



Public understanding and perceptions of data practices: a review of existing research

Helen Kennedy, Susan Oman, Mark Taylor, Jo Bates, Robin Steedman

May 2020



Contents

1. Introduction	3
2. Findings from the review	5
2.1. People have some knowledge and understanding of data practices	5
2.2. People are concerned about data practices	10
2.3. People are not <i>only</i> concerned: they have some agency, they negotiate or resist data, and sometimes have contradictory views.....	13
2.4. Feelings play an important role in public understanding and perceptions of data practices	17
2.5. Trust or distrust?	19
2.6. Acceptance or resignation?	23
2.7. What fair data practices look like, from public perspectives	25
2.8. State and industry need to enact change, according to the research	29
2.8.1. Systems-oriented literature recommends design and systems change	30
2.8.2. Critical academic literature recommends changes to regulation of markets and industry	33
2.8.3. Policy and practice-oriented grey literature recommends a range of changes, often involving state or governmental actors	34
2.9. Attention to differences is growing, but more is needed	37
2.10. Methods and framing shape findings	44
3. Conclusions	48
4. Recommendations	50
5. Review methodology	52
5.1. Introduction to review methodology	52
5.2. The systematic search of online databases.....	52
5.3. The snowball search (of academic and grey literature).....	54
5.4. The review of survey reports	54
5.5. Finalising bibliographic lists	54
5.6. Analysing the literature	56
6. References	57
6.1. Academic literature	57
6.2. Grey Literature.....	61
6.3. Other literature referenced (reviews and syntheses, international reports and non-empirical references)...	63
6.4. References from Introduction and Review methodology sections	64
7. About this report and its authors	65

Preferred citation: Helen Kennedy, Susan Oman, Mark Taylor, Jo Bates & Robin Steedman (2020) *Public understanding and perceptions of data practices: a review of existing research*. Living With Data, University of Sheffield. <http://livingwithdata.org/current-research/publications/>

1. Introduction

The widespread collection and use of digital data is said to have wide-ranging effects: benefits such as more effective service provision, and harms such as more surveillance, less privacy, and new forms of inequality and injustice. In this report, we use the term ‘data practices’ to refer to the systematic collection, analysis and sharing of data and the outcomes of these processes (this is sometimes also called ‘datafication’¹). The data at the centre of such practices is often personal data, and research into public perceptions often focuses on personal data, although sometimes the type of data under discussion is not specified. Personal data is defined as data ‘related to an identified or identifiable person’ by the General Data Protection Regulation (or GDPR, European Union regulation about data usage and rights²).

Concern about potential harms has led to measures to influence data governance, such as: the establishment of the government Centre for Data Ethics and Innovation (CDEI) and the independent Ada Lovelace Institute (Ada) in the UK; parliamentary inquiries (eg into data ethics, responsible uses of data and AI, and digital government); social and civil society initiatives (eg Doteveryone’s Society In the Loop event, the Open Data Institute’s ‘data trust’ pilots); and practical experiments with ethical data-driven systems. In this context, interest in how the public perceives data practices has also begun to grow, amongst academic researchers focusing on public/citizens’ views of the new role of data in society and amongst policy-makers and professionals keen to establish positive perceptions of their data-related policies and practices. Understanding public views of data practices is said to be at the heart of initiatives like CDEI and Ada, to ensure that data works ‘for people and society’ (Ada’s mission) and is ‘a force for good’ (a CDEI aim). Consequently, research into public understanding and perceptions of datafication has flourished in recent years.

We conducted a review of original empirical research into public understanding and perceptions of, attitudes towards and feelings about data practices and related phenomena (such as AI and facial recognition), in order to synthesise existing evidence and evaluate whether patterns or generalisable findings emerge. We reviewed academic and grey literature published between 2015 and the end of 2019. By grey literature, we mean literature produced by independent, civil society, third sector, governmental or commercial organisations or by academics for non-academic audiences. This document assesses the claims that are made on the basis of existing research and reflects on methods, analyses and findings. It identifies gaps in the existing research that future research can address.

