- ‘*Social and entertainment media users*’ are 3.4 times more likely to be under 34 than over 55+, and are 1.4, 2.7 and 4.1 times more likely to be in NRS social grades C1, C2 and DE respectively. They also have

a higher likelihood (1.4 times) of being in high deprivation. They are 4.5 times more likely to have left education or planning to leave education before the age of 21, though they are the one group more likely to still be in education compared to extensive users (4.9 times).

- *'Limited' users (who use some social media)* are 3.5 times more likely to be 55+ than under 34, and are 1.7, 3.4 and 5.0 times more likely to be in NRS social grades C1, C2 and DE respectively. They are more likely (3.9 times) to be retired. They are 5.9 times more likely to have left education before age of 21. They are more likely to live in rural rather than urban areas (2.4 times). *'Limited' (who do not use social media) users* are 4.9 times more likely to be 55+ than under 34, and are between 1.6, 3.2 and 4.8 times more likely to be in NRS social grades C1, C2 and DE respectively. But they are much more likely (7.5 times) to be retired. They are likely to have left education before age 21 (4.7 times). Though they are not statistically significantly likely to live in more rural than urban areas.
- *Non-users* are the most distinct from extensive users. They are 10.2 times more likely to be 55+ than under 34, and are between 1.5, 3.1 and 4.7 times more likely to be in NRS social grades C1, C2 and DE respectively. But they are much more likely (8.4 times) to be retired. They have a high likelihood (2.9 times) of being in deprivation and are likely to have left education before age 21 (8.0 times). They are also more likely to live in rented or social housing (2.3 times) and more likely to be not working (3.4 times).

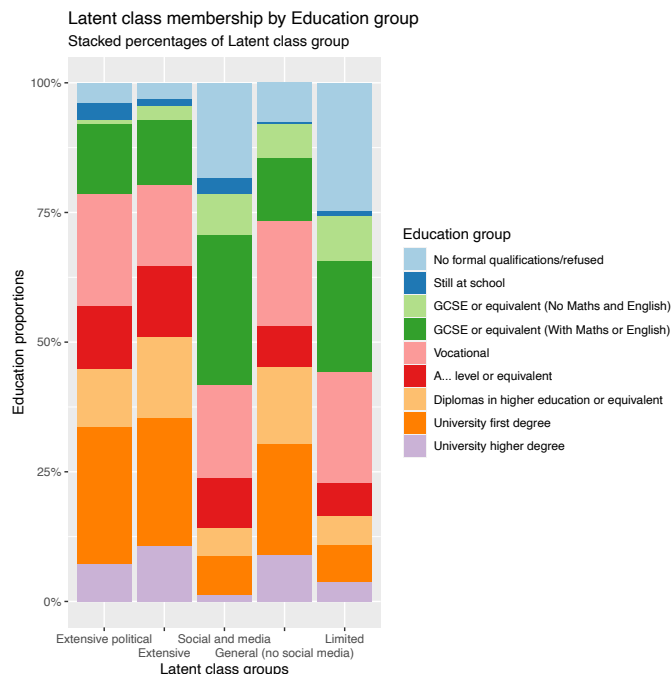


Figure 4: Latent class by Education

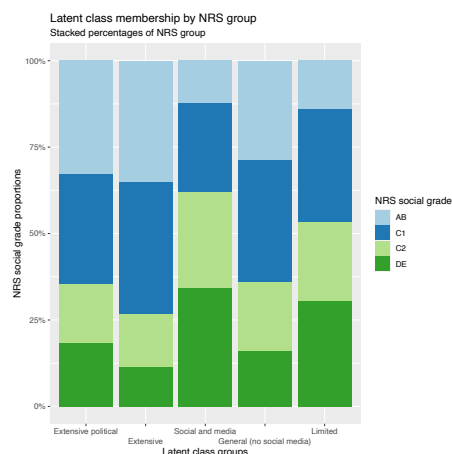


Figure 5: Latent class by NRS social grade

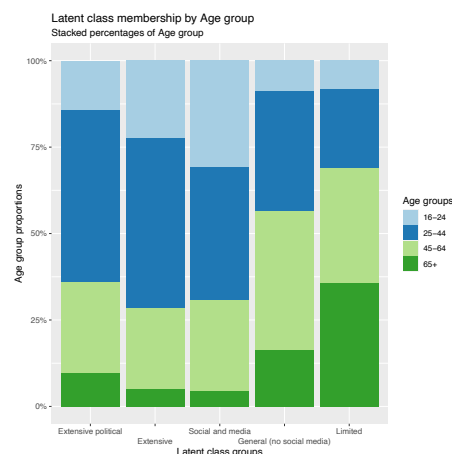


Figure 6: Latent class by age

3.3 How do our user types differ in terms of data literacies?

As noted in section 3.1.1, our survey collected data on 133 individual behaviours, items of knowledge and opinions designed to assess citizens data literacy. These varied in topic from questions about practical activity with data through questions about platforms use of data to issues of perception and trust. As noted above in section 2 current research indicates that data and digital literacy levels are not as high a might be assumed. Many citizens appear not to be aware of the types of data collected, uses to which it may be put, nor are they likely to use data in their everyday lives. In this section we explore how these measures vary across the five digitally active groups of our six user types.

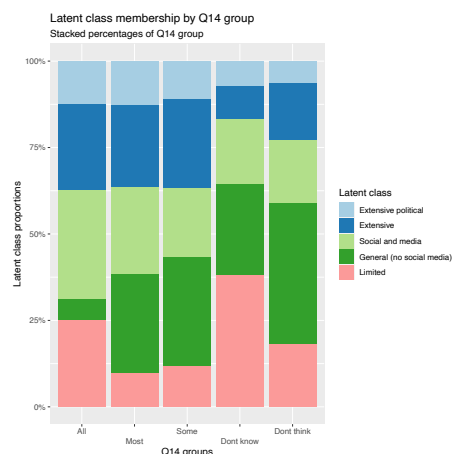


Figure 7: Latent class by trust in social media content

3.3.1 Checking data and content found online

If we look first that those in our survey who used social media, we find a pattern that is replicated throughout the analysis. We asked respondents if they trusted the information that they found on social media platforms or via similar apps. Figure 7 details the results. We find that most users only trust some of

the content. But when we ask respondents to indicate which of nine possible methods for checking content in social media they undertook, we find some notable differences between the groups. Most respondents are using less than 3 types of check (Figure 8 and Figure 55) with most common action being to check if the information was provided by a known or trusted organisation. Our 'Limited' users and our 'Social and entertainment media' user group are far less likely to check content (Table 16). They are more likely than other groups to use checks that rely on other people (trust in the poster, check with friend, check comments on post) than to evaluate the content (Figure 33). This pattern of a general concern about or awareness of digital and data literacy issues, yet limited action in response to it, can be found across all the following issues we studied.

If we next look at search engine use, we find a similar pattern. Once again there is a big difference in the extent of checking between our two types of 'Extensive' users and the rest of the respondents. Yet, as above, the overall range of checking remains low across all groups. Checking of content online, especially facts, figures, news, or claims forms a key critical component of our Data Doing and thinking dimensions. The results clearly demonstrate two patterns. First, the range of methods of checking of data and information found online is low across all respondents in our data. Second, it is very low, if not absent for some groups. We will explore why this in our discussion of our focus group findings (see section 4).

3.3.2 Knowledge of data overtly shared or passively “given off” and the reasons behind it

We asked our respondents about the types of data that platforms of all kinds tend to collect about users. In other words, the data that we directly or indirectly share with platforms as we go about our daily digital lives. We also asked respondents about the reasons why platforms and organisations might collect this data. The majority of respondents have “some awareness” of data that is collected and some awareness of the reasons for this (ticked at least one correct answer in each case). Though once again our 'Social and entertainment' media user group and our 'Limited' user groups had a notable proportion who were unaware (predominantly “Did not know”). Figure 10 and Figure 11 and Table 18 and Table 19 detail these responses. When we look at the range of awareness, as measured by the number of data types and reasons noted by respondents, we find a similar pattern to that for checking data and content online. Respondents are more aware but are still only aware of half of the available options, with our 'Social and entertainment media' and our 'Limited' user groups again scoring lowest.

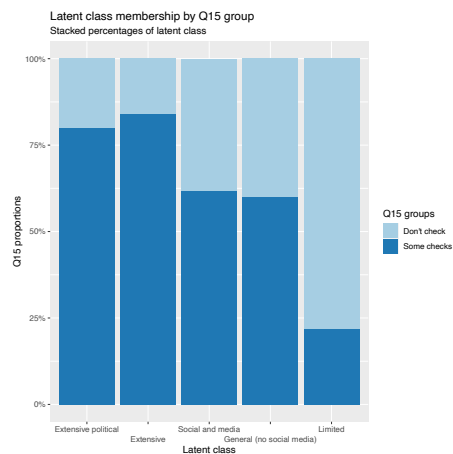


Figure 8: Latent class by likelihood of checking social media content

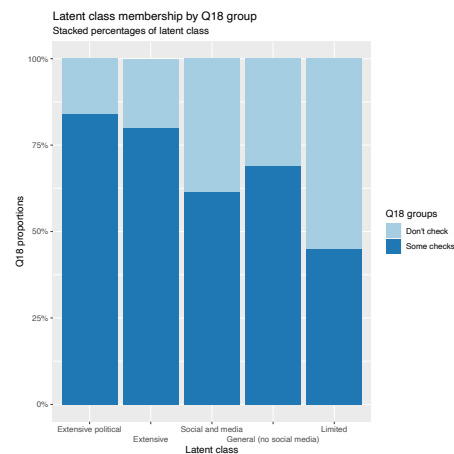


Figure 9: Likelihood of checking search engine results

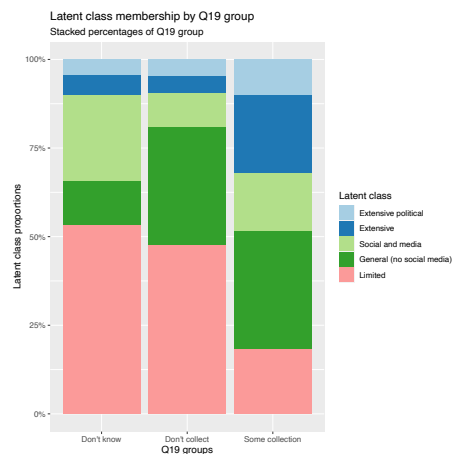


Figure 10: Latent class by awareness of data collected by platforms

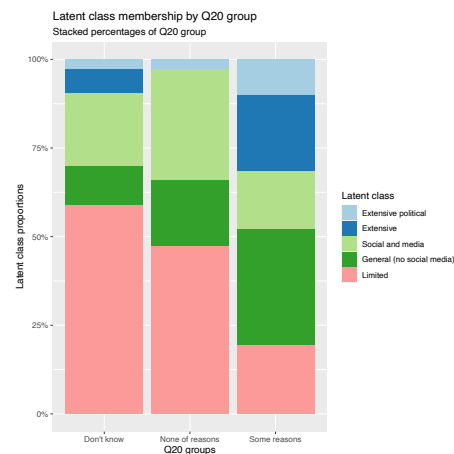


Figure 11: Latent class by awareness of reasons for data collection

3.3.3 Acceptability of data collection and third-party sharing of data by platforms

If we next look at how acceptable data collection is for respondents, we see clear variation between context and user types on some but not all measures. There is a clear distinction between respondents accepting the collection of data for their consumer benefit but not accepting the use of collected data to benefit the platform or company. This is a very understandable position but contains the inherent contradiction that to deliver the benefits that consumers find acceptable, platforms and companies have to (they would argue) undertake the behaviours that respondents find less acceptable (Figure 12 to Figure 15). This disparity between higher levels of acceptance for the benefit of the consumer vs benefits platforms holds for all user types (Figure 59).

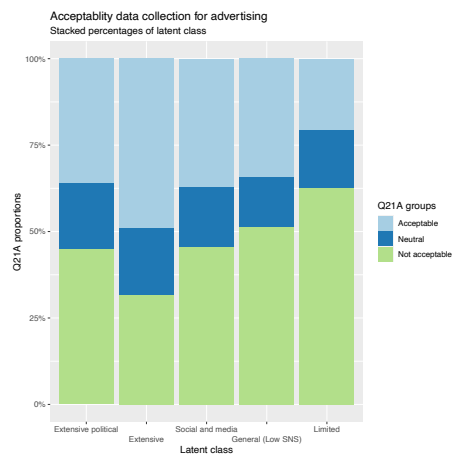


Figure 12: Acceptability of advertising data collection

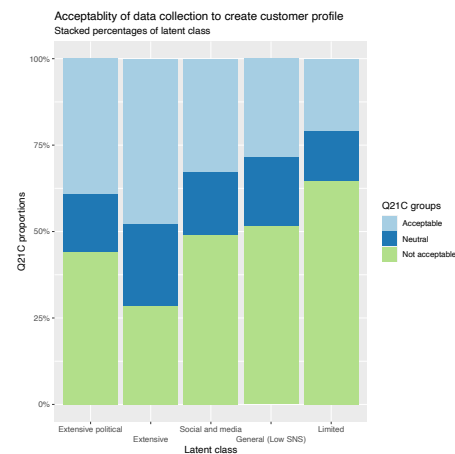


Figure 13: Acceptability of profiling data collection

This disparity potentially implies two things. First, it may imply a lack of understanding of the links between the uses to which platforms put citizens data and platforms' ability to deliver personalisation or tailored services. Second, it may imply a response to the inability to make "digital rights claims" in relation to what platforms do with their data. In particular, the ability to assert the right that platforms only use data to deliver specific citizen needs and that they do this transparently. That said, the fact that there is greater acceptance of the use of data to create customer profiles implies some understanding that some data is needed to make personalised services available. We will further explore what this differentiation means in our discussion of the focus group findings (section 4).

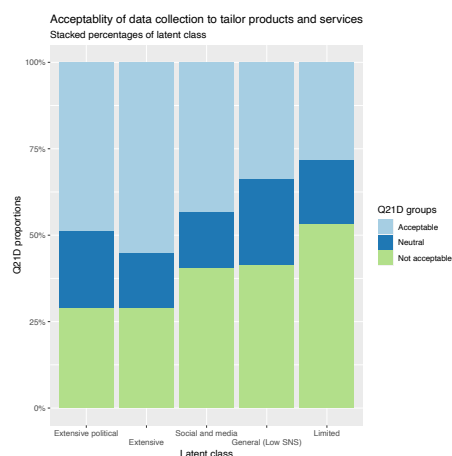


Figure 14: Acceptability of tailored services data collection

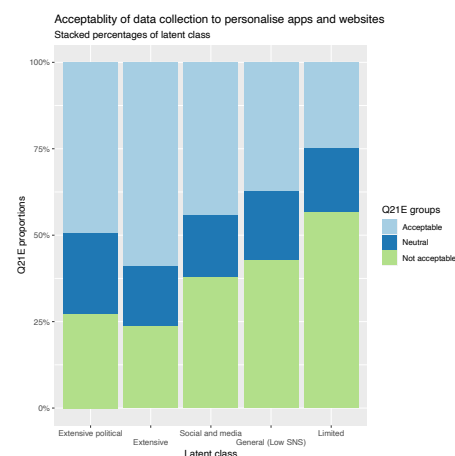


Figure 15: Acceptability of personalized apps data collection

Given the concerns most respondents have with platforms and companies' collection of data, we explored attitudes towards the "third party" sharing or selling of user data. We asked which if any kinds of data that platforms and companies regularly collect respondents would be comfortable being shared or sold to third parties. Once again, most respondents were uncomfortable with allowing any "third party" sharing or sale of data (see Figure 19 and Table 27). Where respondents were comfortable to share, the two least

concerning items were their name and email address (Figure 19). It is notable that 'Social and entertainment media' user group are most comfortable with third-party sharing of their data.

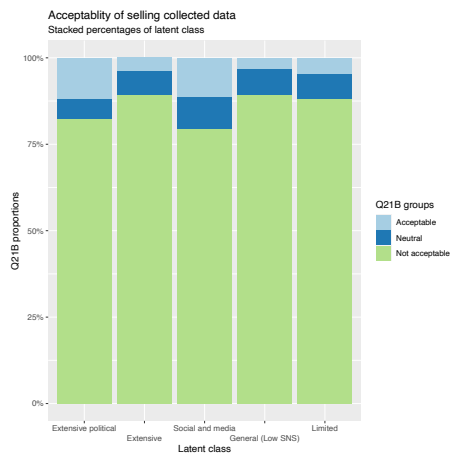


Figure 16: Acceptability of data collection for sale

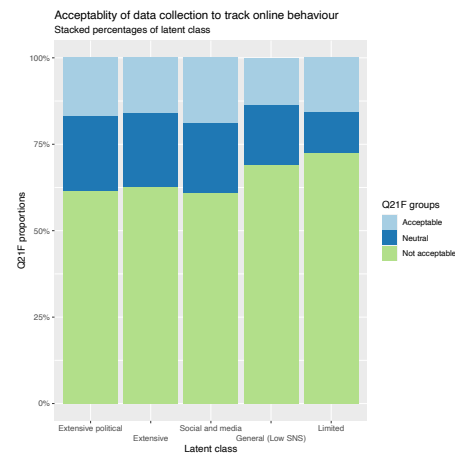


Figure 17: Acceptability of data collection for online tracking

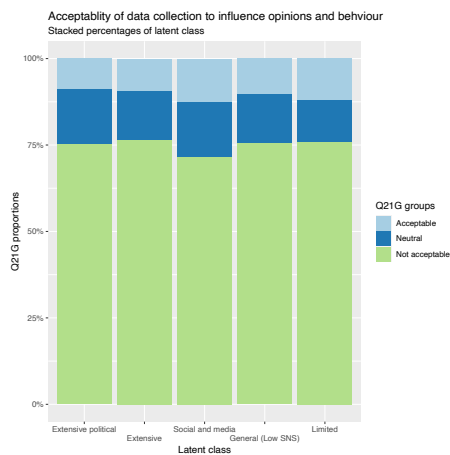


Figure 18: Acceptability of data collection to influence behaviour

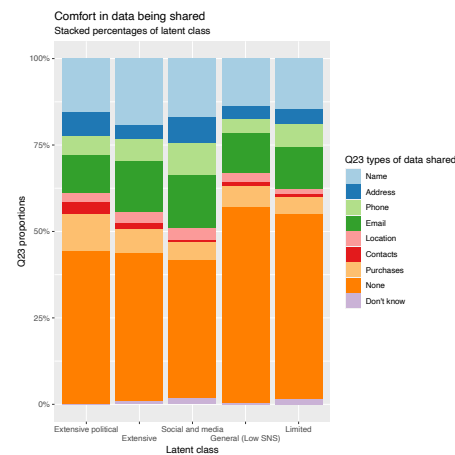


Figure 19: Comfort with different types of data being shared with 3rd parties

3.3.4 Attitudes to companies and platforms use of and protection of data

An argument is made, by some writers, commentators, and platform providers, that citizens are happy to share their data if the service is free. In fact, we find the opposite is true with the majority of our respondents disagreeing with this assertion (Figure 23). The majority of respondents are also clear that they feel that they “have no choice” but to share data with platforms in order to gain access (Figure 24). The majority of respondents also feel that companies don’t make managing privacy and data protection settings on platforms easy (Figure 22). When it comes to more personal actions respondents show greater efficacy. The majority do not think changing such settings is “too much effort” but they are split on whether making such changes is worthwhile (as companies and platforms will find ways round these) (Figure 20).

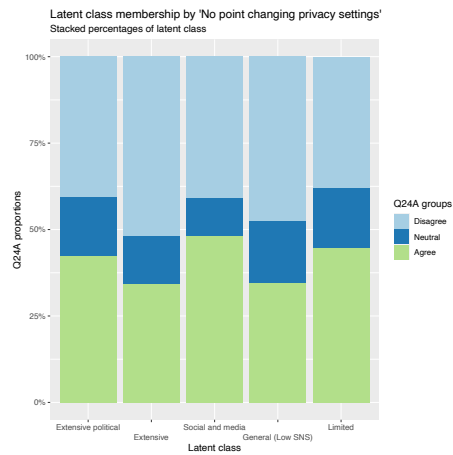


Figure 20: 'No point changing privacy settings'

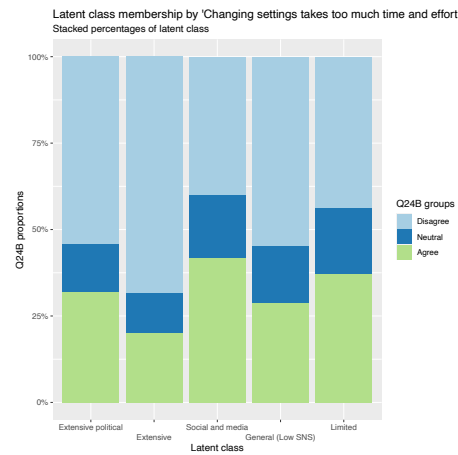


Figure 21: 'Changing settings is too much effort'

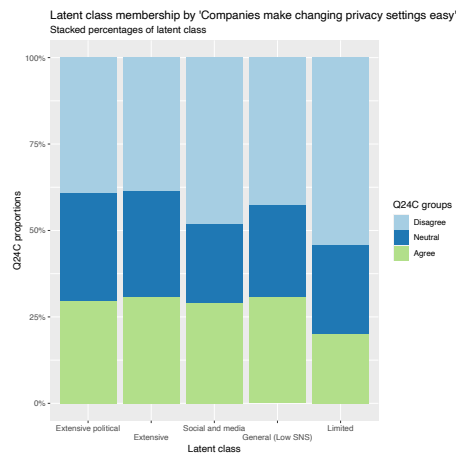


Figure 22: 'Companies make privacy settings easy'

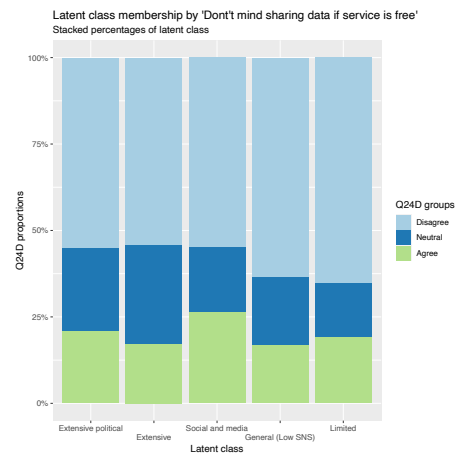


Figure 23: 'Don't mind sharing if service is free'

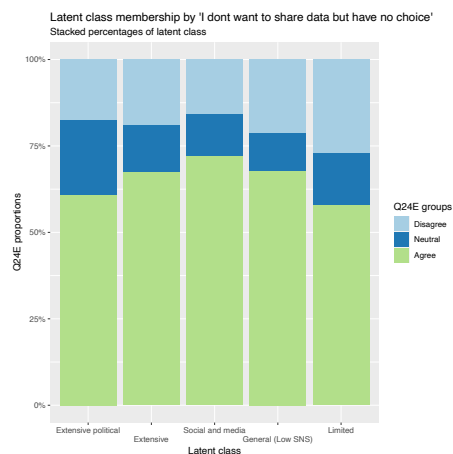


Figure 24: 'Don't want to share but have no choice'

3.3.5 Citizens confidence with their own data protection actions

We then asked about their confidence with individual actions that might help protect their data such as strong passwords or changing privacy or tracking settings we find a similar pattern. Our extensive users are

more confident than the other groups (Figure 60). That said on average they are only confident in 4 or 5 of the seven activities examined. All groups, other than 'Limited' users, show confidence in setting strong passwords (Figure 26).