This review will inform the original empirical research to be undertaken on *Living With Data: knowledge, experiences and perceptions of data practices*,³ a project funded by The Nuffield Foundation⁴. *Living With Data’s* research questions are:

- *What do different people know and feel about specific data-related practices in different domains of everyday life?*
- *What do fair data practices look like, from non-experts’ perspectives?*

¹ Mayer-Schoenberger, V & Cukier, K (2013) *Big Data: a revolution that will transform how we live, work and think*, John Murray Publishing

² European Union (2016) *General data protection regulation*, Off J Eur Union 49: L119. <https://gdpr-info.eu>

³ Living With Data: <https://livingwithdata.org/current-research/>

⁴ The Nuffield Foundation: <https://livingwithdata.org/current-research/>

Because of the proliferation of research on this topic in recent years, we have limited the research that we review in this document in the following ways:

1. **By dates:** We review literature published between 2015 and 2019, and we include some literature published in early 2020, of which we became aware whilst undertaking the review.
2. **By geography:** The grey literature we review is UK-focused (either UK-only or based on international research that included the UK). We include selected international studies in the academic literature we discuss.
3. **By populations researched/research subjects:** We exclude literature about children's understandings and perceptions of data practices, although studies of adults' perceptions of data practices relating to children are included.
4. **By domain, especially with regard to research about public perceptions of health data:** We focus primarily on domains other than health because more is known about public perceptions of health data practices and high quality syntheses have already been undertaken.
5. **By subject matter/focus:** We exclude research which focuses on perceptions of privacy, surveillance and security from our review, except where these also address attitudes to data practices.
6. **Existing evidence syntheses and reviews:** We do not carry out analysis of the literature covered by existing evidence syntheses in this document.

Section 5 Review Methodology discusses our rationale for applying these criteria in more detail.

The research that we review in this document was identified through two main search strategies:

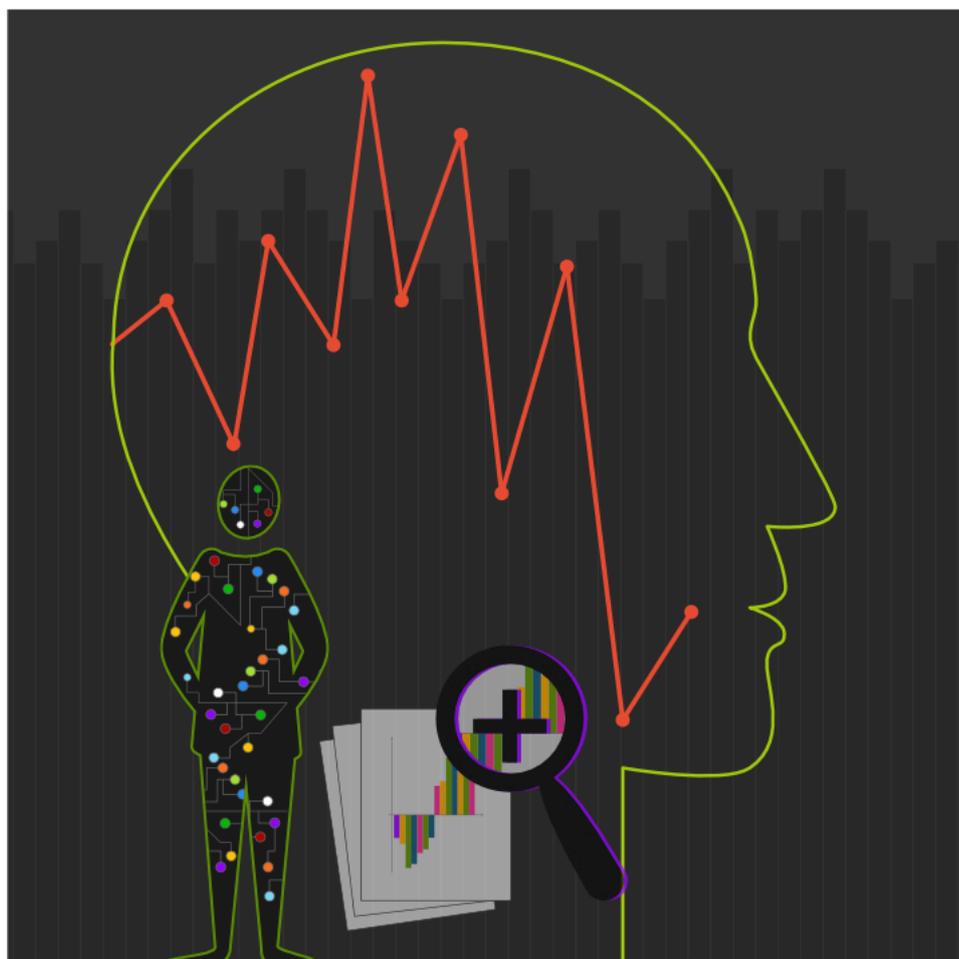
- We carried out a **systematic search** of online citation databases, using multiple keywords relating to how people feel about data practices and what happens to their personal data, focusing primarily on Web of Science.
- We carried out a **manual search** that began with grey and academic literature with which we were already familiar, and then snowballing out (eg searching bibliographies and relevant websites, observing Twitter discussions and building on word-of-mouth recommendations).

The literature that we identified through these processes was reviewed iteratively by us, the authors of this report, according to one principal inclusion criteria: does it report empirical research about how people feel about data practices and what happens to their personal data? Answers to questions like these will vary across researchers, and so will the knowledge which forms the starting point for manual searches. Searching databases also has limitations: different databases retrieve different results for equivalent searches⁵, a database such as Web of Science only retrieves results for sources that are indexed by the database, and keywords selected by author and searcher shape what is and is not found. In short, all literature and evidence searches are partial, and ours is no exception. We discuss these issues and our methodology in full in section 5 below.

We carried out the research for this report before the Covid-19 pandemic of 2020, and the research that we reviewed was also, of course, carried out before the Covid-19 crisis. The pandemic context and the new data practices that resulted from it may change people's understanding and perceptions of data practices. This is something that future research will need to address.

⁵ For a detailed discussion, see Martín-Martín, A, Orduna-Malea, E, Thelwall, M & Delgado-López-Cózar, E (2019) 'Google Scholar, Web of Science and Scopus: which is best for me?', *LSE Impact Blog*.
<https://blogs.lse.ac.uk/impactofsocialsciences/2019/12/03/google-scholar-web-of-science-and-scopus-which-is-best-for-me/>

2. Findings from the review



2.1. People have some knowledge and understanding of data practices

The authors of a report entitled *About Data About Us* claim that ‘We are often led to believe, by the press, parliament and in published surveys, that the UK public’s knowledge and understanding of data protection is low’⁶. We begin this review by assessing whether this is the case. The research we review does not refer to ‘data practices’, as this term is not widely used or commonly understood. Other terms are used, such as data use, data sharing, data privacy or data management. We bring these together in the term ‘data practices’ that we use here, but we would not expect to see this used in research with the public.

Some quantitative research has addressed the question of what people know and understand about data practices. Assessment of people’s knowledge about uses of their personal data is often undertaken with reference to GDPR, which was introduced in May 2018. Some surveys focus on awareness of the rights enshrined in GDPR, while others are broader, including questions about

⁶ RSA / The Royal Society for the Encouragement of Arts, Manufactures and Commerce Royal Society of Arts - Renate Samson, Kayshani Gibbon & Anna Scott (2019) *About data about us*. <https://www.thersa.org/discover/publications-and-articles/reports/data-about-us>, p19

understanding of concepts such as open data, and about what people think companies do, or can do, with personal data.

Doteveryone, a think tank focusing on responsible technology, undertook research in 2018 into understanding of and attitudes towards personal data and technology in general, in which a range of questions about people's knowledge of how their data are used were included⁷. While most people know that data are collected about their searches, the sites they have visited and their online purchasing history (68%, 68%, and 70% respectively), far fewer people know that data about their internet connection, or IP (38%), other sites visited (28%) and information other people share about them (17%) are also collected. These findings may reflect different levels of understanding of the questions posed, with 'internet connection' potentially less clear than 'searches', for example. These figures are comparable with the finding from a 2016 Ipsos Global Trends survey that 83% of UK respondents were unsure what data companies held about them⁸. Doteveryone also found that 6% of people believe that the internet has made their own lives worse, and 20% believe that it has had an overall negative impact on society. This suggests that a large fraction of their respondents simultaneously recognise that data about their online behaviour are being collected, do not necessarily recognise the extent to which this is the case, yet believe that the overall effect of these and other online activities on them individually and on society as a whole is positive. These findings are based on a sample of 2,538 people, of which responses from 500 people were collected via phone.