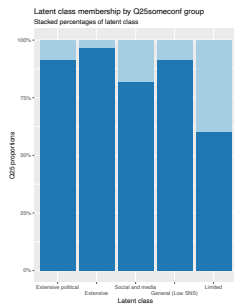


Figure 25: Q25 Some confidence at all ($\chi^2(4, 1322) = 176.203, p = 0.000, CV = 0.365, \text{Medium eff.}$)

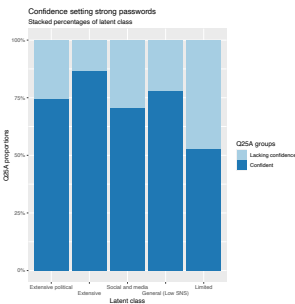


Figure 26: Q25A Strong password confidence ($\chi^2(4, 1322) = 90.873, p = 0.000, CV = 0.262, \text{Medium eff.}$)

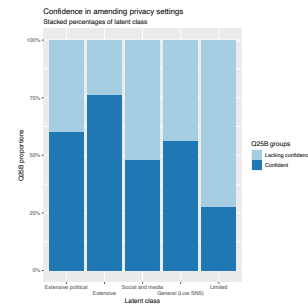


Figure 27: Q25B Confidence amending privacy settings ($\chi^2(4, 1322) = 138.82, p = 0.000, CV = 0.324, \text{Medium eff.}$)

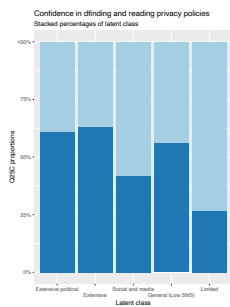


Figure 28: Q25C Confidence findings and reading policies ($\chi^2(4, 1322) = 99.92, p = 0.000, CV = 0.275, \text{Medium eff.}$)

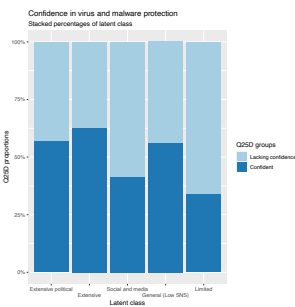


Figure 29: Q25D: Confidence with virus and malware protection ($\chi^2(4, 1322) = 63.81, p = 0.000, CV = 0.220, \text{Medium eff.}$)

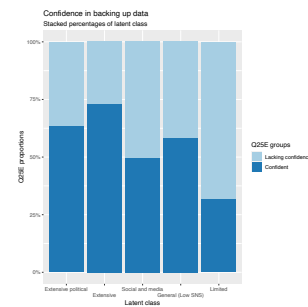


Figure 30: Q25E Confidence backing up data ($\chi^2(4, 1322) = 106.572, p = 0.000, CV = 0.284, \text{Medium eff.}$)

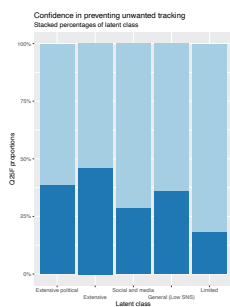


Figure 31: Q25F Confidence preventing tracking ($\chi^2(4, 1322) = 55.131, p = 0.000, CV = 0.204, \text{Medium eff.}$)

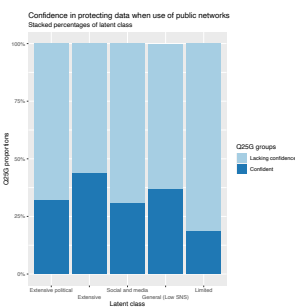


Figure 32: Q25B Confidence protecting data on public networks ($\chi^2(4, 1322) = 45.443, p = 0.000, CV = 0.185, \text{Medium eff.}$)

We then see a pattern of diminishing proportions of confident respondents across all remaining areas (Figure 27 to Figure 32). 'Extensive' users remain proportionally more confident but overall levels drop from 75% to 40% for 'Extensive' users and from 45% to 25% for our 'Social and entertainment media' user group. 'Limited' users hover between 20% and 25% of respondents being confident in all these activities (Figure 27 to Figure 32). For all groups the lowest levels of confidence are in protecting against data

tracking across devices and platforms and in protection of devices and data when on public wifi networks
(Figure 31 to Figure 32).

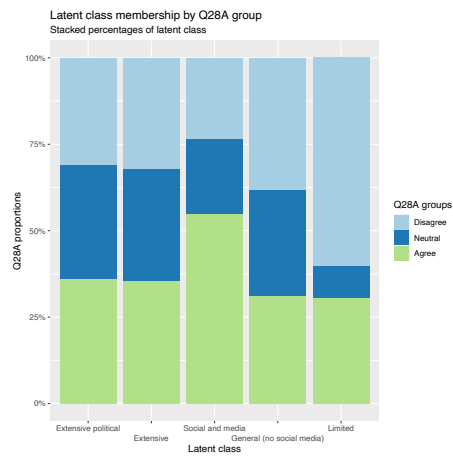


Figure 33: Trust in friends social media posts

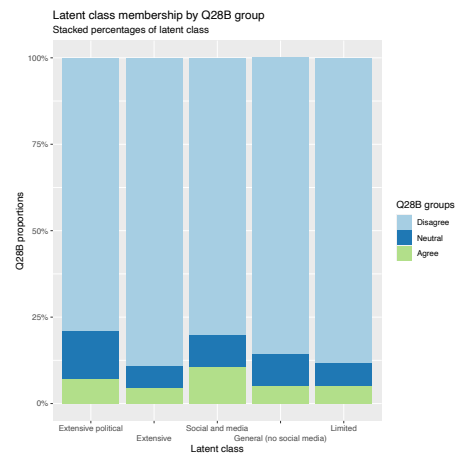


Figure 34: Overall trust in social media posts

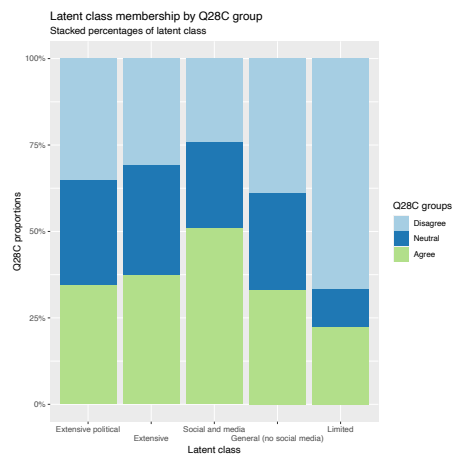


Figure 35: Mostly read social media that shares own values

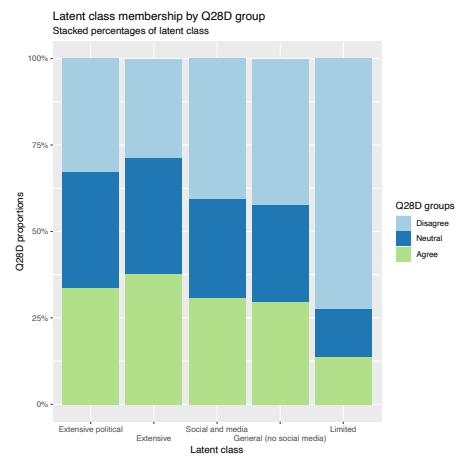


Figure 36: Read social media with different political perspectives

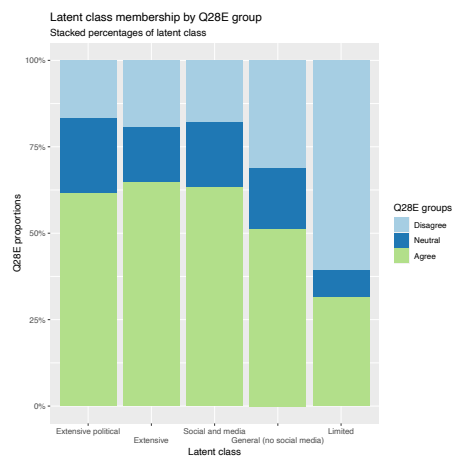


Figure 37: Information sources depend on social media links

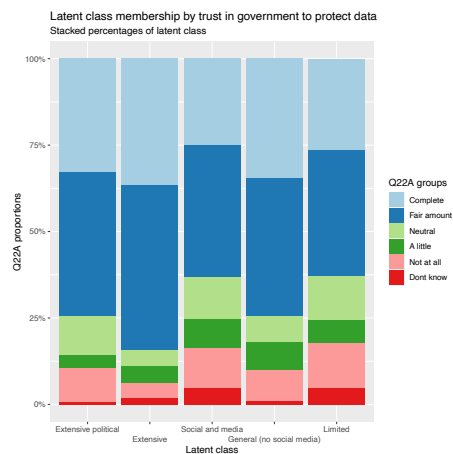


Figure 38: Trust in government to protect data by latent class ($\chi^2(20, 1322) = 60.375, p = 0.001$, Cramer's $V = 0.107$, medium eff.)

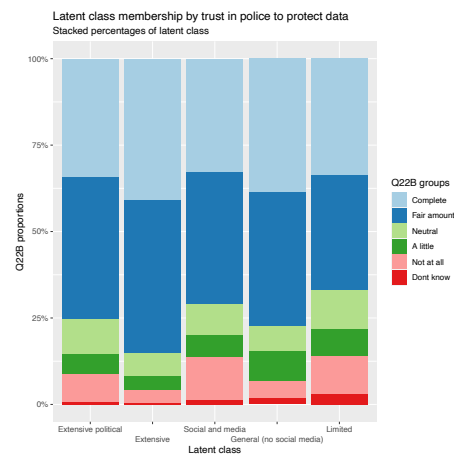


Figure 39: Trust in police to protect data by latent class ($\chi^2(20, 1322) = 44.573, p = 0.001$, Cramer's $V = 0.092$, small eff.)

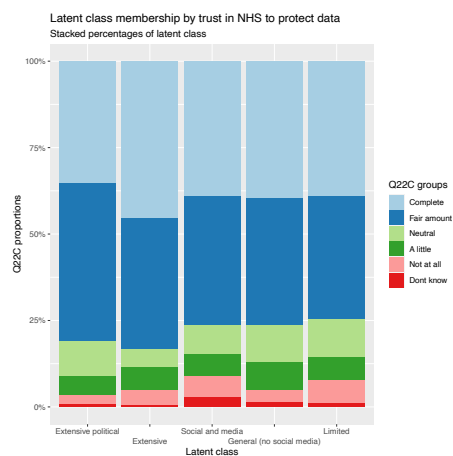


Figure 40: Trust in NHS to protect data by latent class ($\chi^2(20, 1322) = 25.098, p = 0.198$)

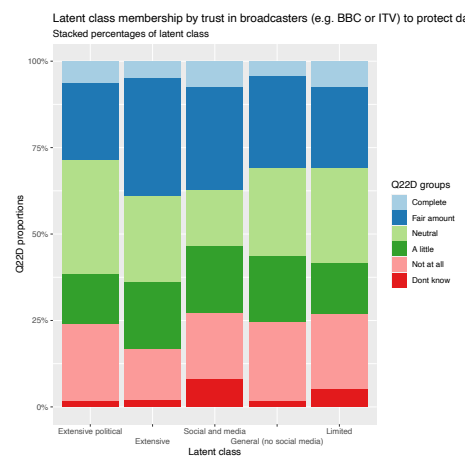


Figure 41: Trust in broadcasters to protect data by latent class ($\chi^2(20, 1322) = 93.262, p = 0.000$, Cramer's $V = 0.133$, medium eff.)

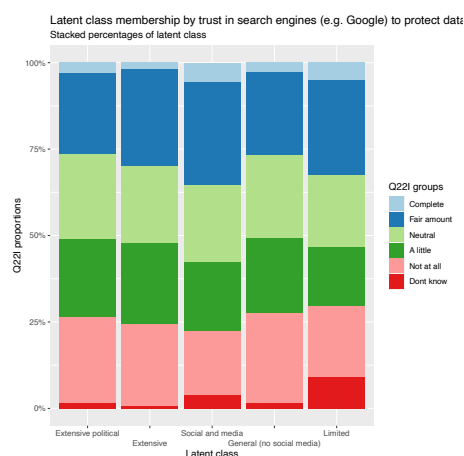


Figure 42: Trust in search engines to protect data by latent class ($\chi^2(20, 1322) = 58.488, p = 0$, Cramer's $V = 0.105$, medium eff.)

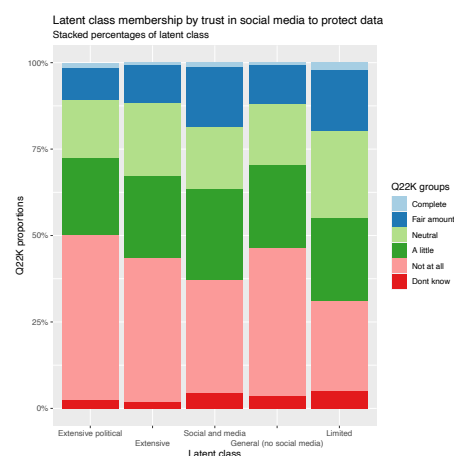


Figure 43: Trust in social media to protect data by latent class ($\chi^2(20, 1322) = 31.185, p = 0.053$).

3.3.6 Citizens trust in digital platforms and organisations

We asked respondents to rate how much they trusted different organisations to protect their data. Digital platforms such as social media and search engines had the lowest levels of trust. It is not surprising that most respondents had trust in major institutions such as the government, police, and NHS. It is interesting to note that both broadcasters and search engines generated very similar levels of trust. Social media had by far the lowest levels of trust. There was limited statistically significant variation among our user groups. Where it was present it reflected the higher levels of trust that our 'Extensive' users had in long standing institutions such as the police and NHS.

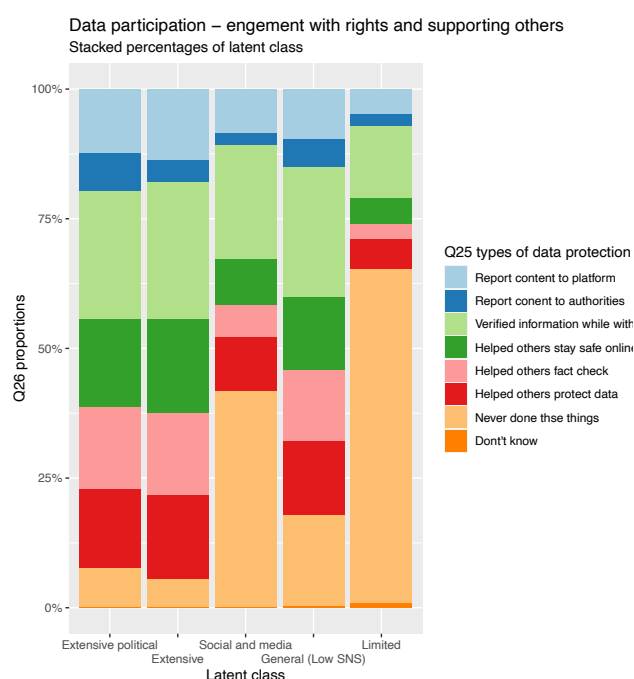


Figure 44: Data Participation – data protection activities

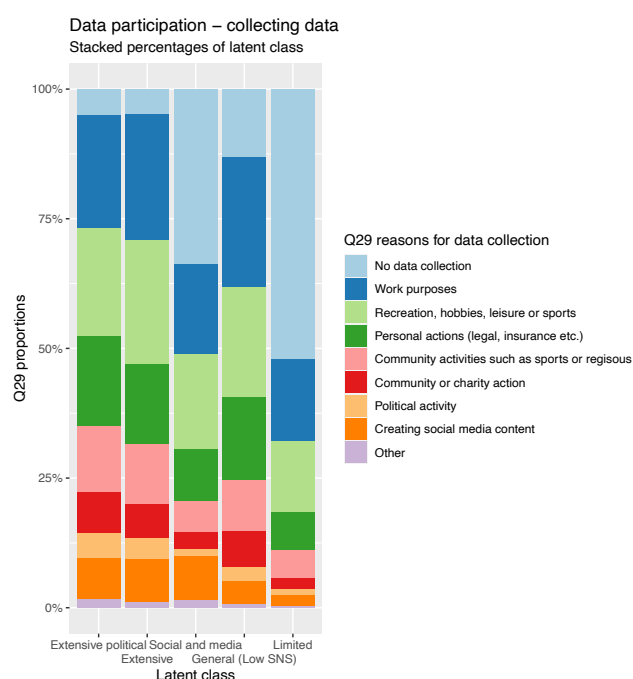


Figure 45: Data Participation - collecting data

3.3.7 Supporting others and using data

We asked respondents about activities such as supporting each other and using data to support their community, that fall under our Data Participation dimension. We looked at three main aspects:

1. Data Participation - actions that include engaging with digital rights or supporting others. These are individual actions that are indicators of proactive participation in our 'datafied' society:
 - a. Actions that respond to data and digital rights such as reporting harmful content or verifying information as part of social or community interaction.
 - b. Supporting others to protect their privacy or data online or to verify information
2. Using data in everyday life: from work, through personal or community activities to politics and digital media content creation. These split across personal activities, even if involving others, and community

activities. These activities include aspects of *Data Doing* in and of themselves as they require the collection and manipulation of data and information. They involved *Data Thinking*, especially in terms of the analysis or presentation of data. In terms of Data Participation they involve the active social and community use of information found online.

3. Whether or not the data had been analysed, re-edited or visually presented in any way.

Looking first at the collection of activities around engagement and support – Data Participation – we find that our ‘Limited’ and ‘Social and entertainment media’ user groups who are least likely to have undertaken any of the activities (Figure 44). The most common activity for all groups has been to verify, via the internet, data or information pertinent to ongoing interactions with friends, family or colleagues. Splitting the results by ‘rights’ and ‘helping’ actions we once again see that our ‘Limited’ and our ‘Social and entertainment media’ user groups are the least likely to engage in these activities (Figure 46 and Figure 47). Turning next to respondents’ active use of data for their personal community or civic activity the pattern reported in previous sections is repeated. Our ‘Limited’ and ‘Social and entertainment media’ user groups show almost no use of data for any of these activities. Though in this case even our ‘Extensive’ user groups average just three of the eight activities surveyed (Figure 44). Looking at the spread of activities, work and personal uses are the most common. Likelihood of manipulating data in some manner was lower across all groups with our ‘Limited’ and ‘Social and entertainment media’ user groups being the least likely to undertake such activity.

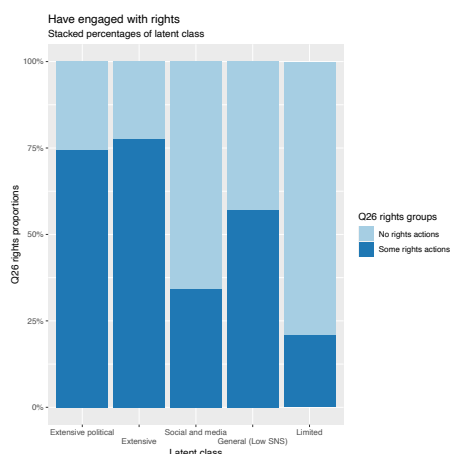


Figure 46: Data Participation - rights actions

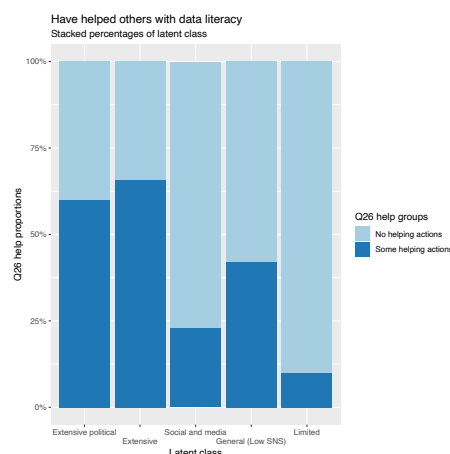


Figure 47: Data Participation - helping others

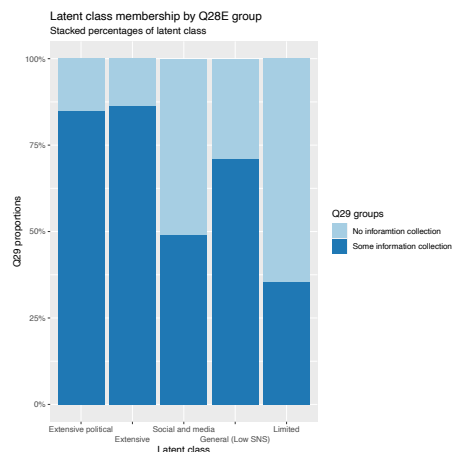


Figure 48: Data Participation - data collection

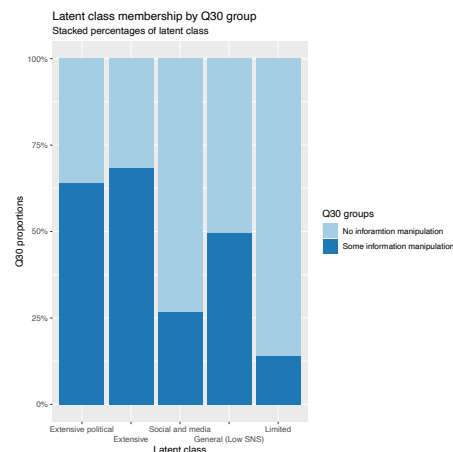


Figure 49: Data Participation - data manipulation

3.3.8 Measuring Data Citizenship

It is clear from the results presented so far in this section, that levels of Data Citizenship across our three dimensions of *doing*, *thinking*, and *participation* vary considerably across our five user types. But this variation is as much within as across our three dimensions. As section 3.3.7 indicates the analysis and manipulation of data – in general or for personal or civic action – is low for the majority of users. This combines aspects of *Data Doing*, *thinking* and *participating*. As section 3.3.2 points out, understanding of the ecology and economy of our 'datafied' society and understanding actions to mitigate use of data or risks are also unevenly distributed. This also combines aspects of *Data Thinking* and *Data Doing*. As we argued at the close of section 3 nearly all actions taken using or through digital media or systems involve some aspects of *Data Doing*, *thinking* or *participating*. We therefore looked to explore if the results from our various individual measures could be combined into a broader measure of Data Citizenship.