Digital technology innovation centre Digital Catapult found that while 96% of respondents to their survey claim to understand the term 'personal data', just 64% picked the correct definition from a series of options⁹. 65% reported that they were unsure about whether their personal data is being shared without their consent. These figures can be compared with a more recent survey reported by Kennedy and others, where rates of accurate definition of the term personal data were largely higher, perhaps reflecting a growth in understanding over time¹⁰. Kennedy and colleagues found that rates of knowledge about other data-related issues, such as open data, were much lower, at less than 50%. Similarly, in this survey, most people were knowledgeable about the basic tenets of GDPR, but more specific details about GDPR were less well understood.

Software company PEGA undertook a survey about public attitudes to GDPR before it came into effect¹¹. Most of the conclusions are at a Europe-wide level, with UK level details unavailable. PEGA found that across Europe, 82% of people said they were likely to ask to see, limit or erase their personal data once GDPR was in effect. One UK-specific statistic that they cite is that 74% of people in the UK said they were likely to want to see all the personal data a company had on file about them if they had the opportunity, the lowest of the seven countries surveyed. This suggests some understanding of the data practices which GDPR seeks to govern, alongside widespread intention to

⁷ This research was reported over two publications: Doteveryone (2018) *People, power, and technology: the 2018 digital understanding report*.

http://understanding.doteveryone.org.uk/files/Doteveryone_PeoplePowerTechDigitalUnderstanding2018.pdf and Doteveryone (2018) *People, power and technology: the 2018 digital attitudes report*. <https://www.doteveryone.org.uk/report/digital-attitudes/>

⁸ Cited in The British Academy & The Royal Society - Franck Fourniol & Fiona McLaughlin (2017) *Data governance: public engagement review*. <https://www.thebritishacademy.ac.uk/sites/default/files/Data-Governance-public-engagement-review.pdf>

⁹ Digital Catapult (2015) *Trust in personal data: A UK review*. Digital Catapult

¹⁰ Kennedy, H, Hartman, T, Steedman, R & Jones, R (2020b) *UK public unhappy with the ways their data is managed*. <https://livingwithdata.org/project/wp-content/uploads/2020/03/Views-on-Data-Management-Full-Report.pdf>

¹¹ PEGA (2019) *GDPR: Show me the data survey reveals EU consumers poised to act on legislation*. <https://www.pega.com/system/files/resources/2019-07/GDPR-Show-Me-The-Data-eBook.pdf>

act on the rights that are enshrined within it. These figures are surprisingly high, and should be compared with surveys undertaken after the introduction of GDPR, in order to explore whether people have acted in ways they indicated that they would.

In one such survey, in sharp contrast to PEGA findings, the Information Commissioner's Office (ICO) measured awareness of data rights in a survey, and found it to be relatively low¹². People were asked which rights they think they have under law with regard to personal data held about them by companies and organisations. The best-known right is the right to access one's own personal data, at 59%, while 43% are aware of the right to be forgotten, and 31% are aware of the right to not be the subject of automated decision-making.

A Eurobarometer survey covering GDPR (amongst other areas) found that 47% of people in the UK knew what GDPR was, a relatively high figure compared with other countries in Europe, but consistent with ICO findings¹³. It also found that 25% of people in the UK had exercised the right to access their data, and another 24% knew about this right but had not exercised it – together, these figures are slightly lower than the ICO estimate of 59%. Awareness of other rights was found to be similar to levels reported in the ICO report, but, in each case, only around a third of the people who reported awareness of these rights claimed to have actually exercised them: 24% claiming to have exercised the right to correct data about them, 20% to have used the right to be forgotten, and 15% to have had a say about when decisions are automated. These estimates are all relatively high compared with other countries in Europe, but may reflect social desirability effects – that is, respondents may have felt they should be aware of these rights, and should have acted on them, whether or not this was in fact the case. 74% of UK respondents reported that they had tried to change their privacy settings from the defaults on online social networks, the highest of all countries, which suggests significant knowledge of the ways in which personal data shared publically may be accessed and used. This is based on face-to-face research in each of the EU member states in 2019, with a total of 1,021 responses in the UK.