Given that much of our data is binary (actions were or were not done) or very polarised (strong agreement or disagreement) we applied multiple correspondence analysis (MCA) to the question results reported in section 3.3. MCA techniques allow for nominal categorical data to be examined in a manner akin to a principal components exploratory factor analysis. The results are graphical and can be used to inductively to detect and represent underlying structures in a data set (see Le Roux and Rouanet 2004; Blasius and Greenacre (eds.) 2006). As with factor analysis dimension scores can be allocated to individuals from their responses. We find four statistically relevant dimensions accounting for 96.3% of the variance in the data (Figure 61). Looking at the items that underpin these dimensions (Figure 62 to Figure 65) we interpret these as measures of:

1. Data literacy across all three elements
2. Trust in content from friends and on social media
3. Trust in and use of 'offline' media

4. Lack of engagement with nor neutrality on data literacy elements

Dimension 1 produces a positive score for low data literacy and a negative score for high. In the results and graphs we have reversed axes for this score to show increasing data literacy as positive and moving to the right. Differences in average scores by user types for all four dimensions are statistically significant (Figure 66 to Figure 69). Plotting the density of cases across dimensions 1 and 2 we can see how the combination of data literacy and trust in friends and social media function for three key variables (Figure 70 to Figure 73).

Figure 71 shows that NRS social grades AB most strongly correspond with higher data literacy and average to below average trust in friends, shared information, and social media. Conversely NRS social grades C2 and DE are more likely to correspond with lower data literacy and above average trust in friends shared information and social media. Figure 72 indicates that data literacy is notably lower in over 65-year-olds but that trust in friends, shared information and social media falls significantly for older groups (45+ years old). Figure 73 indicates that increasing data literacy corresponds with increasing levels of education.

Finally, we have undertaken a regression analysis of our data literacy dimension score against the following key demographic variables:

- Education
- Age
- NRS social grade

Given the strong correspondence of levels of data literacy with our 5 digital user types we were not surprised to find that all three variables are strong predictors of data literacy levels (Table 9, Figure 74 to Figure 76). These results indicate that younger respondents with a university education, in NRS social grades AB will score more highly than older people with lower levels of education of a lower socio-economic grade. Taking three possible respondents and applying our regression equation we find starkly different scores:

- A 30-year-old, higher university degree graduate in social grade AB would score on average 2.458 ($= -0.726 + 3.196 - 0.012 + 0$) in the upper quartile of our data (score above 2.256)
- A 24-year-old, with basic secondary school education in social grade DE would score on average 1.196 ($= -0.726 + 0.498 + 0 - 0.968$) in our second lower quartile (score between -2.256 and -0.167).
- A 65-year-old, with no formal education in social grade DE would score on average -3.590 ($= -0.726 + 0 - 1.896 - 0.968$) in the lowest quartile of our data (score below -1.945)

These scores are of course *relative* not absolute measures in relation to the data from our survey. As we have noted above, overall data literacy measures for individuals or collectively across groups are rarely very high or close to the potential maxima.

Table 9: Data literacy linear regression final model

Dependent variable: Data literacy score	
Education – baseline no-education	
Still at school	2.230*** (0.618)
GCSE or equivalent (No Maths and English)	0.498 (0.341)
GCSE or equivalent (With Maths or English)	0.671** (0.262)
Vocational	1.568*** (0.260)
A' level or equivalent	2.166*** (0.313)
Diplomas in higher education or equivalent	2.406*** (0.301)
University first degree	2.650*** (0.292)
University higher degree	3.196*** (0.365)
Age – baseline 16-25	
25-44	-0.012 (0.214)
45-64	-0.480** (0.219)
65+	-1.896*** (0.256)
NRS social grade – baseline AB	
C1	-0.113 (0.198)
C2	-0.891*** (0.233)
DE	-0.968*** (0.243)
Variables	
Constant	-0.726** (0.341)
Observations	1,322
Log Likelihood	-3,052.455
Akaike Inf. Crit.	6,134.911
Note: *p<0.1; **p<0.05; ***p<0.01	

3.4 Conclusions: personas

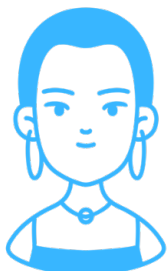
From our analysis of the survey data, we have identified some key characteristics of respondents in relation to data literacy:

1. We find that on many specific points of data literacy knowledge, practice, and awareness few respondents score highly. Except for a few more basic items, the proportions of any of our user groups showing specific knowledge or practice are rarely ever above 70%.
2. Therefore, only a small proportion of respondents have the deep and broad data literacy envisioned in sections 2 and 3.
3. Low levels of use ('Limited' users) and narrow use ('Social and entertainment media' user group) consistently correlates with low awareness and limited actions.
4. Responses to specific aspects of data literacy and our overall data literacy measure strongly correspond to our five user types.
5. Not surprising given this correspondence, data literacy scores are statistically correlated with key demographics such as education, age and socio-economic status (NRS social grade).

To understand these results better, and building on our five types of users, we have developed "Data Citizenship" personas. These 'ideal type' descriptions of users, their demographics and their specific data literacy capabilities and practices also provided a basis for our further qualitative workshops and our development of educational guidance. Sections 3.4.1 to 3.4.5 provide the core details for each persona.

3.4.1 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" too 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

3.4.2 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	

3.4.3 Social and entertainment media 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	
3.4.4 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

3.4.5 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 are 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

4 Understanding data citizens

4.1 Introduction

In this section we explore the findings from our focus groups. These were designed to allow us to further explore our survey insights, focusing on the understanding and experience of citizens as they navigate our digital and ‘datafied’ society. Originally these were designed to be workshops and involve engagement with awareness raising activities and more complex data collection tools. Sadly Covid-19 intervened and limited our ability to run face-to-face and more complex sessions. We ran the focus groups between September 2020 and onto April 2021. The focus groups were recorded using Zoom and fully transcribed. The team then undertook a thematic analysis of the focus groups informed by our theoretical framework and survey findings. This section provides an overview of the themes and presents examples of respondent comments that aligned with these themes. The section concludes with reflections on the focus group results that build on the survey and theory presented in sections 2 and 3.

4.2 Focus groups

All but two of our focus groups were recruited via local community centres and groups that were part of the Good Things foundation national network. This represents several thousand local centres that support digital inclusion intervention, support, and training. The final two groups were recruited from students attending a variety of universities across the UK. Table 10 and Table XX detail the groups and their demographics. The groups were designed to reflect the core demographics of our user types but with an over sampling of those groups with lower data literacies. We selected this balance to gain a greater understanding of the challenges experienced by these groups as they navigate digital systems and media. In addition, the literature review revealed that there are very few qualitative accounts focusing mainly on these groups.

Due to covid restrictions the original plan for longer workshop sessions with interactive elements could no longer be supported as this required close face-to-face interaction. During planning these focus groups, the UK moved out of the first lockdown and a socially distanced set of activities were designed. Unfortunately, the return of lockdown restrictions required a shift back to remote or mediated focus groups. Putting the health and wellbeing of our participants first, the team had to undertake more standard focus group interviews using Zoom and working with local centres to support respondents taking part in the sessions. Details of the methodological challenges of running focus groups with low-digital skill participants via remote or mediated interactions will be reported more fully in a forthcoming paper. The focus group interview schedule therefore had to be significantly reduced to fit the shorted time frame and the more

limited Zoom format. In particular, the mapping activities had to be reduced, a request to pause reflect and make notes on the questions. The focus group schedule is presented in Table 44.

Table 10: Focus group participants

Focus group	Target group	Likely User Types	Format
A	Younger people (<30 - preferably under 25) who may have some digital skills and/or are social media focused	Social and entertainment media users or Extensive users	All respondents on Zoom
B	Younger people (<30 - preferably under 25) who may have some digital skills and/or are social media focused	Social and entertainment media users	All respondents on Zoom
C	Younger people (<30 - preferably under 25) who may have some digital skills and/or are social media focused	Social and entertainment media users	All respondents on Zoom
D	"Older adults (55+) who are offline or with limited digital skills"	General or Limited users	All respondents on Zoom
E	"Older adults (55+) who are offline or with limited digital skills"	General or Limited users	Respondents socially distanced in centre – researchers on Zoom via data projector Limited demographics provided
F	"Older adults (55+) who are offline or with limited digital skills"	General or Limited users	Respondents socially distanced in centre – researchers on Zoom via data projector Limited demographics provided
G	Older adults with digital skills	General users	All respondents on Zoom
H	Older adults with digital skills	General or Extensive users	All respondents on Zoom
I	Older adults (55+) who are offline or with limited digital skills	General or Limited users	All respondents on Zoom
J	Older adults (55+) who are offline or with limited digital skills	General or Limited users	All respondents on Zoom
K	Adults with limited digital skills	Limited users	All respondents on Zoom
L	Adults with limited digital skills	General or Limited users	All respondents on Zoom
M	Post-18 education students with higher digital skills	Extensive or Extensive political users	All respondents on Zoom
N	Post-18 education students with higher digital skills	Extensive or Extensive political users	All respondents on Zoom

4.3 Themes

We undertook a thematic analysis of the focus group data underpinned by our Data Citizenship framework. We have grouped the focus group findings under three themes:

1. Limited or little understanding of how digital systems and media work
2. Feeling disempowered in the face of platforms and broader digital society trends
3. Verifying information and seeking skills online – including a reliance on “networks of literacy”

4.4 Limited or little understanding of how digital systems and media work

When setting up the idea of the Data Citizenship framework, we argued that the ability to make digital rights claims is key for citizens of a digital society. This is clearly something that forms part of ‘Data Participation’ but is highly reliant on developed and robust ‘Data Doing’ and ‘Data Thinking’. In particular, an understanding of and ability to reflect on our digital society and the nature of data collection and use by platforms (see section 3). As our findings in section 3 indicate, though some of our user groups demonstrate relatively higher levels of data citizenships and literacy, these differences are relative. Very few of our respondents consistently provided responses that indicated a depth of awareness, understanding or best practice across all aspects. This lack of awareness and understanding of digital society technologies, limited depth of critical Data Thinking nor ability or willingness to assert digital rights claims is apparent throughout the focus group interviews. We have identified five main issues under this theme in the focus group discussions. Put starkly these are:

1. Limited or no understanding of how digital systems and media actually work
2. Limited or no understanding of basic definitions such as ‘data’
3. Limited or no understanding of the kinds of data platforms can and do track
4. Limited or no understanding of the potential consequences of data tracking and trading
5. Limited or no understanding of which organisational and technological entities are involved in the digital ecosystem

In many cases one or more of these issues were described together or in combination. Our focus groups began with a set of questions and a discussion to establish what we might collectively mean by ‘data’. It became very clear that the use of the term data in the academic literature – to refer to the data platforms collect about users – was not understood by many respondents. In particular those with lower levels of digital skills and literacies. Two particular definitions were offered:

Data as a resource – your data allowance on a phone as noted by I3:

“The first thing I think of data is it's flashing up on my phone telling me I'm running out”
(Participant I3: F; 60 years old; post-18 education).

Data as numerical information or evidence, as described by G3:

“Maybe data research, maybe research information something you're researching yourself so you're checking out data evidence” (Participant G3: F; data not provided).

Once we had established an understanding of the meaning of data in the context of the focus group, respondents were able to articulate the types of information and data that social media platforms or other

digital systems might collect. The only groups to immediately focus on the same definition as used in much of the academic and policy debate were our younger ‘Extensive’ users. For example, M4 stated:

“Same as the other guys to be honest, it'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” (Participant M4: M; 20 years old, undertaking post-18 education).

This obviously an issue of language use – terms like information or personal information were more easily discussed. We interpret some of this initial discussion and establishing of language as reflecting a lack of understanding of what is going on behind the scenes when respondents use digital systems and media. This lack of understanding, a lack of Data Thinking, has direct impact on behaviour. For example, N2 appears resigned to just saying ‘yes’ to tracking cookies:

“When you go on websites and it asks you if you're ok with Cookies or whatever, most of the time I'm like yes just to get it out of the way because it's annoying because I never like seeing the effects of it so to me I'm thinking like since I never see what actually happens with it myself, I'm just like yea, OK it just happens” (Participant N2: M; 22 years old; post-18 education).

This behaviour could be viewed as simply a ‘lazy’ response but the need to click on and accept cookies, partly driven by legislation such as GDPR, is a very poor process for gaining meaningful consent (Carmi, 2021). Nor does this method allow for any kind of nuanced digital rights claims – one cannot say “yes you can use my data for this in this context but not for anything else”. Overall, many respondents admitted to a lack of knowledge about how systems work or what platforms do with data they collect. This is clear in a comment by G4:

“I'm not sure, but I'm always a bit wary as to what they actually do, so 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” (Participant G4: F; 84 years old, post-18 education)

In our focus group discussions, we noticed that the development of greater critical awareness tended to be driven by specific issues or acute incidents. For example, N3 stated that they “didn't really think about my data and where it was going for a long time”. But once she became aware that her passwords had appeared in a data leak her awareness and data protection behaviours changed:

“I was told by Apple - because I have saved passwords on my phone - that my passwords had appeared in a data leak or something and I had to change all my passwords and that freaked me

out and I was like oh my god so it was like my passwords had appeared somewhere on the dark web or something, so you need to change them. So I did and now it's something I think about a lot more like I will never use the same password for something now and I don't know I'm just conscious of it more it has freaked me out a little bit" (Participant N3: F; 23 years old; post-18 education)

This indicates that once citizens understand the direct consequence of data harms on their everyday life experience, they change their data citizenship practices and increase their literacy levels.

We view these comments as further evidence, along with the survey results, that many citizens have a very limited understanding of how both the technologies themselves work nor do they have detailed understanding of the economics and technical ecosystems that underpin platforms.

Respondents could talk about key technological applications such as "encryption" or storage in the "cloud" whilst also admitting a lack of knowledge about what these things mean or even if or how they could check these things. This is illustrated in a comment from F1:

"I've often wondered where it goes but I know it's stored somewhere like Facebook Google they allegedly say it's anonymised 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" (Participant F1: F; 45+ years old; no post 18 education).

Similarly, there is awareness of tracking and third-party data sharing but a lack of deeper understanding of the processes by which data are collected, linked, used, and sold. As I3 notes:

"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 realise 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" (Participant I3; F; 60 years old; no post-18 education).

These issues are compounded for some citizens by the 'language' around digital systems and media. This is notably true for citizens coming new to platforms or often returning after a break away from digital technology. In these cases, new ways of talking and thinking about digital systems have to be learned, things that long term users may take as givens or obvious. As J1 points out:

“And the language that's so important, it hasn't been mentioned before but the first difficulty I had was learning the different language which words mean one thing on the internet but something quite different to me when I hear it so you've got to know the language and be able to see it when it is written” (Participant J1: 90+ years of age; post-18 education).

These comments reinforce the results from the survey (section 3) indicating that knowledge of the details of both overt and passive data collection and sharing may be quite limited for many users, this does not mean they are not aware it is happening. In all the focus groups respondents expressed an awareness that data around use of platforms is collected; but what, how, and why were often poorly understood. Similarly, nearly all respondents had not read nor understood what forms of data collection and processing they had agreed to through platform “terms and conditions”.

This lack of awareness and resignation to the complexity of the issue is illustrated by a comment from M3:

“I think I'm sharing the obvious stuff like what the lads have mentioned like what apps you're on, what you scroll past, how long you're on certain pages for, but I think there's a lot of stuff that you just don't can't even think about you just don't know about because you're not in these big social media companies or whatever that record the data. No-one reads through all the terms and conditions so they'll be reading something that you just there's obvious stuff like what we've said that you can think of but there'll be things you can't think of and again god knows what that is and how scary it is” (Participant M3; M; 20 years old; post-18 education)

Respondents noticeably occasionally confused data security features such as encryption with the idea that data were safe from inappropriate use. For example, that secure encrypted websites implied the data could not be hacked nor could it be used for data trading purposes. This confusion is a key one to explore in the development of data literacies. In particular the confusion that encrypted data or using secure sites (e.g. with the ‘padlock’ symbol) implied data was secure from unwanted use or user profiling by the platforms of companies themselves as opposed to when ‘hacked’. Many programmes designed to support citizens moving online rightly teach that users should only use secure platforms and systems. This is taught through a technical definition of security, that when being transmitted over the internet encrypted data will be hard to intercept or will not be useable if acquired. In this study we are concerned with citizens understanding of how their data may be used by platforms and systems however technically secured from potential hacking. This understandable confusion is clear in the comment from G5:

“To add to what I said before, I'm sure they sell your information so they can make money by targeting you with advertising mostly but I think you have to make sure you use secure sites with the padlock and so if you stray onto other ones I'm not sure what they do with your data but I imagine it's more secure and they have rules how long they can keep your data in the bona fide sites” (Participant G5: F; 71 years old; no post 18 education)

Respondents were also aware of how differences in experience and knowledge may lead to different attitudes. With more experienced users being more comfortable with aspects of data being collected and produced. This position reflects the findings in section 3, in particular Figure 19. Yet even the more experienced respondents felt somewhat resigned to the current ‘new normal’, as much to facilitate the pragmatics of getting on with digital systems and media use. This point is clearly made by E2:

“I think the longer you have it the more willing you are you're just constantly just putting information in without perhaps questioning why you're doing it. So, I've used internet and computers for years and years and years and I'm a lot more willing to put data in I think than somebody who is new to it.” (Participant E2: F; 45+ years old; no post-18 education).

Similarly, the lack of specific knowledge about what happens with data, how it is processed to shared leads to a complex position to be comfortable with this ‘new normal’. A6 describes this as being ‘cynical’ whilst ‘liking to think’ collected data are put to ‘good purpose’:

“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 years old; post-18 education)

These comments reinforce the survey findings (section 3) where experience and ‘user types’ clearly correspond to some variations in attitudes. What they bring to the fore are the ways in which this experience plays out for different groups. This includes: the ability to understand the nuances of risks inherent in sharing and “giving off” data when online; the ability to differentiate (or not) “secure data exchange over a network” from “data being secure from unwanted use”; and ways in which citizens manage their experience of and often resignation to the realities of the ‘new normal’ of data being collected and processed.

4.5 Feeling disempowered

Another consistent theme in the focus group data is a feeling of being disempowered. This disempowerment is mostly associated with four issues:

1. Feeling disempowered in relation to and by social media platforms
2. Background feelings of worry and anxiety over uses of their data by platforms and systems
3. Lack of confidence in their digital 'efficacy' in regard to managing and protecting personal data
4. Concerns over the potential to be manipulated

4.5.1 Feelings of disempowerment and worry

As noted above in section 4.4, all respondents, even those with greater confidence, noted some aspect of resignation and disempowerment in relation to managing, protecting and controlling their data. Very often constructed as a personal failing or lack of skills (see section 4.4). This feeling of disempowerment and of identifying this as a personal failing fit with the points made in section 2 where we noted how 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 internalisation of these prescriptions (Isin and Ruppert's "conventions", see section 2) that digital citizens are subjected to and follow. Part of the 'new normal' is an internalised assumption that it is citizens responsibility to manage settings, know how data may be used and protect their data. As such our respondents often talked about it being hard to "fight back" or "resist" the practices and processes used by platforms. As N4 notes:

"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 you're 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&C's 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" (Participant N4: M; 25 years old; post-18 education).

This resignation is not passive it has led to changes in behaviour even total avoidance as we will present later. Importantly, in all the focus group sessions respondents talked about feeling resigned but also 'uneasy' with data being tracked or shared in many ways. Very often this unease was expressed in quite forceful terms such as "creepy", "scary" and "horrid". M4 stated:

"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" (Participant M4; M; 20 years old; post-18 education)

Showing the complex trade-offs people often feel from their 'datafied' experiences, C3 acknowledges getting some useful things from these platforms, but also described platforms tracking of data as "creepy":

"It is useful, and it is creepy at the same time because it's kind of like they're spying on you" (Participant C3; M; 24 years old; no post 18 education)

Others described it as “scary”:

“I think it is scary about how much information the companies and governments have about me. It is scary” (B1; M; 26 years old; no post 18 education)

And M2 described this as “horrid”:

“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” (Participant M2: M; 20 years old; post-18 education).

Throughout the focus groups there was a general expression of anxiety and unease, what we might call “wrongness” about how platforms collect, use and trade data about citizens. Overall, we might describe the respondents’ position as one of general background and occasional overt “worry”, linked to a lack of confidence or knowledge, as described by K2:

“It worries me a lot because you always think who's got it and they get your details ... as you said you don't know who can see it so that is worrying.” (Participant K2: F; 52 years old; no post-18 education)

4.5.2 Lack of personal digital efficacy

This feeling of disempowerment is often reflected by respondents as a personal failing. For example, statements placing the fault or responsibility on themselves not on platforms or policy makers: “I don't know how to do X”, “It is my fault” or “I should do X but I don't”. As N1 notes:

“But it is horrible the fact that they're taking all your data, all your data's out there but I guess we're doing it but then that all comes down to the terms and conditions as well which I don't really read when it comes to websites so I guess that's my own fault” (Participant N1: F; 21 years old; post-18 education)

They go on to say:

“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 me knowledgeable or unknowledgeable like I don't really know other steps that I should be taking like other than the basically change your password” (Participant N1: F; 21 years old; post-18 education)

Personal responsibility was also tied up with “confidence” and the “type of user”. In many areas of practical activity to keep their data safe online respondents talked in terms of their confidence with platforms,

systems, or specific issues. For example, respondents often talked of “older users” in order to capture a set of behaviours and circumstances that included:

- Lack of knowledge
- Lack of confidence
- Fear of using digital systems and media

Similarly phrases such as “being a geek”, “being a teacher”, or “having used computers for a long time” were used to explain why a respondent took time to address privacy settings or deal with cookies. As H1 notes:

“I do think because I'm a geek I tend to do things that people online don't do tend to - wipe out history all the time, remove the cookies weekly, start again next week if they want.” (Participant H1: M; 64 years old; post-18 education)

4.5.3 The problem of practical action

The obvious question to raise is this: why does this worry and concern not translate into action? As the survey data show (section 3) few respondents undertook broad proactive steps to protect their data online. As noted above confidence and knowledge are also clearly barriers. This said we also noted multiple descriptions by respondents of effectively “giving up” on following through the steps needed. As we found in the survey few engage with the terms and conditions of use for platforms and systems. Implying that they had no knowledge as to what rights over their data they have handed over to these platforms and systems. This is clearly described by E1:

“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”

In our survey results (section 3) we found that respondents were relatively evenly split on the ease of changing settings, but few were confident with terms and conditions. More generally, respondents found that the tasks of checking terms and conditions, cookies, privacy settings, were often “contextually impractical”. As F2 notes, 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 of I will take care of myself and read the small print and all that but now it's like even 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 or want to watch that video. And sometimes I just get fed up and say OK I agree but we don't really know what we're agreeing with” (Participant F2: F; 45+ years old; no post-18 education).

Others were simply resigned to the fact that it was too much work to change settings and that there was in fact no real choice but to go along with the demands of the platform or system. D3 put this somewhat bluntly:

“If you don't agree to their T&C you can't use the service then you won't be able to do anything on line all the apps and programmes 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”
(Participant D3; M; 54 years old; no post 18 education).

In with these issues of disempowerment and worry respondents also raised concerns about the possibility of being manipulated. The touch point for such concerns was usually Cambridge Analytica:

“But I mean the Facebook some organisations 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” (Participant H3: M; 41 years old; post-18 education)

This concern was not just expressed in terms of politics but also purchasing and lifestyle. G7 talked about how this might be done “surreptitiously” without the knowledge or understanding of the users. This issue was also constructed as one of rights “you really want to know if somebody else is controlling your lifestyle or your life in a way and you don't really want to be pushed in directions that you don't want to move”.

Though the same respondent also thought such manipulation was:

“Quite possible because not everybody realises that they're likely to be pushed or are being pushed. Some people don't have a questioning attitude or cynical attitude but lots of people don't have that” (Participant G7; M; 78 years old; post-18 education)

Once again implying that resistance to such manipulation requires a specific attitude or approach from individuals.

These discussions go beyond the findings of the survey (section 3). Though the survey points to concerns, ambiguities, and differences in attitudes to the uses of data by platforms the survey results do not highlight the responses to these concerns. The focus groups make clear how citizens maintain both background and overt anxieties and concerns. Yet even the most engaged find acting on these concerns a challenge – often limited by the practical need to “get on with the job” or though lack of knowledge. Very often these limitations are presented as personal failings as much if not more so than failings of the platforms or regulators.

4.5.4 Hacking

Some, but certainly not all, of the worries raised by respondents concerned illegal access to data. Mostly referred to as ‘hacking’. It is interesting to note that ‘hacking’ is framed in three ways in the discussions. First, it is just something to be feared. A danger of being on online. Second, it is framed as one of the ways in which citizens data is at risk from being stolen, shared, or accessed without consent. Third, it is framed as something that citizens should guard against through proactive action to ensure that their data is secure. I2 argues it is:

“really down to you to be extra vigilant I think and I think the most frightening thing is the hacking side ... but I say it's down to the individual” (Participant I2: F; 74 years old; no post-18 education).

The fear of hacking and the need to protect against it appears most often in relation to loss of data that might lead to clear material harm such as banking details. As C1 notes:

“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” (Participant C1; F; 29 years old; no post-18 education).

We noted that none of the respondents linked the data collection and 3rd party data sharing by platforms with similar behaviour by hackers. Where they did bring these issues together it was often in relation to “trust” of digital platforms and systems, as described by N5:

“Even though companies lay out terms & conditions, and stuff I don't really trust anything like I don't trust anyone or anything online so it doesn't really matter to me even if Google tried to be all we're keeping you safe, because I just feel like it's all - I know I sound very pessimistic but I just think at any given moment even the most secure systems can be hacked” (Participant N5: F; 26 years old; post-18 education).

Respondents clearly maintain 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.

4.5.5 Rights

A small number of respondents framed the feeling of disempowerment in terms of rights. As we argued in section 2, a key part of Data Citizenship is the ability to formulate and make digital rights claims. These

arguments were presented as the breaking of ‘rights’ or the inability to assert rights particularly around privacy. I3 argued that the behaviour of platforms is inherently a breach of rights to privacy:

“It's an invasion of privacy at the end of the day because they know what you're doing, where you're doing it and when you're doing it” (Participant I3; F; 60 years old; no post-18 education).

E3 formulated this in terms of a lack of civic rights, arguing that both awareness of and some of the very principles of civil liberties, such as a right to privacy, are being eroded:

“My dad's generation - he's passed on sadly - but his generation were much more aware of what was a civil liberty, much more aware of it than people are now, so I'm very pessimistic the whole concept of privacy. I think is really under threat cos nobody's questioning [this] kind of community” (Participant E3: M; 45+ years old; no post-18 education).

4.6 Searching for information and help online

In the focus groups we discussed how respondents asked for support, verified information, or learned new skills. It became very clear that respondents all depended on what we have termed “networks of literacy”, which are “the ways people engage with others, where and with which media to gain the digital understanding, skills and competencies in ways that fit them” (Carmi et al., 2020). They are reaching out through both their physical and digital social networks of family, friends, acquaintances, work colleagues and other community members when needing support or advice. If these routes are not available, or timely enough, then they often select specific social media. This pattern of behaviour is clearest in three areas we explored:

1. Verifying news and information
2. Trust in content, media, platforms and others
3. Learning new skills

4.6.1 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 as E5 notes in relation to getting help:

“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” (Participant E5: F; 45+ years old, no post-18 education).

Social media themselves, of course, can support this behaviour as they are often one of the main means by which people rapidly connect across their social networks. L4 describes how they seek out knowledgeable friends and family in tier network:

“I do a bit of that on WhatsApp if I'm not sure about something I'll get in touch with someone who knows a bit more than I do so just to give me that bit of reassurance as to what's going on so we can discuss it... I've got a couple of friends and I speak to my son all the time on WhatsApp so just little things in general just to get some reassurances, so if you're not sure about something it's always best to ask or talk to someone about it isn't it?” (Participant L4: M; 57 years old; no post-18 education)

This quote also points out a key feature of these networks as routes to “reassurance” or “confirmation”. As we will discuss in a moment, this reliance both on physical social networks and digital social networks, may have very different results depending on levels of different literacies, such as media literacies or data literacies. Alternatively, others actively use 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 online centre] 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” (Participant F2: F; 45+ years old; no-post 18 education)

Others use digital media to build networks for overt community activism as F1 describes:

“When we were fighting to save the library, I went out with sheets for people to sign to save the library and got their name and email which we collated and put together to send into various places” (Participant F1: F; 45+ years old; no post-18 education).

Though they go on to comment on the trusting and uncritical manner in which people shared key personal information such as email addresses and post codes:

“I was, always incredible how willing people are to give you their name and email, oh yea that's fine, we'll do that and their postcode and they do it, they just sign it without even thinking ... I was just amazed that everybody just said oh that's fine, name email and postcode” (Participant F1: F; 45+ years old; no post 18 education).

Only one other respondent highlighted how these personal or community networks might pose data sharing risks to others:

“When I use YouTube I clear all the data all the time but my other family members are connected to the YouTube to my Gmail account so I get their data instead of my own data so I try and clear them out but not all the time, sometimes I do” (Participant D3; M; 54 years old; no post 18 education).

A number of respondents talked about having key “go to” individuals in their networks, or even being that person in other people’s networks, as I2 describes:

“I’m inundated by my neighbours here the ones who say you don’t need any lessons but they’re always knocking on my door whenever anything comes through” (Participant I2: F; 74 years old; no post-18 education).

We would argue that these networks, operating on different scales and with different levels of skill and knowledge among their members are in fact key to citizens data and digital literacy. They provide the basis for their navigation of digital content, their acquisition of skills and knowledge, their verification of information and underpin their community engagement. They therefore underpin and cut across all three of our data-citizenship dimensions. They support citizens in “doing” things with data, they support their “thinking” about and with data, and they underpin their Data Participation.

4.6.2 Verifying news and information

Respondents’ discussions of how they verify news and digital content provide multiple examples of these networks in action. N2 provides a clear example of how news and information flows through respondent’s social networks (both physical and digital). Such behaviour is not new, it reflects an understanding of how media content flow and influence citizens going back to Lazarsfeld’s (1968) “two-step slow model”. In this model social media content is often “filtered”, “mediated” or “framed” by key people in citizens social networks, this can also happen when people feel overwhelmed by the pace of information on these platforms:

“I talk to my friends normally whenever I see news because I actively try to disconnect myself from news because I always see my own health deteriorate if I try and keep up with the news regularly there’s so much happening every single day that I get overwhelmed by it so I just stay out of it, but then if something really does happen then either my friends tell me or my mum tells me or something like that” (Participant N2: M; 22 years old; post-18 education)

Alternatively, digital media themselves function as the filter and mediator as described by G3:

“I use Google search but I do have the BBC news app on both my phone and iPad and it's quite good because I get instant news flashes of anything that's considered to be big news so I use that quite a bit” (Participant G3; F; data not provided – older group 70+ years old)

We also found that some respondents would also go beyond their immediate social network of friends and family to their digital social networks. F1 stated:

“I Google it put in whatever the words are to see if it comes up and then I might go on WhatsApp to the group and ask them if they know of this and I do sometimes go on Facebook and put it up and ask if anybody has experienced this or done that what the results have been which I found has been pretty good way of getting a cross section of answers, I don't always trust the BBC or the news because it can all be manipulated to fit the facts as we all know so I tend not to believe everything that comes on the news or anything else, I'm very cynical about it I try and find out other facts if I can before I accept stuff” (Participant F1: F; 45+ years old; no post 18 education)

It is important to note that in this response F1 is highlighting a reliance on and trust in content from friends and generated on social media, over that from mainstream media (BBC). This reflects the findings in section 3.3.6 on trust in organisations and media – though here we were asking about trust in protection of data not trust in content. We will return to the issue of trust in content in section 4.6.3 below. It also shows one of the key problems that arise from a lack of deeper data and digital literacies. This is the circular and uncritical process of going back and forth between multiple digital sources such as Google and Facebook to check “facts”. With other formal and broadcast media sources being less trusted. Potentially reflecting once again a greater trust in material derived through social networks (physical and digital) than through other routes.

Many of these points made by respondents, especially around accessing and verification of information, therefore highlighted key media, digital and data literacies issues. Some respondents demonstrated clear data and information literacies being able to articulate different levels of reliability in data sources:

“I always look at an alternative source to check any facts ... there's a hierarchy isn't there in terms of evidence base to hearsay and it's about really on that hierarchy how well can you say well this is where it sits it sits as hearsay oh actually it's got proven source so like some government websites you know that for instance the ONS that data has been sifted and checked and verified and so on and so forth whereas somebody on Facebook they're just saying something for whatever so it's about understanding that I think” (Participant H2: M; 60 years old; post-18 education)

Others articulated this in terms of broader media literacy talking about how they cross checked information online with that found in broadcast media. This was clearly articulated by G7:

“Well the Times, Telegraph, maybe the Spectator, but I wouldn't give any of them 100% clearance as to the truth because they're all politically biased and you just have to look at the people who own them to realise that so I think when you get to our age you tend to use a lot of common sense and not believe everything you read or hear. On the BBC App you do get fact checks on certain things that have appeared in the press or in the media which at times is quite illuminating so I tend to take those with more belief than the general stories that come out” (Participant G7: M; 78 years old; post-18 education)

Other respondents spoke about cross checking with broadcast media, even if it did not fit their political position (either left or right wing). Some respondents did not check and often relied heavily on social media or web searches as exemplified in the following responses:

“I go to Google for everything, and the worst thing is I've not got people at home as well I could talk to so I go on internet, what's the weather going to be like tomorrow. Google's my friend”

“If I need to know about the news, I just use Google and YouTube. Youtube have a live translation and every video whichever one you need just for information and if you write about something in Google will show you videos. I use Google and YouTube” (Participant E4; F; 45+ years old; no post-18 education)

“I love sky news, all there is a lot of channels like if you watch Facebook like personal they send fake news or real news you don't know. But if you go to YouTube you know the real news I use it a lot but I use it to waste my time not like to use it if I want to see something I need to go to YouTube or sky news” (Participant C4; M; 18 years old; starting post 18 education)

Though they may view some digital sources as better or more reliable than others as described by C3:

“If I want the best information I Google it, Google doesn't lie it gives you all the information you want its better than Facebook because now Facebook owns Instagram, it owns WhatsApp it owns so many companies so all of them companies like this are still going to feed you with fake news good news so something that I won't take seriously” (Participant C3; M; 24 years old; no post 16 education)

We again see evidence of “networks of literacy” at work with family members or friends helping each other navigate these issues of data, digital and information literacy. An overt example is F3 who described talking to their children about these issues:

“I like I say to my boys if it's not true if you haven't got the facts and you haven't seen it for yourself, don't post it because it's only going to start an argument only going to escalate with other people but just make sure it's true factual and you know it's for real then you can't argue with

anybody. I can't post because I never know what's right" (Participant F3: F; 45+ years old; no post-18 education).

As noted in section 4.2 our Focus groups are not a representative sample of the UK population. They were selected to approximate groups with different Data Citizenship / data literacy scores and our user types. That said, we would argue that we can see a particular pattern in these responses that fits with survey results presented above:

1. The use of social media as a primary source of information and route to verification appears only in focus groups with mainly younger users with lower digital skills.
2. Overt distrust of mainstream broadcast media (as opposed to critical media literacy) again only appears in these groups.
3. The attempt to apply broadcast media literacy back onto social media content only appears in our focus groups with older users.
4. Our younger Extensive users show considerable information literacies and good awareness of digital media eco-systems and economies but do not articulate the same media literacies as the older users.

We therefore have at least three very distinct sets of overlapping perceptions and behaviours around accessing and assessing information (Data Doing), thinking through, and using this information (Data Thinking) and proactive support of others and community data and digital literacy (Data Participation). This leads us to question how we might socially and educationally address this more complex set of interactions around digital media.

4.6.3 Trust in content, media, platforms, and others

Trust in sources of information clearly underpins much of the above discussion. It underpins the evaluation of content as well as creating data or data traces as discussed in section 3.3.2. This issue of trust has both a technological and a broader social component. Technical trust mixes together a variety of issues already presented in section 3 and so far in section 4. These include:

- A lack of knowledge about how digital systems work and where data 'go'
- A lack of personal efficacy and confidence with digital systems
- A lack of trust in specific technologies or organisations – often these are synonymous (as in Facebook or Google)

The social component is much more about a lack of trust (or not) in organisations and groups and trust in (or not) personal social networks. M3 describes issues of technical trust:

"What is a bit concerning is its data about my health and myself and I don't know where that's going, I'm pretty sure I clicked all the buttons when I set it up so it doesn't share it with anyone, but

yea, I don't really... do I trust them? I don't know if I do" (Participant M3; M; 20 years old; post-18 education)

Social aspects of trust are again tied up with aspects of data, digital and information literacies. Here H4 describes not trusting social media content from a 'random person', and the need to assess this against broadcast media:

"If I got something through on Facebook, I'd check the news sites first rather than just believing some random person that's put a cure for Covid or something I wouldn't just take that at face value I'd look on the usual news websites, I know there's a bit of bias on somethings but they're generally accurate... The main ones like BBC ITV Sky that kind of thing I wouldn't just take someone on Facebook especially if I didn't know who it was. If someone posted something major had happened in the area on Facebook, I'd just go on Manchester Evening News and see if it said anything on there" (Participant H4: F; age not provided – retired; no-post 18 education)

These two aspects become entwined as users seek to verify content through further internet or social media searches whilst looking for technical 'markers' of reliability. N5 provides an example of this:

"If I'm uncertain, I suppose maybe I'd Google it and see if there are other articles that are saying the same thing but I would say that I've probably become quite used to making a decision about whether an article's legitimate or not based on how they present themselves so if I think it looks a bit click baity or a bit gimmicky I might not trust it or if in the URL it doesn't have one of those padlocks I might not trust it or I think if it's a bit sensationalist I might not trust it. So I think I've become quite attuned to knowing what looks legit and what doesn't but if I'm unsure then I'll Google and cross reference" (Participant N5: F; 26 years old; undertaking post-18 education.)

In this example N5 talks about a range of activities that help them assess content. These include:

- Further searches
- The 'look' of the content – not looking "click baity" or "gimmicky"
- Assessing the URL
- Checking for encryption (pad lock)

We would argue that the focus groups provide the same mixed bag of evidence around trust and also assessment of content as the survey. With many respondents:

- *Only doing 'some things' to check content*
- *Having very different levels and areas of trust in technologies, platforms and organisations*
- *Only a very few having deep and broad data and digital literacies*

- *As noted in section 4.5.3 all respondents face the ‘challenge of practical action’ in using their knowledge, understanding and skills*

4.6.4 Learning new skills

Having established that citizens are dependent on their underpinning networks of digital and data literacy, but that knowledge, understanding and skills vary greatly, we need to assess where citizens gain new skills and knowledge. Any successful awareness raising or education strategy, and any attempt to build critical democratic digital education interventions, will need to understand the routes citizens already take. When we explored these issues with respondents, they discussed two clear routes to solving problems, getting advice and in particular learning new skills. The first route, as already pointed out throughout this section, is citizens existing social networks (physical and digital). For many lower skilled users this is family and friends. Given our focus groups were recruited through digital support centres, formal and informal learning of new skills through such centres was also heavily mentioned. The other route was social media, predominantly YouTube. In fact, YouTube was the ‘go to’ digital location for nearly all new skills learning. As noted by A3:

“I think in terms of actually learning other skills, most people tend to use YouTube because everything you need is pretty much on YouTube these days you don’t have to read anything it’s literally a video and a person telling you how to get from A to B and that’s literally the best way to do it. And I think for myself if I need to learn anything it’s always on there and if it’s not then it’s just a blog on Google, so YouTube and Google are my main sources of information” (Participant A3; M; 24 years old; post-18 education).

This could be for almost any topic as described by J4:

“I find YouTube very useful for learning how to do things step by step for myself it can be either sewing or knitting you can use it for car ‘mechanic-ing’ whatever you want to do somebody has done it and filmed it and the great thing about YouTube you can freeze it yourself at a certain stage then play the next bit and go back over it again and again, you can watch somebody doing it saying this is how you do this. I find it wonderful for that, it’s an excellent teacher” (Participant J4: F; age not provided – older adults group; likely post-18 education).

As described by J4 many respondents talked about the visual and personal nature of the explanations on YouTube and the ability to “follow along” because you can pause and re-watch as many times as you need. This reliance on YouTube was highlighted and discussed in all the focus groups. One respondent talked clearly about the advantages of internet access for someone both recently moved the UK and as a new mother:

“I did that just when I arrived in the UK because in my country the internet is very poor until I came to the UK. At the beginning I look for somewhere for writing, doing my homework I search it I Google it, I use YouTube to have an idea how to write an article or letter or anything like that now when I have my baby anything that can maybe help my baby I read articles about how to look after him, his health, maybe it's like the internet was the first guide for me during my pregnancy and now when I look after my baby” (Participant C1; F; 29 years old; no post-18 education).

It was also clear that respondents made judgements about whether to seek information or support online via sources such as YouTube or via their social networks as described by E5:

“Where do I go to? I suppose to find out things I would go to YouTube I suppose if it's something very practical nothing to do with technology I'd go to YouTube, for fixing radiator or something like that. If it was something to do with actually fixing the software on a computer I would, I do know people personally who would be able to help me” (Participant E5: F; 45+ years old; no post-18 education)

4.7 Conclusion

We draw the following conclusions from our focus group work.

- The depth and breadth of knowledge about the collection, tracking and use of data by platforms is best described as “patchy”. It is especially limited for those people who have low digital skills and are limited or narrow users of digital systems.
- All respondents feel unease and differing levels of disempowerment around how platforms collect data, the uses to which it is put and the potential of third-party trading. This unease is 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”.
- The processes to verify information, the learning of new skills or the development of better awareness are tied to users “networks of literacy”.

4.7.1 Conclusion: limited or little understanding of how digital systems and media work

Both our survey and our focus group analyses show that respondents have some, but *not extensive* understanding of data that they overtly share or passively “give off” when using digital systems and media. This ‘Data Thinking’ is clearly patchy with different groups and individuals showing different patterns of awareness. A key issue is language. As noted in section 4.4 coming to an agreed understanding about the nature of shared/“given off” data or information was the first challenge for our focus group discussions.

Limited or narrow users were the least able to articulate clearly the nature and types of data shared/“given off” and clearly did not actively think about these issues – unless a specific fear (surveillance) or a specific incident (scam/hack) had raised their awareness. Though much academic and policy work talks about ‘data’, ‘sharing of data’ or ‘data protection’ only more experienced users talked in this way and, importantly, could articulate clearly what they meant. Others were clearly more comfortable talking about ‘information’ or specific items of data/information.

These findings make clear to us that despite a lot of ‘surface level’ use of terminology one of the challenges preventing citizens gaining a deeper understanding of our ‘datafied’ society is a lack of shared language tied to shared understandings that can allow for a more robust discussion. This lack also makes it much harder for citizens to express and put forward digital rights claims. For example, when it comes to various citizens rights in the GDPR, how can people demand their data to be erased or changed when they do not know what data are and who is involved in the data ecosystem?