In general, the findings from quantitative studies suggest that people's knowledge about what happens to their personal data is mixed. UK residents' knowledge of the data practices enshrined in GDPR is above average compared with other EU countries, and a sizeable minority claim to be not just knowledgeable, but also active in exercising their rights. Given that GDPR is complex, new and how it will be implemented remains unclear, the high numbers of respondents who claim to have exercised their rights are somewhat implausible¹⁴. This is especially the case in relation to having a say about automated decision-making, given how opaque and black-boxed such processes are¹⁵.

Findings from qualitative research also paint a mixed picture. In 2015, Eslami and others explored 40 Facebook users' understanding and perception of Facebook's algorithms use of their data to target content to them¹⁶. They concluded that their participants had limited awareness of data-driven,

¹² ICO / Information Commissioner's Office (2019) *Information Commissioner's annual report and financial statements 2018-19*. <https://ico.org.uk/media/about-the-ico/documents/2615262/annual-report-201819.pdf>

¹³ The European Commission (2019) *Special Eurobarometer 487a. Summary - The General Data Protection Regulation* <https://ec.europa.eu/commfrontoffice/publicopinionmobile/index.cfm/Survey/getSurveyDetail/surveyKy/2222>

¹⁴ Hinz, A & Brand, J (nd) *Data policies: regulatory approaches for data-driven platforms in the UK and EU*. <https://datajustice.files.wordpress.com/2020/01/data-policies-research-report-revised.pdf>

¹⁵ Dencik, L, Hintz, A, Redden, J & Warne, H (2018) *Data Scores as Governance: Investigating uses of citizen scoring in public services*, Cardiff University. <https://datajustice.files.wordpress.com/2018/12/data-scores-as-governance-project-report2.pdf>

¹⁶ Eslami, M, Rickman, A, Vaccaro, K, Aleyasen, A, Vuong, A, Karahalios, K, Hamilton, K & Sandvig, C (2015) 'I always assumed that I wasn't really that close to [her]: reasoning about invisible algorithms in news feeds', *CHI '15: Proceedings of the 2015 CHI Conference on Human Factors in Computing Systems*, April 2015. <https://doi.org/10.1145/2702123.2702556>

algorithmic curation. This finding is contradicted by later qualitative studies, such as Bucher, who found understanding of uses of data by social media platforms, evidenced for example in participants' behaviours on these platforms¹⁷.

Other research has found that knowledge and understanding vary across demographic groups. In a focus group study (n = 68) of perceptions of the data practices of the UK public service broadcaster the BBC, Kennedy and others found greater awareness and understanding amongst younger participants (aged 16-34), and less understanding amongst older participants (over 65) and participants with mild learning disabilities¹⁸. Woodruff and others carried out a series of workshops and interviews with 44 participants from marginal communities in the US¹⁹. They found that most participants were not aware of data-driven, algorithmic intervention and what authors describe as 'algorithmic unfairness' prior to their participation on the project, but that on learning about it, some participants believed they had experienced it. In contrast again, in a broader study of how low socio-economic status youth in the US view online data sharing, Marwick and colleagues found that their 28 participants were aware of risks, perhaps because many had experienced offline surveillance²⁰. Consequently, many avoided social media, self-censored or obfuscated. In a focus group study in Australia involving 48 participants whose demographic differences are not mentioned, Lupton and Michael found awareness of what they describe as 'obvious dataveillance' by commercial and some government actors, but less awareness of national security, police or illegal data uses²¹.

In the grey literature, Citizens Advice commissioned Illuminas to address what they identified as a gap in knowledge about how consumers experience data-driven, digital markets, through a mix of qualitative methods. They also found differences in understanding of data practices across consumers and other stakeholders, which they argue is due to differing interpretations of concepts such as transparency and trust (a topic we discuss in more detail below). The RSA, Open Data Institute (ODI) and Luminate carried out a collaborative project to address their concern that the public is often written off as 'complacent or ignorant' when it comes to data practices²². They found that this was not the case in the focus group discussions that they ran.