It is also clear that respondents had an awareness of but lacked detailed understand of how their data were being extracted, tracked, how the underlying technologies work nor how the economics of digital platforms and systems.

They also did not understand what the consequences of those practices are. Another key confusion being a mixing of data security in the form of “encryption” (technically security) and being secure from harm that might be caused by use of data by platforms. None of our respondents had read or understood the terms and conditions of the platforms they used. These points again outline a very shaky basis upon which citizens could coherently and clearly put forward digital rights claims about the collection or use of their data.

4.7.2 Conclusion: feeling disempowered

Our focus group results with UK adults are strongly in line with similar prior studies with US young adults attitudes to the “privacy paradox” (Hargittai & Marwick, 2016). The argument that citizens care about privacy while both actively sharing and passively “giving off” data and personal information through their use of digital systems and media. Our focus group results indicate that respondents were aware of and cared about the potential risks associated with disclosing information online. Our survey findings show that they engage in at least some privacy-protective behaviours. Like the US young adults our respondents also feel a loss of control once data is released such that it is in the end many feel it is fully out their control.

What we found notable about our results was the language used to describe this. Terms such as “creepy”, “horrible”, and “scary”. All of these are words with connotations of fear and particularly of unwanted surveillance and overall, a pervasive feeling of ‘wrongness’. These are very emotive responses and make clear that the “privacy paradox” is far from just a cognitive or behavioural contradiction. It leaves citizens feeling unsafe and at risk.

We find that respondents did not translate this unease 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 interpret the unease expressed by our respondents as a response to what Hargittai and Marwick (2016, p.3737) have described as “networked privacy”, a situation where citizens exist in digital and social contexts where others – be they individuals, organisations, corporations, or the state – can and do violate their privacy. A situation where civic rights to privacy have been technologically eroded with limited civic resistance – as articulated by our respondent E3 (see section 4.5.5).

4.7.3 Conclusion: “networks of literacy” and verifying information

A core finding from the focus groups is the dependence of respondents on their “networks of literacy” for support with digital and data issues and also as routes to verifying information. Very often these consist mainly of close family, friends, or acquaintances. These may be in part managed or mediated by digital systems and media but are mainly built on every day in person interaction.

As we noted in prior work on mobile phone use (Yates and Lockley, 2008) most mediated relationships are not in fact “remote” but digital interactions form part of normal ‘local’ interaction. Though we also find evidence of respondents going out to wider networks via social media if local connections cannot provide the support they need. These locally situated “networks of literacy” can be beneficial providing access to advice, ongoing support, or skills. They can also be limiting if the needed or best support is not available in the network. They also run the risk of becoming self-reinforcing, preventing members from being able to critically assess content and information. For some groups, especially our ‘Social and entertainment media’ users and more ‘Limited’ users, verification and assessment of data and information can become highly

reliant on narrow, potentially self-reinforcing interactions with a narrow network and narrow social media sources.

We also found evidence of older users applying broader critical media literacy ideas onto social media and digital content – applying the idea that sources are biased and need to be verified against other traditional (broadcast) sources or even books and libraries! Though this was done without the deeper understanding of how digital platforms serve up material algorithmically that we found expressed by some of our younger ‘Extensive’ users. Overall, we would argue that these networks are a key to engaging users and raising awareness, especially for those who are outside formal contexts, such as education.

4.7.4 Conclusion: seeking skills

Our focus group results also make clear that respondents are dependent on these networks when seeking new skills. A clear alternative both in general and where these networks cannot provide support, is YouTube. It is clearly a major source of support for everything from new digital skills to activities in other areas of respondents’ lives. Though it has the risk of taking citizens on to further algorithmically suggested content, and hence problematic content such as conspiracy theories and disinformation. Without deeper critical data and digital literacies citizens may not be able to evaluate the quality and veracity of the information or content and the processes that have delivered it to them.

5 The challenge of critical educational intervention

Data has become the core mediating element of our cultural, social, political, and economic worlds, our Data Citizenship model aims to create a framework that explores links between “data, power, and positionality” (Philip et al., 2016, p.365). It outlines the importance of a critical and proactive individual as well as a collective stand, at a time when society’s datafication and algorithmically driven decision making has become normalised. One aspect of successful Data Citizenship is to foster greater agency – which we understand following Couldry et al (2014a; 2014b) as ‘the longer processes of action based on reflection, making sense of the world so as to act within it’ (p.891) in increasingly data-driven environments. To do so, it is crucial to cultivate Data Thinking as part of both formal and informal education and civic culture. Thus, our Data Citizenship framework aimed to extend the concept of democratic education and position it in the context of the challenges of the ‘datafied’ society.

Furthermore, our conceptualisation of Data Citizenship departs from narrow understandings of citizens as individuals with responsibilities and rights, to a broader conceptualisation of citizenship as enacted individually and collectively. Indeed, through Data Citizenship, citizens are encouraged and supported to carry out an individual *and* collective critical inquiry in order to fully participate in their communities in ways that are proactive and meaningful.

While we acknowledge the complexities of data literacy education (Deahl, 2014) we believe that the current power imbalance between data-processors and data-subjects should be at the centre of any such educational initiatives and be framed more broadly as Data Citizenship education. Data Citizenship education should be positioned within a wider discourse on the value of citizens democratic participation, which aims to liberate citizens “from powerless positions and places them in a position to construct their own future” (White 2003, p.37). In line with the work of democratic educators, Data Citizenship emphasises the importance of individual and collective proactive knowledge creation, collaboration, and action to resist tech-domination, oppression, and surveillance in data-driven societies. The question to be answered is therefore: how we match these goals with the evidence from the survey and focus groups?

The results imply that considerable additional work is needed to provide citizens with the Data Doing, thinking and participating to proactively act as data and digital citizens as envisioned by the authors discussed in section 2. When the project was conceived it had been our plan to develop educational materials to support greater data literacy – somewhat naively believing that there would be straightforward set of issues to address. This is clearly not the case. Also, as noted in section 1 much has happened in the intervening three years since the proposal for the project was submitted. The Covid-19 pandemic made our reliance on ‘datafied’ technologies even deeper, forcing many to use more digital

services for their everyday life activities. Data and digital literacy have become key topics as scholars, policy makers, practitioners and communities have sought to address issues of mis-/dis-/mal-information. Further data breaches and public debate following Cambridge Analytica, issues of bias in AI and machine learning have all brought issues of data surveillance, control, and manipulation, data rights to the fore for many communities and groups. These debates were touch points in some of our focus group discussions, even if the depth and breadth of a detailed understanding was not completely evident.

A result of this, there has been a growing academic industry of education and training tools (mostly online) that seek to provide the critical data and digital literacies citizens need to address these issues. It is also the case that some companies and organisations in the digital economy are changing their behaviours. The move by Apple in 2021 to allow users to prevent 3rd-party data exchange and cross-application tracking being one and the pushing out of ‘cookies’ in newer more privacy friendly browsers (e.g. Brave) being another. Similarly, the provision of greater visibility around the data being collected and tracked by apps, the data we “give off” when living our daily digital lives, could potentially help citizens potentially make more informed choices (see Figure 75). We have therefore taken a different route with the final part of our project. We have separately developed guidance for educators, practitioners, and policy makers on the issues we believe need to be addressed in designing or selecting educational resources and the issues around their deployment.

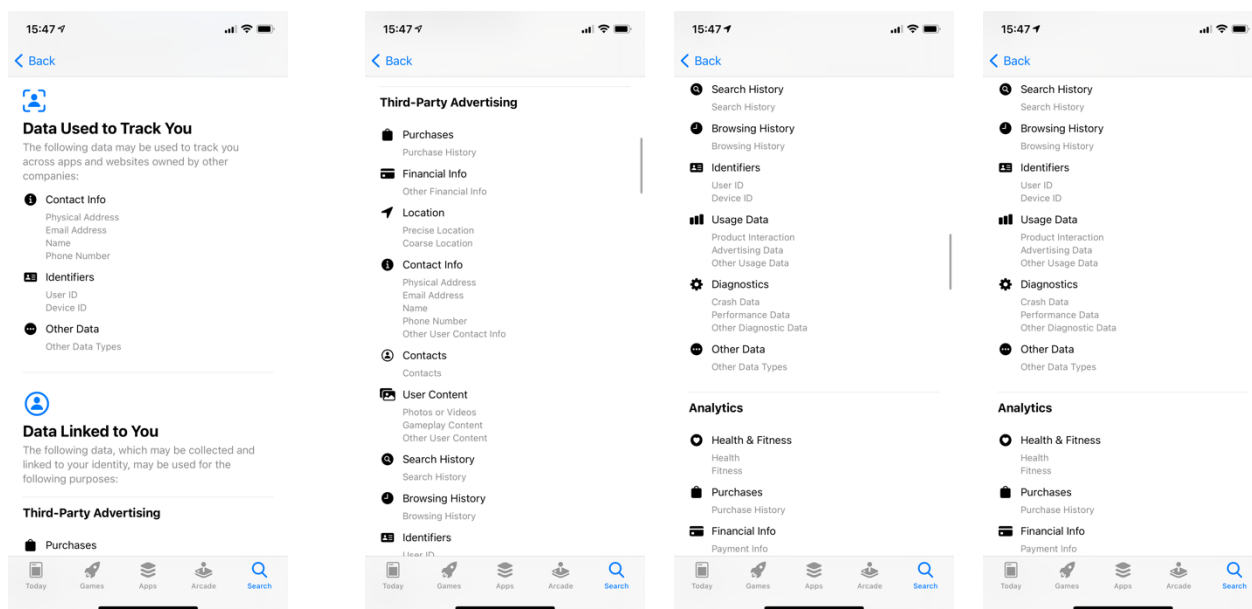


Figure 50: Example of information on Facebook app data tracking as presented on iPhone

This guidance is provided in an accompanying report. Here we layout the seven issues that we feel flow from our findings and that underpin our guidance.

1. Fear and worry

2. The challenge of practical action
3. Responsibilities
4. Audiences
5. Routes to engagement
6. Skills vs understanding
7. Critical education

5.1 Fear and worry

As section 4.5.1 makes clear citizens are uncomfortable with “networked privacy” – the knowledge that in many respects online privacy is only partial. As our survey results show, citizens are not happy with being forced to accept this in order to use digital systems and media – getting a free service is not a bargain they are happy or comfortable making. This discomfort in part derives from their feeling of disempowerment and lack of options. There is a delicate line to walk in supporting citizens to develop greater data literacy without increasing this concern and feeling of disempowerment. It is well documented that fears over such things as cybercrime, data loss or tracking, and exposure to unwanted content are barriers to moving online for non-users or to greater use by limited users. Effectively excluding them from the many benefits of digital systems and media use. As we noted in our discussion of the focus group findings these feelings of unease and concern do not always translate into practical action.

- *Any educational or awareness raising interventions must ensure citizens feel more empowered and have practical and alternative routes to enact that empowerment.*

A way possible way is to **create data stories** that provide explanatory narratives for citizens. As our focus groups show, only when things happened to respondents in their everyday lives did they changed their behaviour. **Data stories need to be contextual and meaningful to citizens’ lives have a potential to increase their data literacies.** This requirement leads to two clear issues around practical action and responsibilities for address citizen concerns.

5.2 The challenge of practical action

It is clear from our focus groups, and implied in the data form our survey, that citizens do not follow through on the practical actions needed to address their concerns nor to make their data safe. More work is needed on this but recent research on the situated nature of technology use and our reading of the focus group findings all appear to point in a similar direction. Recent research by Allman and Blank (2021) has

highlighted the fact that many users, especially those with lower digital literacies, tend to follow quite prescribed paths through digital systems. They “follow the script they know”. This finding is very similar to ethnomethodological work on the interaction with technologies (see Suchman’s (1987) foundational work on this topic).

Here again many people have very situational and local understanding of how to use technologies to achieve specific goals but lack the breadth and depth of Data Thinking and Data Doing needed to tackle new circumstances. We can see this situation described in our focus group work where either limited knowledge or the need to practically achieve tasks quickly means that people agree to terms and conditions, accept cookies, or share data or information they may be concerned about. This is rather than take the time to select cookies, change privacy settings, review or control data being shared or take time understand the consequences of using specific systems.

There is an argument that such constraints are partly intentional, what have been called “dark patterns” (Carmi, 2021), designed to make the actions most useful to the platforms or companies ‘simplest/default’ and to make asserting different digital rights difficult or impossible if the service is to be used (Forbrukerrådet, 2018). This behaviour of ‘taking the easiest route’ or accepting through not really consenting was common across all our user types.

- *Any educational or awareness raising interventions need to 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.*

5.3 Responsibilities

When designing the project, building on experience in supporting digital inclusion interventions, the project lead (Yates) formulated the projects educational goals around the training or development of citizens. This places a very strong emphasis on individual, potentially community action. As both the project research and the external environment have developed it has become very clear that there are multiple stakeholders who need to be engaged, change practices or policy to address the issues that have come to the fore.

5.3.1 Citizens

Citizens retain a key role in this debate – but it is more than just having the data and digital literacies to navigate our ‘datafied’ society. There remains an onus on them to be active and engaged citizens and therefore they need to develop their Data Thinking and be supported to undertake Data Participation. This

is no different a position than that put forward in much media literacies education work. But as our research shows having a depth of data literacy, being a knowing and active Data Citizen, is only part of the equation. The design of systems, the legal frameworks in place and regulation also need to play a part. As we will discuss below, active Data Citizenship must include an element of making digital rights claims of industry and governments/regulators. Though this does not absolve industry or regulators from their role in addressing citizen concerns.

- *Any educational or awareness raising interventions need to make clear to citizens their responsibilities in regard to their own data but also the role of industry and regulators.*

5.3.2 Industry

We do not have space to explore in depth the full role of industry in these issues. Recent non-academic publications have sought to raise awareness of the underlying technologies and algorithms (O’Neil, 2016), their differential impacts on social groups such as poorer citizens (Eubanks, 2018) and their broader social consequences (Arthur, 2021). The much-publicised conflict between Apple and other companies such as Google and Facebook over Apple’s restrictions on Apple devices for cross-app and 3rd party sharing is one example of industry change. Reports³ indicate that over 90% of Apple users have opted out of apps such as Facebook continuing with cross-app data sharing. Though of course Apple are still collecting user data for their own purposes and within App data collection by Facebook and Google remains. Google’s removal of cookies from further Chrome browsers and replacement with the Federated Learning of Cohorts technology appears to be halfway house between greater privacy and keeping open the targeted advertising industry. Apple’s restriction on cookies sharing data is a further example. These actions are presented as supporting “privacy” but in many respects they only represent a reduction in the extent of citizens “networked privacy” – the extent to which and the range of organisations that can access private information.

This said, such actions leave two issues for citizens to navigate. First, making visible the data being traded and limiting how it may be used help citizens understand the data environment they are in, but they do not necessarily vastly increase the range of actions they can take. Second, companies such as Apple provide

³ <https://arstechnica.com/gadgets/2021/05/96-of-us-users-opt-out-of-app-tracking-in-ios-14-5-analytics-find/>

high-cost devices and services. There is a potential scenario that those who can afford such devices and services will gain ever greater control and privacy. Whereas those citizens using lower cost or free devices and services will be most exposed the risks associated with data sharing. The ‘new normal’ may end up being different depending on your economic capital. This argument that privacy is directly linked to systems of power has been articulated in more detail by Véliz (2020).

- *Any educational interventions or awareness raising must make clear to citizens the role of industry and the practical options provided by but also limitations of platforms.*

5.3.3 Governments, policy makers and regulators

There is a clear role for government and regulators to intervene in two areas:

1. *Support for and provision of data, digital and media literacy education and training*
2. *Democratic regulation and effective enforcement of digital industries*

These are obviously major interventions. The second being beyond the scope of this report to discuss in depth. In terms of the provision of digital literacy support. Our findings show that in many cases respondents blamed themselves for ‘failing’ their approach to dealing with and protecting their data.

We would argue that this approach of focusing on individual responsibility within an often exploitative and harmful data ecosystem is highly problematic. This is especially the true for those citizens that are outside of formal education or similar contexts where education and training is readily available. Citizens can hardly be blamed for not having skills or knowledge where there are few places to where this can be effectively gained.

We would argue that responsibility of allocating appropriate resources to increase data literacy lies first with government. In particular to support those outside of formal education more key accessible spaces, such as libraries, are needed where professionals can assist and support citizens to develop greater data literacy.

In regard to regulation the UK government is currently working on an Online Harms Bill. While the draft Bill proposes approaches of ‘care by design’, it is clear that the big technology companies unlikely to themselves develop programmes the depth of critical education of citizens we are proposing here. Nor are they likely to design platforms that put people’s rights at the centre when their business model relies

exclusively on the extraction and monetisation of individuals’ data. Hence, governments need to reconsider the tendency to look for answers from the companies that created the problems. As OfCom will receive more powers as part of the Online Safety Bill, it is crucial to make sure that citizens rights are maintained and importantly that they are aware of them and know how to claim them.

Furthermore, legislators need to make sure that the mechanisms that are offered to people to make rights claims are easy to use and straightforward. Organisations such as the Information Commissioners Office must do more to engage with citizens from different backgrounds and make the process of making rights claims easy and fast.

As part of this the UK government has released while this report was being prepared a new “Online media literacy strategy”⁴. The Five principles outlined in the strategy directly mirror the points made in this report and presented in our accompanying policy guidance report. We would though agree with some of the first assessments of this strategy (for example see Sonia Livingstone⁵). In particular, we agree that the principles are sound, but the focus is very much on citizens as consumers – about improving their digital experience as consumers. There is much less in the strategy about Data Thinking that seeks to provide citizens with the tools to challenge both government and industry so as to make substantive and actionable digital rights claims. We explore this strategy further in our accompanying policy report.

- *Any educational interventions or awareness raising must 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.*

5.4 Audiences

The findings for this research make clear that there is no ‘one size fits all’ approach to developing citizens data literacies. Importantly, we would argue that data and digital literacy levels are in fact likely lower than

⁴ <https://www.gov.uk/government/publications/online-media-literacy-strategy>

⁵ <https://blogs.lse.ac.uk/medialse/2021/07/22/a-missed-opportunity-dcms-announces-its-new-national-media-strategy/>

many official statistics may indicate. We see in our survey data that respondents often have some awareness of issues, but hardly complete awareness of the data ecosystem and importantly, what are the consequences of that on their life. We also note that such awareness does not necessarily translate into action. In our focus groups we can see that awareness of issues, does not imply depth of understanding of the technical, economic or organisational processes underpinning the data economy. You may be aware that your data are tracked, but the additional crucial understanding of by whom, how, and in what ways and this data is used, may not be known. Awareness is not the same as knowing the detail.

Our Data Citizenship personas are a starting point for understanding audiences for educational or awareness raising interventions. They imply that for all but our two types of 'Extensive' users the foundation points for developing Data Citizenship are in fact quite limited. As we will discuss in a moment, many of our survey and focus group respondents are operating at the lower levels of current digital skills models. As such these citizens will need support on a much longer journey than say our extensive users. The challenge remains of doing this in a way that engages citizens, but which does not add to their worries and anxieties over the collection and use of data. Importantly, developing Data Citizenship needs to be part of citizens personal and civic circumstances. It is more than just a set of practical skills.

- *Any educational interventions or awareness raising must 'meet citizens where they are' in terms of their digital and social experience and context.*

5.5 Routes to engagement

A key finding from our survey is the role of education, especially post-18 education, in developing greater data literacies. This obviously provides a direct route for educational and awareness raising interventions. What is not clear from our research is whether it is simply greater exposure to technologies and digital systems through higher education, aspects of curriculum content, critical academic practice, or a combination of all these things that underpins higher data literacies in those with a post-18 education. This question needs further research. Especially if the key elements can be translated into other contexts such as pre-18 education or informal learning contexts.

But our research clearly shows that for those who do not experience a post-18 education do not seem to develop very strong data literacies. They are also clearly constrained in their Data Citizenship and more broadly as citizens in a digital and 'datafied' society. It is not clear what the best 'informal' routes might be to support these citizens in developing stronger data literacies. Citizens own "networks of literacy" maybe a route through which such support can be provided. This again requires thinking about the social and

community contexts of citizens, their intersectionality (race, gender, socio-economic status, age, mental and physical abilities, etc.) and the need to ensure interventions appropriately ‘meet citizens where they are’ socially as well as in terms of digital awareness.

- *Any educational interventions or awareness raising must address the challenge that those adults most in need of support are very likely outside formal educational settings.*

5.6 Skills vs understanding

There are well established digital skills frameworks that have worked well to underpin digital inclusion interventions and practice. For example, the Basic Digital Skills Framework (Doteveryone, 2015) covers five skills areas only two of which go beyond our definition of ‘Data Doing’. This focus on more practical skills and not on the broader issue of Data Citizenship can be found in many similar examples of digital skills frameworks. The focus on skills tends to place less, if any, emphasis on a broader understanding and critical evaluation of the social and economic context of digital systems and media. It is also clear that much of our ‘Data Thinking’ and ‘Data Participation’ dimensions fall into the higher-level digital skills categories. Our survey data and our focus groups discussions indicate that these are the skills that many respondents lack or struggle with.

These results are not particular to the UK. Van Deursen and van Dijk (2008) surveyed and tested a sample of the Dutch population on a digital skills measure. They found that Dutch population had on average a fairly high level of operational and formal Internet skills. Yet they found that levels of information skills and especially strategic Internet skills attained are much lower. Information Internet skills are only completed in 62% of cases and strategic Internet skills are accomplished by only 25% of cases. These strategic internet skills map closest onto our ‘Data Thinking’ and ‘Data Participation’ categories.

As with our results educational level attained was the most important correlating factor with skills performance. Skills are therefore key, and as noted above in meeting citizens ‘where they are’ there may be core skills training needed to support their Data Citizenship journey.

We therefore argue that skills are only part of the story, and Data Citizenship requires more than just skills, it requires the development of greater critical awareness and a more holistic understanding of the data ecosystem.

As noted in section 2 discussion of the participatory and engaged aspects of data literacy and citizenship is missing from many studies and policy documents – which focus on practical skills (Data Doing).

- *Any educational interventions or awareness raising must support skills development, but must be more than skills encompassing key elements of Data Participation.*

5.7 Critical education

As we theoretically argued in section 2 data literacies need to be more than “basic digital skills”. It needs to combine these with broader critical thinking and understanding of the digital eco-systems in which citizens operate. We therefore argue that the overall approach to data literacies and Data Citizenship needs to be grounded in a broader model of “democratic education” (Dewey, 1930; Freire, 1970/1996). This needs to equip citizens with skills and knowledge to enable them to effectively exercise citizenship in a ‘datafied’ society. The objective has to be more than just skills to allow them to be effective consumers (or safe online while consuming) but it has to empower citizens to critically examine their positions within the social, economic and power structures of a ‘datafied’ society. The goal being to achieve a more conscious, informed, and active way of digital living. A key element of this has to be equipping citizens with the ability to identify and critique the new “conventions” of our ‘datafied’ society that platforms, technologies or organisations subject them to and expect them to follow. They need to be able to critique and feel empowered to challenge the ‘new normal’ or our “networked privacy”.

- *Any educational interventions or awareness raising must 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.*

6 Conclusions

The Me and My Big Data project had three main goals:

1. 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.
2. To understand the intersectional basis of variations and inequalities in data literacy across a range of demographics factors.
3. To develop policy recommendations for stakeholders on developing citizens' data literacy.

Section 2 of this report present the development of our Data Citizenship model. This model argues for three key dimensions of active data literacies among citizens:

1. **Data Doing** – the accessing, assessing, use and protection of data by citizens.
2. **Data Thinking** – problem solving with data as well as critical understanding the social, economic, cultural and technical aspects of our 'datafied' society.
3. **Data Participation** – both using data to achieve personal and civic goals and also working with and helping others to achieve the same.

We have used this model to assess citizens data literacies through both a national survey and citizen focus groups. Along with developing a measure of Data Citizenship our survey and focus group work also addressed the following questions:

- How do citizens respond to the power imbalance between these platforms, organisations and themselves?
 - How do citizens feel in relation to this power imbalance?
 - To what extent do citizens trust digital media and systems – especially major platforms?
- To what extent do citizens trust the organisations that aggregate and use their data?
- To what extent have the conventions around data extraction and use by data-processors become 'the new normal'?
 - To what extent are citizens comfortable with this 'new normal'?
 - What actions do citizens take in response to these conventions – to protect their data or assess content that is exchanged?

6.1 Overall data literacies

Our evidence paints a picture of low data literacy levels for the majority of the UK population. Specific analyses demonstrate that knowledge of the details of data extraction and data use by digital systems and media is low for the majority of UK citizens. This does not mean citizens are unaware of the broad issues or

are disengaged from them, rather it reflects the extent to which gaining a deeper and more detailed understanding is difficult. This said there is clear variations in levels of data literacy across different social groups. Our overall model shows three key factors that impact levels of data literacy:

- Age – older users tend to have lower data literacies levels, though some younger citizens with lower education attainment also score lower on our data literacies measure.
- Socio-economic status – citizens from lower socio-economic status groups (NRS grades C2, D and E) tend to have lower data literacies levels.
- Education – citizens with lower educational attainment, especially the lack of a post-18 education, have lower data literacies.

These factors intersect and we have developed five Data Citizenship personas drawing on prior work exploring inequalities in digital access and skills to reflect these interactions. These personas are based on differential levels and types of digital systems and media use. These personas are described in detail in section 3.4. Our 'Extensive' and 'Extensive political' users show the highest levels of data literacies and show greater knowledge of both data extraction and data use by digital systems and media. They are the groups most likely to undertake Data Participation activities. Our 'General users' show lower levels of data literacies and show less understanding of data extraction and data use by digital systems and media. Our 'Social and entertainment media users' are younger adults, mainly without a post-18 education and from lower socio-economic households. This group has lower data literacies levels than our extensive users, but also show a much higher trust in content and material derived from their social networks (their friends or via their social media). Our 'Limited' users only marginally engage with digital systems and media. They tend to be older adults, with lower educational attainment and to be from lower socio-economic grade households. They show the lowest levels of data literacies and are the least likely to trust online content.

Therefore, we can see notable difference and inequalities in data literacies and the ability to be active and engaged Data Citizens across our personas.

It is not surprising that these inequalities exist and match onto our personas. We have demonstrated in prior research (Yates and Lockley, 2018) that these groupings, their use of digital systems and media and their levels of knowledge are underpinned by notable differences in economic, social and cultural capital. In particular, we would note the higher cultural capital (education) of our Extensive users.

This pattern of greater awareness and a greater likelihood of relevant action by our two types of ‘Extensive’ users and lower levels for our “Social and entertainments media’ and ‘Limited’ users is found across nearly all the survey data.

There are some commonalities across the groups as well. All groups are much happier with data being collected to provide benefit to themselves, but not benefit to the platforms and companies. Here our ‘Extensive’ users show more comfort with data collection if used to deliver services than other user types. None of our groups have a majority who are comfortable with organisations undertaking third party exchange or sale of data. Importantly none of our groups are happy for their data to be used just to get a free service. Rather, they feel they have “no choice” but to allow platforms to operate “networked privacy”. This is a situation where citizens digital and social contexts are such that others – be they individuals, organisations, or corporations – can and do effectively violate their privacy on an ongoing basis through the extraction and use of data. One of our focus group respondents (Participant E3: M; 45+ years old, no post-18 education) described the development of this situation as one where civic right to privacy have been technologically eroded with limited civic resistance.

We would argue that these survey results point to two issues:

- 1. Users are not comfortable with the manner in which their data is used to drive the economics and practices of big-tech platforms. That said they do want some of the services, especially the personalised services, that these platforms provide.*
- 2. We can interpret this as either a failure to understand the technical and economic processes that underpin the current design of algorithmic delivery of content by platforms or a lack of an ability to make digital rights claims of platforms about what is done with their data. As we found in our focus group data it appears to be a mix of both issues.*

We also found that ability to undertake and also user confidence in, practical action to protect data or manage such things as privacy settings were also generally low across all user groups. This varied by type of activity – different types of Data Doing. Some simpler actions such as ‘strong passwords’ were broadly undertaken by all. Much lower confidence was found for more complex activities such as preventing data being tracked across apps, protecting devices on public wifi, or key activities such as anti-virus protection. Within this variation by activity, we once again found that our ‘Social and entertainment media users’ and ‘Limited’ user groups are likely to score lower than the other user types. Though ‘Extensive’ users are more active and use a greater variety of systems they are also more likely to engage in some practices to protect

their data and privacy. Other groups may be less active or use a narrower set of systems but are more likely to be at risk of online harms.

We would argue that these findings point to an unequal distribution of risks across our user types.

There is one area where the ‘Social and entertainment media users’ stand out. This in their trust of content and data provided by friends in social media. They are also the least likely to have a varied engagement with information sources on social media. They along with our ‘Extensive’ users are more likely to say that their sources of information depend on their social media networks. We view these results as pointing to quite significant issues of data literacy – in particular Data Thinking – where different groups may use similar platforms, or are reliant on them, yet have very different perceptions of their veracity and reliability. In terms of Data Participation this was again split in the same pattern across our user types but with greater statistical difference between the groups. Only our ‘Extensive’ user group had undertaken any of the digital rights activities we surveyed – such as reporting problematic content or verified information for offline action. They were also the group most likely to have collected data from digital sources for their own or community use and the group most likely to process or manipulate this data in some way (analysed, edited or shared). We would argue that our survey results therefore show a clear differentiation in Data Participation across our user groups. This differentiation by user types and the three core demographics noted above was also clear in our focus group discussions.

6.2 Citizens response to the power imbalance between platforms and themselves

Our results make clear that citizens feel very disempowered in regard to big-tech platforms extraction and use of their data. This came through very clearly in our focus groups. Citizens described the unease this made them feel in quite strong terms such as “creepy”, “horrid” and “scary”. We were struck by the extent to which all respondents – no matter their level of digital skills or demographics – express this concern. This feeling of disempowerment came from a range of sources, but we would argue the lack of transparency in how digital systems and media work and the lack of answers to questions about by whom, where, how and why their data are extracted and processed is key. It is therefore no wonder that citizens have the lowest levels of trust in social media companies and platforms to protect their data.

Our survey and focus group evidence clearly point to the data extractive behaviours of digital systems and media have become the ‘new normal’. The ‘conventions’ imposed on users but the power imbalance the experience with data processors are very difficult to escape. Our results clearly indicate that citizens are not at all happy nor comfortable with the implied ‘deal’ that the cost of a ‘free service’ is the offering up of

their often quite personal data. But their feelings of discomfort go along with their feelings of disempowerment.

It is also clear that citizens therefore suffer from the limits to 'practical action' as they encounter data and privacy issues online. This points to a larger problem of the way many digital systems and media intentionally design 'dark patterns' (Carmi, 2021) to be able to extract more data from people. Even where citizens have high awareness and motivation to protect their data and privacy, the work needed to do this is extensive and limits useability.

The changing of privacy settings, checking cookies for every site visited, clearing out settings on a regular basis and similar activities adds practical barriers to efficient use of systems. Where digital skills and awareness are low this becomes an unsurmountable barrier. There is an argument that these practical barriers are part of the 'dark patterns' and are designed to make the actions most useful to the platforms or companies 'simplest/default'. Conversely, they make asserting different digital rights difficult or impossible if the service is to be used (Forbrukerrådet, 2018). This behaviour of 'taking the easiest route' or accepting through not really consenting is very much the 'new normal'.

6.3 Policy recommendations

Our full discussion of educational and policy recommendations are presented in an accompanying report. We designed and began this project with the intention of developing educational materials to address issues of data literacy. Much has happened in the time since the proposal for the project was submitted. Not least the Covid-19 pandemic which made our reliance on 'datafied' technologies even deeper, forcing many to use more digital services for their everyday activities. Further data breaches and public debate, issues of bias in AI and machine learning and the impacts of mis-/dis-/mal-information have become very clear. A result of this, there has been a growing academic industry of education and training tools (mostly online) that seek to provide the critical data and digital literacies citizens need to address these issues. Some companies and organisations in the digital economy are making some changes to their behaviours. The move by Apple in 2021 to allow users to prevent 3rd-party data exchange and cross-application tracking being one. Drawing on Paolo Freire (1970/1996) we have therefore focused on providing a framework for educators and practitioners to take a critical democratic approach to developing data literacy and Data Citizenship. Underpinning this are seven principles drawn from our research results:

- *Any educational or awareness raising interventions must ensure citizens feel more empowered and have practical and alternative routes to enact that empowerment.*
- *Any educational or awareness raising interventions need to 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*
- *Any educational interventions or awareness raising must 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.*
- *Any educational interventions or awareness raising must ‘meet citizens where they are’ in terms of their digital and social experience and context.*
- *Any educational interventions or awareness raising must address the challenge that those adults most in need of support are very likely outside formal educational settings.*
- *Any educational interventions or awareness raising must support skills development, but must be more than skills encompassing key elements of Data Participation.*
- *Any educational interventions or awareness raising must 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.*

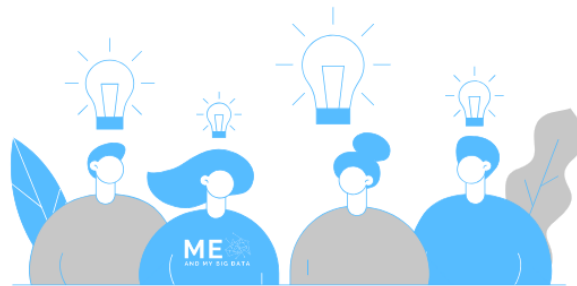
These principles focus on developing citizens Data Citizenship, but the onus cannot only be on citizens. As we discussed in section 5 responses are also needed from policy makers and the technology providers. Data extractive activities underpin the business models of most platforms used by citizens. They are therefore unlikely to change this behaviour unless the social, economic or political landscape changes. Though some (such as Apple) are now ‘selling’ privacy as key feature of their approach this is only a partial technical response and only a limited reduction in the “networked privacy” risks their users are exposed to. Such solutions may also exacerbate inequities if only high-cost devices and services offer higher protection – therefore excluding many citizens. Governments and policy makers have a clear role in setting the social and economic context in which digital systems and media – especially the current major platforms – operate. They have two clear routes for intervention:

1. *Support for and provision of data, digital and media literacy education and training*
2. *Democratic regulation and effective enforcement of digital industries*

The objective of such interventions must be more than just skills to allow them to be effective consumers (or safe online while consuming). They must empower citizens to critically examine their positions within the social, economic, and power structures of a ‘datafied’ society. The goal being to achieve a more

conscious, informed, and active way of digital living. citizens need to be able to critique and feel empowered to challenge the 'new normal' or the current state of "networked privacy".

7 The team



- Professor Simeon Yates (University of Liverpool) has run multiple UK and EU projects delivering stakeholder benefits and policy outcomes around digital media use; digital inclusion; cultural inclusion; and cultural policy. He has supported policy development around Digital Inclusion and Digital Culture both nationally (Department for Digital, Culture, Media and Sport), and regionally with Regional and Local Authorities.
- Dr. Elinor Carmi (city University) was the lead Post-Doctoral Researcher on the project and is now Lecturer in Sociology at City University in London. Elinor is a digital rights advocate, researcher and journalist who has been working, writing and teaching on deviant media, internet standards, cyber-feminism, software studies, sound studies and internet governance.
- Dr. Eleanor Lockley (Sheffield Hallam University) has collaborated with Simeon on several of the above projects. Eleanor has also undertaken research on Cyber Crime and Cyber Terrorism with a focus on human factors and best practice for crisis communication; in close collaboration with Local Authorities, the UK police, first emergency responders and citizens.
- Professor Bridgette Wessels (University of Glasgow) is Professor of the Sociology of Inequalities and founder of the Glasgow Social and Digital Change Group. Bridgette has over 20 years of experience of funded research addressing the social aspects of digital technologies.
- Dr Alicja Pawluczuk (United Nations University Institute in Macau) was a Post-Doctoral Researcher on the project and is now an ICTD Fellow working for the United Nations University Institute in Macau. Alicja's research and community education practice focuses on digital inclusion, digital, and data literacy. She is the founder of the digital inclusion and digital storytelling collective Digital Beez.
- Paul Sheppard of Critical Research supported the project survey. Critical Research is a leading UK independent research agency.

8 References

- Allmann, K. and Blank, G. (2021). Rethinking digital skills in the era of compulsory computing: methods, measurement, policy and theory. *Information, Communication & Society*, 24(5), pp.633-648.
- Andrejevic, M. (2014) Big data, big questions: the big data divide. *International Journal of Communication* 8(17), p.1673-1689.
- Arthur, C., (2021), *Social Warming: The Dangerous And Polarising Effects Of Social Media*, London: Open World Publishing
- Barassi, V. (2019) Datafied Citizens in the Age of Coerced Digital Participation. *Sociological Research Online* 24(3), p.414-429.
- Barocas, S. and Selbst, A.D. (2016) Big data's disparate impact. *Calif. L. Rev.*, 104, p.671.
- Beraldo, D., & Milan, S. (2019). From data politics to the contentious politics of data. *Big Data & Society*, 6(2)
- boyd, D. (2014). *It's complicated: The social lives of networked teens*. Yale University Press.
- Bucher, T. (2020). Nothing to disconnect from? Being singular plural in an age of machine learning. *Media, Culture & Society*, 42(4), p.610-617.
- Carlson, J., & Johnston, L. (2015). *Data information literacy: Librarians, data, and the education of a new generation of researchers*. Purdue University Press.
- Carmi, E. (2020). *Media Distortions: Understanding the Power Behind Spam, Noise and Other Deviant Media*. New York: Peter Lang
- Carmi, E (2021) A feminist Critique to digital consent. Seminar.net, 17(2).
<https://doi.org/10.7577/seminar.4291>
- Carmi, E. Yates, S.J. (2020). "What do digital inclusion and data literacy mean today?", *Internet Policy Review*, 9(2)
- 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*, Vol. 9, Issue 2.
- Cheney-Lippold, J. (2017). *We Are Data: Algorithms and the Making of Our Digital Selves*. New York: New York University Press.
- Civic Switchboard Guide, (2019) *Developing civic data literacy*. Retrieved from: <https://civic-switchboard.gitbook.io/guide/library-roles/developing-civic-data-literacy>
- Couldry, N. and Powell, A. (2014a). Big data from the bottom up. *Big Data & Society*, 1(2), p.1-5

- Couldry, N. Stephansen, H. Fotopoulou, A. MacDonald, R., Clark, W., Dickens, L. (2014b). Digital citizenship? Narrative exchange and the changing terms of civic culture. *Citizenship Studies* 18(6-7), p.615-629.
- 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
- Dencik, L., Hintz, A., Cable, J., (2019). Towards data justice. In: Bigo D, Isin E, and Ruppert E, *Data Politics*, Oxon: Routledge, pp.167-187
- Dewey, J. (1930). *Individualism, Old and New. II: The Lost Individual*. New Republic, 61, 294-296.
- Doteveryone, (2018). *People, Power and Technology: The 2018 Digital Attitudes Report*.
<https://attitudes.doteveryone.org.uk>
- Draper, N.A., Turow, J. (2019). The corporate cultivation of digital resignation. *New Media & Society*, 21(8), 1824-1839.
- Eubanks, V., 2018. *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.
- Forbrukerrådet, (2018). *Report: Deceived by design* (<https://www.forbrukerradet.no/undersokelse/no-undersokelsekategori/deceived-by-design/>)
- Frame, A., Brachotte, G., (2016). (eds.) *Citizen Participation and Political Communication in a Digital World*, London: Routledge
- Freire, P. (1970/1996). *Pedagogy of the oppressed (revised)*. New York: Continuum
- 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.
- Gutiérrez, M. (2019). Participation in a datafied environment: questions about data literacy. *Comunicação e sociedade*, (36), 37-55.
- Hall, S. (1980). Encoding/decoding. In Hall S, Hobson D, Lowe A and Willis P, (eds.) *Culture, Media, Language*, London: Hutchinson, p.28-138.

- 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.
- Hintz, A., Brown, I. (2017). Digital Citizenship and Surveillance: Enabling Digital Citizenship? The Reshaping of Surveillance Policy After Snowden. *International Journal of Communication*, 11, p.782–801.
- 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.
- ICO. (2018) <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/>
- Isaak, J. Hanna, M.J. (2018). User data privacy: Facebook, Cambridge Analytica, and privacy protection. *Computer*, 51(8), p.56-59.
- Isin, E., Ruppert, E. (2020). *Being digital citizens*. 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.
- Law, N., Woo, D., de la Torre, J., & Wong, G. (2018). *A global framework of reference on digital literacy skills for indicator 4.4.2*. Unesco Institute for Statistics.
- Lazarsfeld, P.F., Berelson, B., Gaudet, H., (1968). *The people's choice*. Columbia University Press.
- Lehtiniemi, T., Haapoja, J. (2020). Data agency at stake: MyData activism and alternative frames of equal participation. *New Media & Society*, 22(1), 87-104.
- 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.
- Lloyds, (2018). *Lloyds Bank UK Consumer Digital Index 2018*
- Mandinach, E.B., Gummer, E.S. (2013). A systemic view of implementing data literacy in educator preparation. *Educational Researcher*. 42(1), p.30-37.
- McCarthy, M.T. (2016). The big data divide and its consequences. *Sociology Compass*, 10(12), p.1131-1140.
- Mihailidis, P., Thevenin, B., (2013) Media literacy as a core competency for engaged citizenship in participatory democracy. *American Behavioral Scientist*. 57(11), p.1611-1622.
- Milan, S., Van der Velden, L. (2016). The alternative epistemologies of data activism. *Digital Culture & Society*, 2(2), 57-74.
- Noble, S. U. (2018). *Algorithms of oppression*. New York University Press.

- O'neil, C., 2016. *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.
- 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.
- Schmidt, C.D. (2018). *Google Data Collection. Report*. Retrieved from: https://www.ftc.gov/system/files/documents/public_comments/2018/08/ftc-2018-0074-d-0018-155525.pdf
- Selwyn, N., Pangrazio, L. (2018). Doing data differently? Developing personal data tactics and strategies amongst young mobile media users. *Big Data & Society*, 5(1), 2053951718765021.
- Street, B.V., Street B.B. (1984). *Literacy in theory and practice* (Vol. 9). Cambridge University Press.
- Suchman, L.A. (1987). *Plans and situated actions: The problem of human-machine communication*. Cambridge university press.
- 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.
- 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
- van Deursen, A., van Dijk, J. (2008). Using online public services: a measurement of citizens' operational, formal, information and strategic skills. In *International Conference on Electronic Government* (pp. 195-206). Springer, Berlin, Heidelberg.
- Véliz, C. (2020). *Privacy is power*. Random House Australia.
- White, S.A. (2003). (ed.) *Participatory video: Images that transform and empower*. Sage.
- 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.
- Yates S.J., Lockley, E., (2018),. "Social media and social class", *American Behavioural Scientist*, 62(9), p.1291-1316

Yates, S.J., Blejmar, J. (2020) “Future Research on the Social, Organisational, and Personal Impacts of Automation: Findings from Two Expert Panels” in Yates, S.J., Rice, R., *The Oxford Handbook of Digital Technology and Society*, Oxford: Oxford University Press.

Yates, S.J., Carmi, E., Lockley, E., Pawluczuk, A., French, T., Vincent, S., (2020b), “Who are the limited users of digital systems and media? An examination of UK evidence”. *First Monday*, 25(6). (ISSN:1396-0466).

Yates S.J., Carmi, E., Pawluczuk, A., Lockley, E., Wessels, B., Gangneux, J. (2020a), *Me and My Big Data Report 2020: Understanding citizens' data literacies: thinking, doing & participating with our data*, 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)

Yates, S.J., Wessels, B., Hepburn, P., Frame, A., Weerakkody, V., (2020c), “ESCR Review: Citizenship and Politics” in Yates, S.J., Rice, R., *The Oxford Handbook of Digital Technology and Society*, Oxford: Oxford University Press.

Yates, S.J., Lockley, E., (2008), “Moments of Separation: Gender, (Not So Remote) Relationships, and the Mobile Phone” in Holland, S., (ed.), *Remote Relationships In A Small World*, Jones, S., (ed.), Digital Technologies and Everyday Life Series, Oxford: Peter Lang Publishing

Yates, S.J., Rice, R., (2020), *The Oxford Handbook of Digital Technology and Society*, Oxford: Oxford University Press.

Zuboff, S. (2019) *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. London: Profile Books.

9 Methods and results annex

9.1 Survey sample design and administration

Critical Research undertook the survey work. Critical have both specific and extensive experience of undertaking internet and digital media focused research having undertaken survey work for the Ofcom media literacy research programme. Critical utilised in-home survey work, using a computer-aided personal interview methodology. The survey quota sample was sourced from UK Geographics and broken into sampling points using Census 2011 Output Areas (OAs). 125 sampling points were used to achieve a maximum of n = 1,542 interviews. These points were selected to be a representative cross section of UK addresses. Quotas were set to be reflective of the UK internet using population by age, gender, and household socio-economic group, and urbanity. Post survey weightings for respondents were calculated by Critical and provided with the data set. Survey results were provided to the research team in SPSS and CSV formats.

9.2 Data analysis

All data analyses were carried out by team members. Reported statistical analyses were conducted using:

- R (v 4.0.5 (2021-03-31)) running under RStudio (v 1.4.1106) and MacOS 11.5.2
- IBM SPSS Statistics 27 running under MacOS 11.5.2

Specific key analyses utilised:

- Latent Class Analysis utilised the 'poLCA' R package (v 1.4.1)
- Multiple Correspondence Analysis utilised the 'ca' R package (v 0.71.1)
- Multinomial Regression utilised the 'nnet' R package (v 7.3-16)
- All graphs and charts utilised the 'ggplot2' R package (v 3.3.5)

9.3 Data tables

Table 11: Proportions of user types from prior studies

Group description (percentages)	2013	2015	2018	Current survey (2019)
Extensive political users	4.6	5.2	21.3	8.1
Non-political extensive users	11.2	15.4	15.2	17.1
General (no social media) users	8.9	6.7	7.4	26.7
Social and entertainment media users	9.7	14.5	19.8	14.8
Information seeking limited user	12.7	14.6	N/A	N/A
Formal transaction users	11.8	N/A	N/A	N/A
Limited (social media)	N/A	14.4	10.0	N/A
Limited users	17.2	14.3	11.5	19.1
Non-users	23.8	14.9	14.9	14.3

Table 12: Multinomial regression demographic model of user types

	Dependent variable: baseline 'Extensive political'			
	Extensive	Social and media	General (no social media)	Limited
Education – baseline 'No secondary education'				
Still at school	–1.566 (0.980)	–2.003** (0.871)	–2.363** (1.055)	–1.922** (0.979)
GCSE or equivalent (No Maths and English)	1.254 (1.217)	0.731 (1.137)	1.404 (1.133)	0.654 (1.129)
GCSE or equivalent (With Maths or English)	–0.205 (0.658)	–0.895 (0.561)	–0.871 (0.572)	–1.169** (0.556)
Vocational	–0.366 (0.642)	–1.676*** (0.557)	–0.852 (0.552)	–1.820*** (0.544)
A' level or equivalent	–0.216 (0.684)	–1.864*** (0.620)	–1.245** (0.609)	–2.206*** (0.620)
Diplomas in higher education or equivalent	0.109 (0.680)	–2.232*** (0.646)	–0.580 (0.597)	–2.414*** (0.627)
University first degree	–0.353 (0.661)	–2.613*** (0.616)	–0.985* (0.575)	–2.644*** (0.599)
University higher degree	0.281 (0.742)	–2.864*** (0.870)	–0.673 (0.663)	–2.251*** (0.717)
Age – baseline '16-24'				
25-44	–0.659* (0.347)	–0.872** (0.356)	–0.079 (0.359)	–0.149 (0.399)
45-64	–0.806** (0.383)	–0.503 (0.387)	0.665* (0.380)	0.847** (0.417)
65+	–1.498*** (0.548)	–1.033* (0.571)	0.784 (0.488)	2.014*** (0.515)
Home – baseline 'Outright owned'				
Being bought on mortgage	–0.110 (0.331)	0.610 (0.378)	0.152 (0.308)	–0.396 (0.348)
Rented from Local Authority/ Housing Association/ Trust	–0.157 (0.404)	0.493 (0.424)	–0.563 (0.378)	–0.167 (0.389)
Rented from private landlord	0.248 (0.404)	1.066** (0.443)	0.027 (0.391)	0.499 (0.415)
Something else/Refused/Don't know	–0.031 (0.604)	1.072* (0.616)	0.247 (0.567)	–0.971 (0.699)
NRS social grade – baseline 'AB'				
C1	0.011 (0.294)	0.160 (0.355)	0.274 (0.284)	0.715** (0.343)
C2	–0.275 (0.377)	0.564 (0.406)	0.356 (0.354)	0.830** (0.406)
DE	–0.776* (0.425)	0.247 (0.439)	0.102 (0.394)	0.604 (0.437)
Constant				
	1.753** (0.765)	1.894*** (0.731)	1.544** (0.696)	1.398* (0.725)

Note: *p<0.1; **p<0.05; ***p<0.01

Table 13: Latent class by Education (Percentages)

Latent class percentages	EP	E	S&M	G	L
No formal qualifications/refused	4	3	18	8	25
Still at school	3	2	3	0	1
GCSE or equivalent (No Maths and English)	1	3	8	7	9
GCSE or equivalent (With Maths or English)	14	13	29	12	21
Vocational	22	16	18	20	21
A' level or equivalent	12	14	10	8	6
Diplomas in higher education or equivalent	11	16	5	15	5
University first degree	26	25	7	21	7
University higher degree	7	11	1	9	4

$\chi^2(32, 1322) = 246.669, p = 0.000, \text{Cramer's } V = 0.216$

Medium/Large effect size

Table 15: Latent class by age

Latent class percentages	EP	E	S&M	G	L
16-24	14	22	31	9	8
25-44	50	49	39	34	23
45-64	26	24	26	40	33
65+	10	5	4	16	36

$\chi^2(12, 1322) = 229.418, p = 0.000, \text{Cramer's } V = 0.24$

Medium/Large effect size

Table 17: Likelihood of checking search engine results

Checking social media percentages	EP	E	S&M	G	L
Don't check	16	20	39	31	55
Some checks	84	80	61	69	45

$\chi^2(4, 1322) = 102.414, p = 0.000, \text{Cramer's } V = 0.278$

Medium effect size

Table 19: Latent class by awareness of reasons for data collection

Reasons for use awareness percentages	EP	E	S&M	G	L
Unaware of reasons	2	2	12	4	21
Some knowledge of reasons	98	98	88	96	79

$\chi^2(4, 1322) = 94.223, p = 0.000, \text{Cramer's } V = 0.267$

Medium effect size

Table 14: Latent class by NRS social grade

Latent class percentages	EP	E	S&M	G	L
AB	33	35	12	29	14
C1	32	38	26	35	33
C2	17	15	28	20	23
DE	18	11	34	16	31

$\chi^2(12, 1322) = 110.14, p = 0.000, \text{Cramer's } V = 0.167$

Medium effect size

Table 16: Latent class by likelihood of checking social media content

Checking social media percentages	EP	E	S&M	G	L
Don't check	20	16	38	40	78
Some checks	80	84	62	60	22

$\chi^2(4, 1322) = 259.152, p = 0.000, \text{Cramer's } V = 0.443$

Medium effect size

Table 18: Latent class by awareness of data collected by platforms

Collection awareness percentages	EP	E	S&M	G	L
Unaware of collection	6	3	15	6	28
Some knowledge of collection	94	97	85	94	72

$\chi^2(4, 1322) = 114.217, p = 0.000, \text{Cramer's } V = 0.294$

Medium effect size

Table 20: Acceptability of advertising data collection

Acceptability of data collection percentages: advertising	EP	E	S&M	G	L
Acceptable	36	49	37	34	21
Neutral	19	19	17	15	17
Not acceptable	45	32	46	51	63

$\chi^2(8, 1322) = 64.247, p = 0.000, \text{Cramer's } V = 0.156$

Small effect size

Table 21: Acceptability of profiling data collection

Acceptability of data collection percentages: profiling	EP	E	S&M	G	L
Acceptable	39	48	33	28	21
Neutral	17	24	18	20	14
Not acceptable	44	29	49	51	65
$\chi^2(8, 1322) = 80.538, p = 0.000, \text{Cramer's } V = 0.174$					
Small effect size					

Table 22: Acceptability of tailored services data collection

Acceptability of data collection percentages: tailored services	EP	E	S&M	G	L
Acceptable	49	55	43	34	28
Neutral	22	16	16	25	19
Not acceptable	29	29	40	41	53
$\chi^2(8, 1322) = 66.635, p = 0.000, \text{Cramer's } V = 0.159$					
Small effect size					

Table 23: Acceptability of personalized apps data collection

Acceptability of data collection percentages: personalised apps	EP	E	S&M	G	L
Acceptable	50	59	44	37	25
Neutral	23	17	18	20	19
Not acceptable	27	24	38	43	56
$\chi^2(8, 1322) = 89.266, p = 0.000, \text{Cramer's } V = 0.184$					
Small effect size					

Table 24: Acceptability of data collection for sale

Acceptability of data collection percentages: selling data	EP	E	S&M	G	L
Acceptable	12	4	11	3	5
Neutral	6	7	9	7	7
Not acceptable	82	89	79	89	88
$\chi^2(8, 1322) = 30.205, p = 0.000, \text{Cramer's } V = 0.107$					
Small effect size					

Table 25: Acceptability of data collection for online tracking

Acceptability of data collection percentages: track online	EP	E	S&M	G	L
Acceptable	17	16	19	14	16
Neutral	22	21	20	17	12
Not acceptable	62	63	61	69	72
$\chi^2(8, 1322) = 15.807, p = 0.045, \text{Cramer's } V = 0.077$					
Very small effect size					

Table 26: Acceptability of data collection to influence behaviour

Acceptability of data collection percentages: influence opinions	EP	E	S&M	G	L
Acceptable	9	10	13	10	12
Neutral	16	14	16	14	12
Not acceptable	75	76	71	75	76
$\chi^2(8, 1322) = 4.633, p = 0.796, \text{Cramer's } V = 0.042$					

Table 27: Comfort with 3rd party data sharing

Comfort with third party sharing percentages	EP	E	S&M	G	L
Don't know	0	1	3	0	2
Uncomfortable sharing any items	66	63	61	74	71
Comfortable sharing some items	34	35	36	25	27
$\chi^2(8, 1322) = 22.947, p = 0.003, \text{Cramer's } V = 0.093$					
Very small effect size					

Table 28: 'No point changing privacy settings'

'No point changing settings' agreement percentages	EP	E	S&M	G	L
Disagree	41	52	41	48	38
Neutral	17	14	11	18	17
Agree	42	34	48	34	45
$\chi^2(8, 1322) = 24.94, p = 0.002, \text{Cramer's } V = 0.097$					
Small effect size					

Table 29: 'Changing settings is too much effort'

'Changing settings too much effort' agreement percentages	EP	E	S&M	G	L
Disagree	54	68	40	55	44
Neutral	14	11	18	17	19
Agree	32	20	42	29	37
$\chi^2(8, 1322) = 53.052, p = 0.00, \text{Cramer's } V = 0.142$					
Small effect size					

Table 31: 'Don't mind sharing if service is free'

'Don't mind sharing if service is free' agreement percentages	EP	E	S&M	G	L
Disagree	55	54	55	64	65
Neutral	24	29	19	20	16
Agree	21	17	26	17	19
$\chi^2(8, 1322) = 25.713, p = 0.01, \text{Cramer's } V = 0.099$					
Small effect size					

Table 33: Trust in friends social media posts

Trust in friend's social media posts percentages	EP	E	S&M	G	L
Disagree	31	32	24	38	60
Neutral	33	32	21	31	9
Agree	36	35	55	31	31
$\chi^2(8, 1322) = 124.819, p = 0, \text{Cramer's } V = 0.217$					
Medium effect size					

Table 30: 'Companies make privacy settings easy'

'Companies make privacy settings easy' agreement percentages	EP	E	S&M	G	L
Disagree	39	39	48	43	54
Neutral	31	30	23	27	26
Agree	30	31	29	31	20
$\chi^2(8, 1322) = 22.196, p = 0.05, \text{Cramer's } V = 0.092$					
Small effect size					

Table 32: 'Don't want to share but have no choice'

'I don't want to share but have no choice' agreement percentages	EP	E	S&M	G	L
Disagree	18	19	16	21	27
Neutral	22	14	12	11	15
Agree	61	67	72	68	58
$\chi^2(8, 1322) = 23.289, p = 0.03, \text{Cramer's } V = 0.094$					
Small effect size					

Table 34: Overall trust in social media posts

Overall trust in social media posts percentages	EP	E	S&M	G	L
Disagree	79	89	80	86	88
Neutral	14	6	9	9	6
Agree	7	5	11	5	5
$\chi^2(8, 1322) = 19.197, p = 0.014, \text{Cramer's } V = 0.085$					
Small effect size					

Table 35: Mostly read social media that shares own values

Mostly read social media that shares own values percentages	EP	E	S&M	G	L
Disagree	35	31	24	39	67
Neutral	30	32	25	28	11
Agree	34	37	51	33	22

$\chi^2(8, 1322) = 134.228, p = 0, \text{Cramer's } V = 0.225$

Medium effect size

Table 37: Information sources depend on social media links percentages

Information sources depend on social media links percentages	EP	E	S&M	G	L
Disagree	17	19	18	31	61
Neutral	22	16	19	17	8
Agree	62	65	63	51	31

$\chi^2(8, 1322) = 169.668, p = 0, \text{Cramer's } V = 0.253$

Medium effect size

Table 39: Data participation - rights actions

Data participation rights percentages	EP	E	S&M	G	L
No rights actions	26	22	66	43	79
Some rights actions	74	78	34	57	21

$\chi^2(4, 1322) = 240.991, p = 0.000, \text{Cramer's } V = 0.427$

Medium effect size

Table 41: Data Participation - data collection

Data Participation – data collection percentages	EP	E	S&M	G	L
No information collection	15	14	51	29	65
Some information collection	85	86	49	71	35

$\chi^2(4, 1322) = 214.142, p = 0.000, \text{Cramer's } V = 0.402$

Medium effect size

Table 36: Read social media with different political perspectives

Try to read social media with different political perspectives percentages	EP	E	S&M	G	L
Disagree	33	29	41	42	72
Neutral	34	33	29	28	14
Agree	34	38	31	30	14

$\chi^2(8, 1322) = 127.491, p = 0, \text{Cramer's } V = 0.220$

Medium effect size

Table 38: Data participation activities

Data participation (simple binary) percentages	EP	E	S&M	G	L
No participation	20	14	57	35	72
Some participation	80	86	43	65	28

$\chi^2(4, 1322) = 250.619, p = 0.000, \text{Cramer's } V = 0.435$

Medium to large effect size

Table 40: Data participation - helping others

Data participation helping percentages	EP	E	S&M	G	L
No helping actions	40	34	77	58	90
Some helping actions	60	66	23	42	10

$\chi^2(4, 1322) = 235.798, p = 0.000, \text{Cramer's } V = 0.422$

Medium effect size

Table 42: Data Participation - data manipulation

Data Participation – data manipulation percentages	EP	E	S&M	G	L
No information manipulation	36	32	73	50	86
Some information manipulation	64	68	27	50	14

$\chi^2(4, 1322) = 225.109, p = 0.000, \text{Cramer's } V = 0.413$

Medium effect size

9.4 Figures

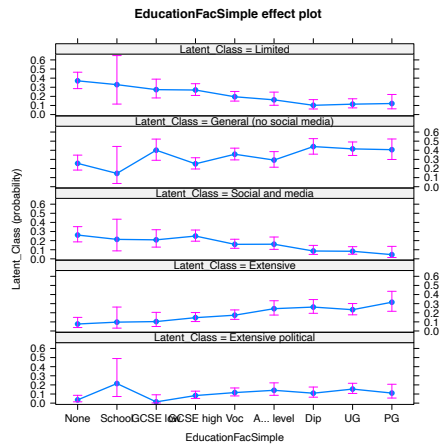


Figure 51: Latent class regression probabilities by education

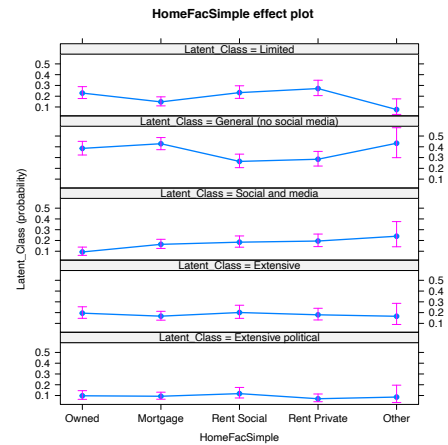


Figure 52: Latent class regression probabilities by home ownership

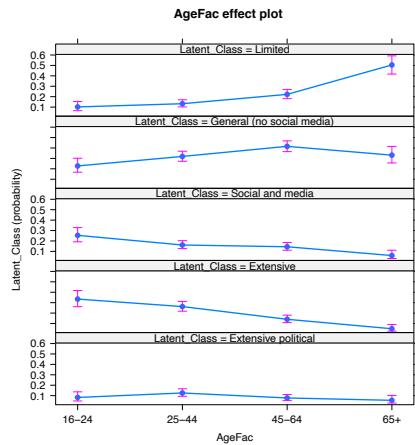


Figure 53: Latent class regression probabilities by age

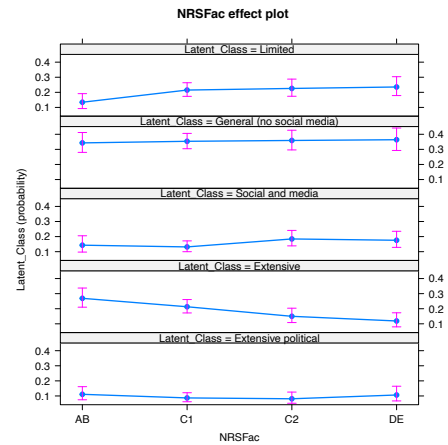


Figure 54: Latent class regression probabilities by NRS social grade

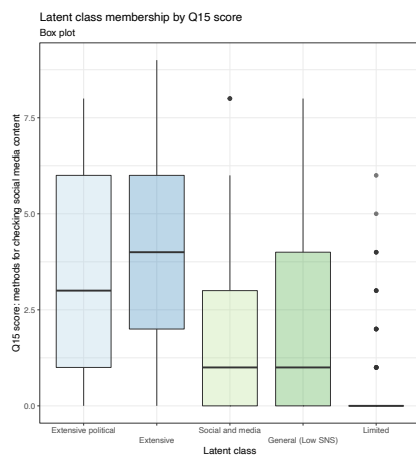


Figure 55: Box plot of the range of checking of social media content by user types

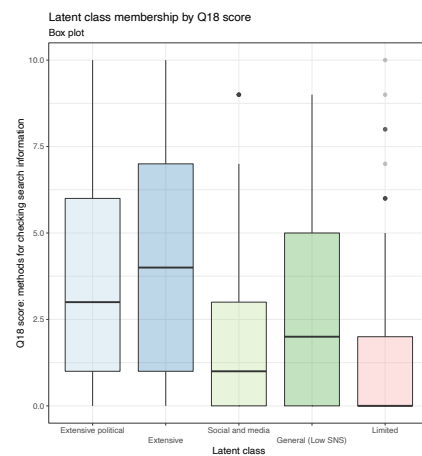


Figure 56: Box plot of the range of checking and search engine content by user types

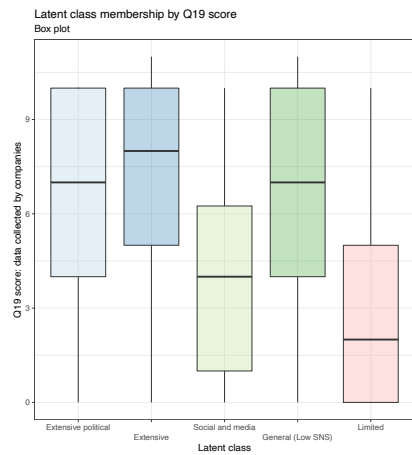


Figure 57: Range of awareness of data collected by companies by user types

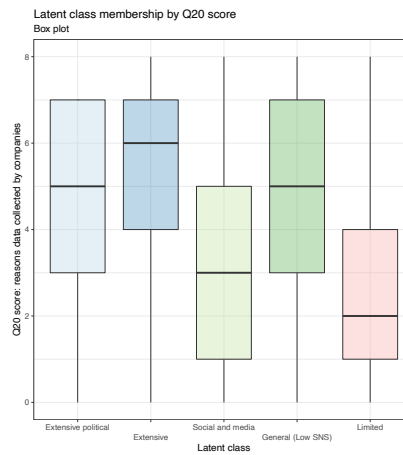


Figure 58: Range of awareness of reasons for collection by user types

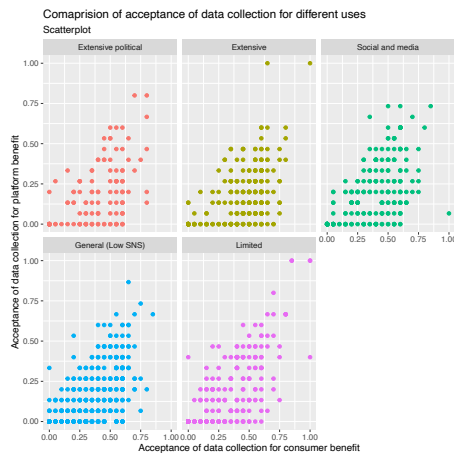


Figure 59: Bias towards acceptability of data collection for customer benefit vs platform benefit

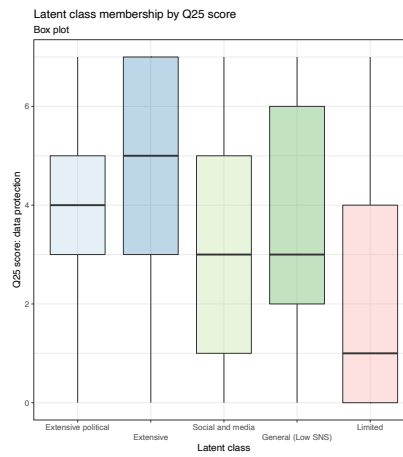


Figure 60: Box plot: Confidence in data protection score by latent class

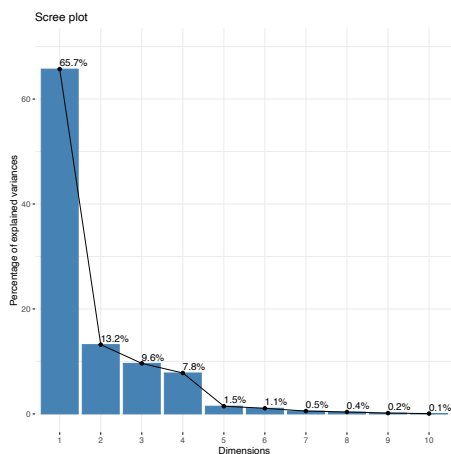


Figure 61: Scree plot of contributing dimensions to MCA

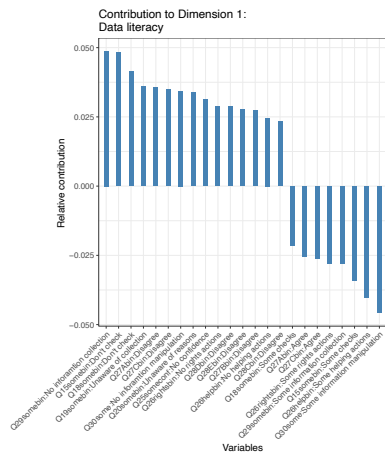


Figure 62: Key contributing variables MCA Dimension 1 (65.7%)

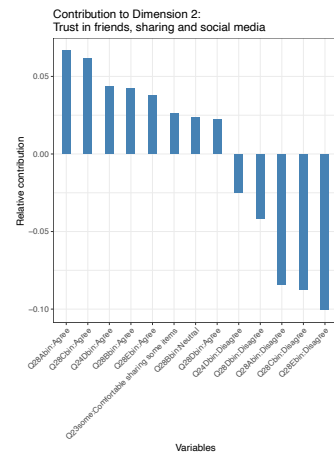


Figure 63: Key contributing variables MCA Dimension 2 (13.2%)

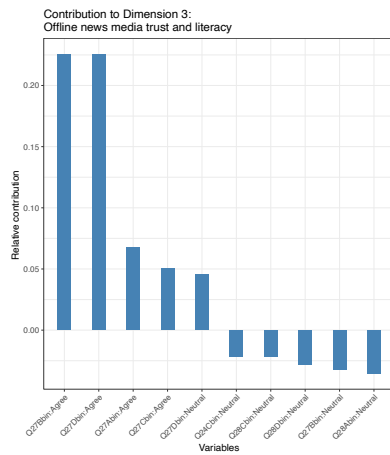


Figure 64: Key contributing variables MCA Dimension 3 (9.6%)

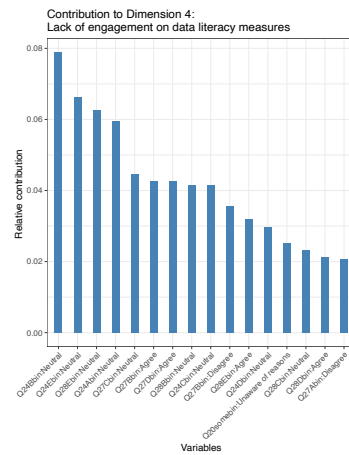


Figure 65: Key contributing variables MCA Dimension 4 (7.8%)

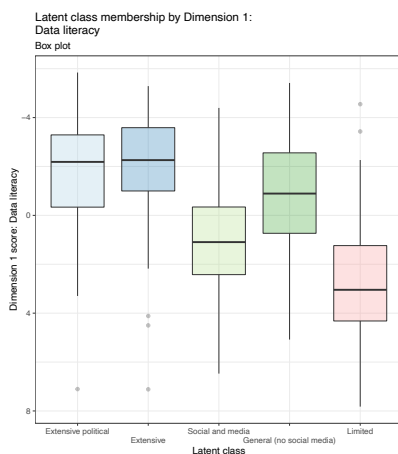


Figure 66: Dimension 1: Data literacy scores by latent class box plot (ANOVA: $F(4, 1317) = 239.44$, $p = 0.000$, $\eta^2 = 0.42$, large effect size)

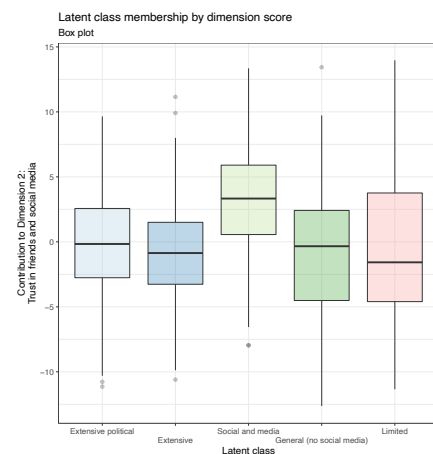


Figure 67: Dimension 2: Trust in content from friends and on social media by latent class box plot (ANOVA: $F(4, 1317) = 740.11$, $p = 0.000$, $\eta^2 = 0.10$, medium effect size)

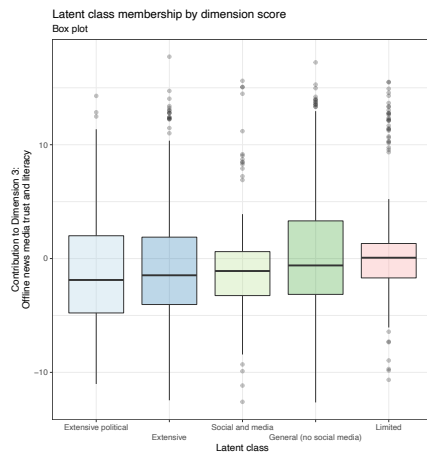


Figure 68: Dimension 3: Trust in and use of 'offline' media by latent class box plot (ANOVA: $F(4, 1317) = 147.39, p = 0.000, \eta^2 = 0.02$, small effect size)

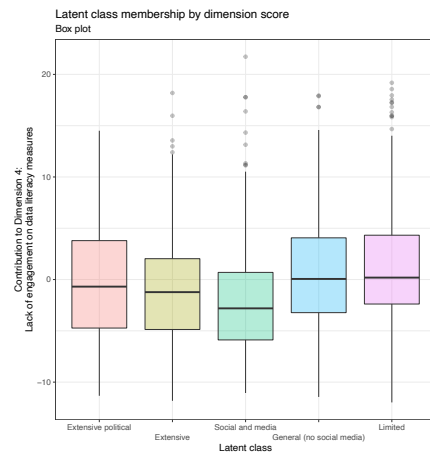


Figure 69: Dimension 4: Lack of engagement with nor neutrality on data literacy elements by latent class box plot (ANOVA: $F(4, 1317) = 13.43, p = 0.000, \eta^2 = 0.04$, small effect size)

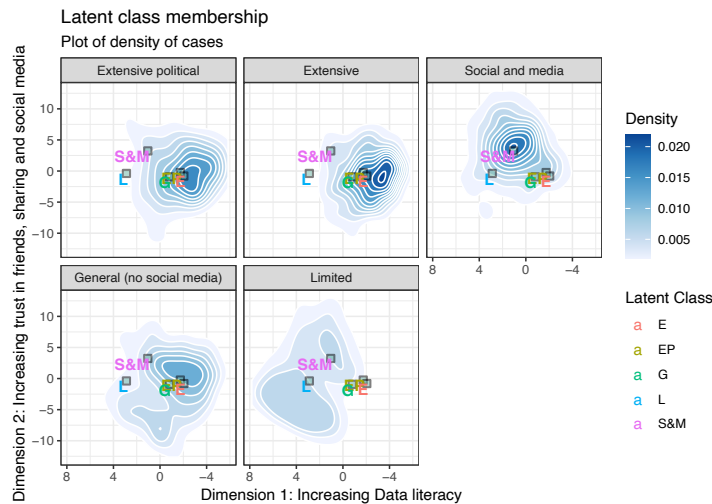


Figure 70: MCA density plots of respondents by data literacy and trust in social media split by user types

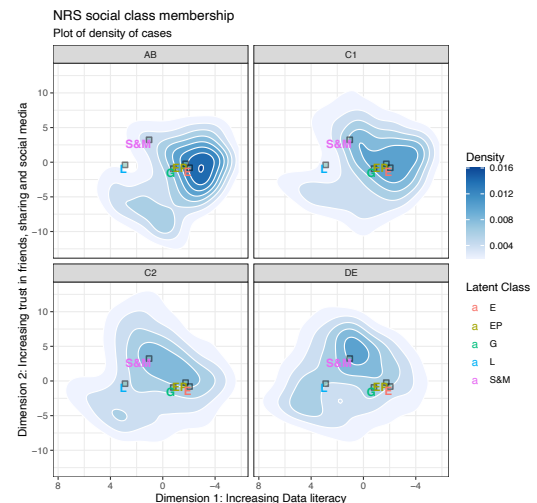


Figure 71: MCA density plots of respondents by data literacy and trust in social media split by NRS social grade

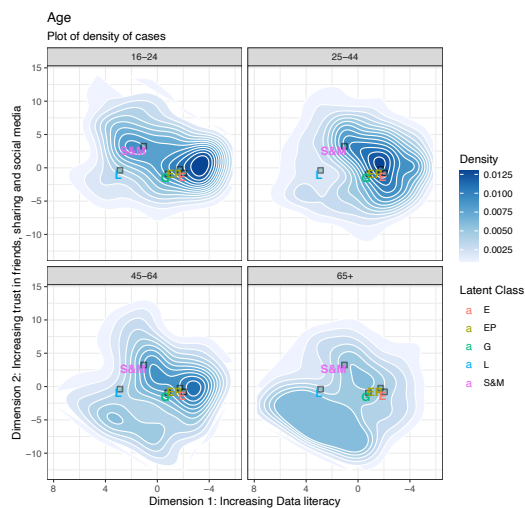


Figure 72: MCA density plots of respondents by data literacy and trust in social media split by age groups

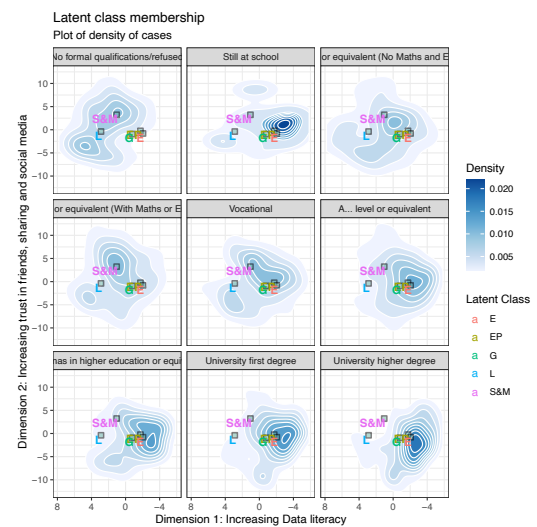


Figure 73: MCA density plots of respondents by data literacy and trust in social media split by education level

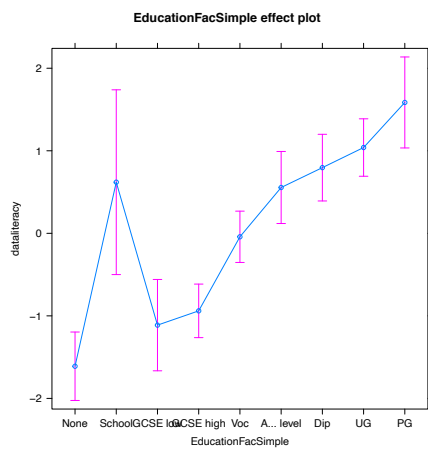


Figure 74: Regression model confidence intervals for data literacy by Education Level

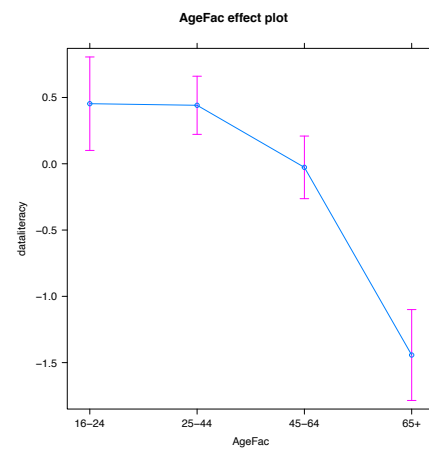


Figure 75: Regression model confidence intervals for data literacy by Age

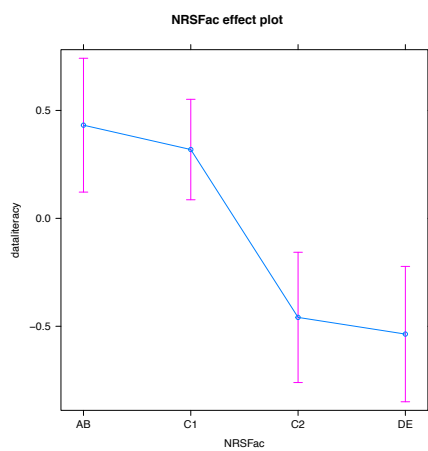


Figure 76: Regression model confidence intervals for data literacy by NRS social grade

9.5 Focus group details

Table 43: Focus group participants

Focus group	Target group	Attendees	Likely User Types	Format
A	Younger people (<30 - preferably under 25) who may have some digital skills and/or are social media focused	- Participant A1: M; 22 years old; post-18 education. - Participant A2: F; 22 years old; post-18 education. - Participant A3: M; 24 years old; post-18 education. - Participant A4: F; 22 years old; post-18 education. - Participant A5: F; 26 years old; post-18 education. - Participant A6: F; 25 years old; post-18 education. - Participant A7: M; 22 years old; post-18 education. - Participant A8: M; 27 years old; post-18 education. - Participant A9: M; 29 years old; post-18 education.	Social and entertainment media users Or Extensive users	All respondents on Zoom
B	Younger people (<30 - preferably under 25) who may have some digital skills and/or are social media focused	- Participant B1: M; 26 years old; no post-18 education - Participant B2: M; 24 years old; post-18 education. - Participant B3: M; 25 years old; post-18 education. - Participant B4: M; 27 years old; no post-18 education. - Participant B5: M; 23 years old; post-18 education.	Social and entertainment media users	All respondents on Zoom
C	Younger people (<30 - preferably under 25) who may have some digital skills and/or are social media focused	- Participants C1: F; 29 years old; no post-18 education. - Participant C2: M; 23 years old; post-18 education. - Participant C3: M; 24 years old; no post-18 education. - Participant C4: M; 18 years old; just starting post-18 education.	Social and entertainment media users	All respondents on Zoom
D	"Older adults (55+) who are offline or with limited digital skills"	- Participant D1: M; 56 years old; no post-18 education. - Participant D2: F; 55 years old; post-18 education. - Participant D3: M; 54 years old; no post-18 education. - Participant D4: M; 62 years old; post-18 education. - Participant D5: M; 52 years old; no post-18 education. - Participant D6: M; 62 years old; no post-18 education. - Participant D7: F; 58 years old; no post-18 education.	General or Limited users	All respondents on Zoom
E	"Older adults (55+) who are offline or with limited digital skills"	- Participant E1: M; 45+ years old, no post-18 education. - Participant E2: F; 45+ years old, no post-18 education. - Participant E3: M; 45+ years old, no post-18 education. - Participant E4: F; 45+ years old, no post-18 education. - Participant E5: F; 45+ years old, no post-18 education. - Participant E6: F; Unknown, facilitator.	General or Limited users	Respondents socially distanced in centre – researchers on Zoom via data projector Limited demographics provided
F	"Older adults (55+) who are offline or with limited digital skills"	- Participant F1: F; 45+ years old, no post-18 education. - Participant F2: F; 45+ years old, no post-18 education. - Participant F3: F; 45+ years old, no post-18 education. - Participant F4: M; 45+ years old, no post-18 education.	General or Limited users	Respondents socially distanced in centre – researchers on Zoom via data projector Limited demographics provided

G	Older adults with digital skills	• Participants G1: M&F: 74 and 70; post-18 educations. • Participant G2: age not provided – retired; post-18 education. • Participant G3: F data not provided. • Participant G4: F: 84 years old; post-18 education. • Participant G5: F: 71 years old; post-18 education. • Participant G6: F: 68 years old; post-18 educations. • Participant G7: M: 78 years old; post-18 educations.	General users	All respondents on Zoom
H	Older adults with digital skills	• Participant H1: M; 64 years old; post-18 education. • Participant H2: M; 60 years old; post-18 education. • Participant H3: M; 41 years old; post-18 education. • Participant H4: F; no age given; post-18 education. • Participant H5: F; 72 years old; post-18 education.	General or Extensive users	All respondents on Zoom
I	Older adults (55+) who are offline or with limited digital skills	• Participant I1: F; 70 years old; no post-18 education. • Participant I2: F; 74 years old; post-18 education. • Participant I3: F; 60 years old; post-18 education.	General or Limited users	All respondents on Zoom
J	Older adults (55+) who are offline or with limited digital skills	• Participant J1: F; 90+ years old; post-18 education. • Participant J2: M; 73 years old; post-18 education. • Participant J3: F; years old; post-18 education. • Participant J4: F; age not provided; likely post-18 education.	General or Limited users	All respondents on Zoom
K	Adults with limited digital skills	• Participant K1: M; 46 years old; no post-18 education. • Participant K2: F; 52 years old; no post-18 education.	Limited users	All respondents on Zoom
L	Adults with limited digital skills	• Participant L1: M; 57 years old; no post-18 education. • Participant L2: M; 61 years old; no post-18 education. • Participant L3: F; 46 years old; no post-18 education. • Participant L4: M; 57 years old; no post-18 education.	General or Limited users	All respondents on Zoom
M	Post-18 education students with higher digital skills	• Participant M1: M; 20 years old, undertaking post-18 education. • Participant M2: M; 20 years old, undertaking post-18 education. • Participant M3: M; 20 years old, undertaking post-18 education. • Participant M4: M; 20 years old, undertaking post-18 education. • Participant M5: F; 20 years old, undertaking post-18 education.	Extensive Or Extensive political users	All respondents on Zoom
N	Post-18 education students with higher digital skills	• Participant N1: F; 21 years old; undertaking post-18 education. • Participant N2: M; 22 years old; undertaking post-18 education. • Participant N3: F; 23 years old; undertaking post-18 education. • Participant N4: M; 25 years old; undertaking post-18 education. • Participant N5: F; 26 years old; undertaking post-18 education.	Extensive Or Extensive political users	All respondents on Zoom

Table 44: Focus group schedule

Description		Questions	Examples of questions in an accessible language	Notes
Discussion Part 1: What do we mean by data?	The aim is to introduce participants into the idea of data and how their undertaking of what data is.	1. What comes to your mind when you think about data? 2. Where do you get information about different things from? 3. What kinds of data do you think you may be sharing?	What is the first thing you think about when you think about the word Data? (It honestly can be anything - this is about you and what you think!)	Facilitators: allow respondents to express things in their own terms. Only provide examples if respondents are struggling. Try to note key examples to help with feedback and setting of baseline.
Provide a baseline definition	Present examples of using and sharing data - building on what has been presented by participants. As a baseline for Part 2.			
Activity 1: Data in Day Mapping (group activity) (Data Doing) 15 Minutes	Creating a graphical representation of your daily data activities	1. Thinking about your data in the day - could tell us about different activities (e.g. checking social media, using school apps) 2. Please down the different activities you undertake during the day. Ask participants to reflex on these.	Lets now think about your normal day - and think about where data fits into it? (Think about when you wake up, things you do in the morning, lunchtime, afternoon, evening - what kinds of things might you do that involve data? This might be using your mobile phone, or connecting with your friends and family via social media or shopping, contacting people - (use other examples already noted by the respondents)	Facilitators: Draw out themes, organisations, and issues identified in the survey and literature - as they relate to Data Doing.
Discussion Part 2:	Exploring citizens everyday networks of data literacy.	1. Thinking about the data in your life - how and who do you share it with? 2. Do you understand what happens to your data? 3. Is there data that you are not comfortable sharing and why? 4. Are there organisations/companies you are not comfortable sharing data with?	Now after you have had a chance to discuss and understand how we are looking at 'data': 1. Who do you think you usually share it with, this could be not only people but also organisations like your supermarket. 2. Do you ever think about what happens to your data? 3. Do you think you understand it? 4. Would you want to understand it better?	Facilitators: allow respondents to express things in their own terms. Be more directive on drawing out examples. Especially around data sharing, trust, helping each other and key people in networks of interaction.
Activity 2: Examining the data maps (Data Thinking) 15 Minutes	Exploring how citizens think about data and its use by organisations and others.	1. How do you feel about your skills when doing these things? 2. Are you confident about what is going on what your data? 3. Do you ever wonder about privacy?	1. How do you feel when you do these things, do you feel confident, worried, unsure, don't care? 2. What does privacy mean to you? Is this something you care about, worried about? 3. Do you think that you know what is happening	Facilitators: Draw out themes, organisations, and issues identified in the survey and literature - as they relate to Data Thinking.

			inside your computer or phone?	
Activity 3: Improving making maps (Data Participation) 15 Minutes	Exploring how citizens use data to achieve things in their everyday lives. Including how they use it for their community?	1. How do you participate in data-driven environments? 2. If you participate online, which types of activities do you take part in?	1. Is anyone sharing data for you or helping you with your data? 2. Posted a picture to show your support for an issue that is important to you. 3. Have encouraged others to share data or use data to help with issues that are important to you. 4. Finding other people who share your views about important issues.	Facilitators: Draw out themes, organisations, and issues identified in the survey and literature - as they relate to Data Participation.
Broad themes	1. Discussion Part 1: What comes to your mind when you think about data? 2. Discussion Part 2: How does data fit into your life? 3. Activity: mapping exercise - could you present your data daily activities? Who is involved in your data (Data Doing)? 4. Activity mapping 2: how do you feel about these activities? Do you feel confident about your skills? (Data Thinking) 5. Activity mapping 3: if you could improve your daily data? Have you done anything to improve your data reality? (Data Participation)			