Awareness of data practices was also identified in a collaborative project commissioned by the Economic and Social Research Council (ESRC), which took the form of a public consultation dialogue about private sector data practices across different stages: acquisition, storage and preservation,

¹⁷ Bucher, T (2017) 'The algorithmic imaginary: exploring the ordinary affects of Facebook algorithms', *Information, Communication & Society*, 20: 30-44. <https://doi.org/10.1080/1369118X.2016.1154086>. See also: Rader, E & Gray, R (2015) 'Understanding User Beliefs About Algorithmic Curation in the Facebook News Feed', In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*: 173–182. <https://doi.org/10.1145/2702123.2702174> for similar findings to Bucher, based on a survey of 464 Amazon Mechanical Turk workers who complete virtual tasks which require human intelligence.

¹⁸ Kennedy, H, Steedman, R & Jones, R (2020a) 'Approaching public perceptions of datafication through the lens of inequality: a case study in public service media', *Information, Communication and Society*. <https://doi.org/10.1080/1369118X.2020.1736122>

¹⁹ Woodruff, A & Fox, S E, Rousso-Schindler, S, & Warshaw, J (2018) 'A qualitative exploration of perceptions of algorithmic fairness', CHI '18: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, April 2018 Paper No.: 656. <https://doi.org/10.1145/3173574.3174230>

²⁰ Marwick, A, Fontaine, C & Boyd, D (2017) "'Nobody sees it, nobody gets mad": social media, privacy and personal responsibility among low-SES youth', *Social Media + Society*, 3(2). <https://doi.org/10.1177/2056305117710455>

²¹ Lupton, D & Michael, M (2017) "'Depends on who's got the data": public understandings of personal digital dataveillance', *Surveillance and Society*, 15:254-268. <https://doi.org/10.24908/ss.v15i2.6332>

²² RSA / The Royal Society for the Encouragement of Arts, Manufactures and Commerce Royal Society of Arts - Renate Samson, Kayshani Gibbon & Anna Scott (2019) *About data about us*. <https://www.thersa.org/discover/publications-and-articles/reports/data-about-us>, p. 19

access and ownership. The authors found general awareness of private sector data collection, but less understanding of the extent of data collected and its potential uses, echoing some of the findings of survey research²³. Oman's research into how cultural sector workers feel about having their personal data gathered to measure inequality was based on focus groups and interviews with around 200 people²⁴. She found that knowledge and understanding of data usage was diverse, even amongst people with similar jobs and backgrounds in the same institution, and many participants did not know the purpose of comparatively mundane data processes such as equality monitoring.

Findings from qualitative and quantitative studies suggest that some people understand some data practices. Quantitative research often focuses on attitudes rather than knowledge and understanding, perhaps because it is methodologically challenging to identify knowledge. For example, when participants are asked if they are aware of certain practices, social desirability bias might lead respondents to feel they should answer 'yes' so that they do not appear ignorant. Factual questions with right or wrong answers can appear intimidating, which may lead to dropout in online approaches or a loss of rapport in face-to-face surveys, and so they are sometimes avoided. Qualitative research has not been primarily concerned with knowledge and understanding either. Where this has been addressed, some studies find that participants have some knowledge of data practices, and others do not. The study we quoted at the beginning of this section, from the RSA, Samson and colleagues, disproves the assumption that people do not understand data-related matters, based on two focus groups and a workshop²⁵. Other studies, which we discuss below, highlight that focus group dialogue can enhance knowledge and understanding for participants²⁶. On the whole, though, it is difficult to be conclusive about degrees of knowledge and understanding about data practices amongst the public.



²³ Hopkins Van Mil: Creating Connections Ltd (2015) *Big data: public views on the use of private sector data for social research - a findings report for the Economic and Social Research Council*. <https://esrc.ukri.org/files/public-engagement/public-dialogues/public-dialogues-on-the-re-use-of-private-sector-data-for-social-research-report/>

²⁴ See Oman, S (2019a) *Improving data practices to monitor inequality and introduce social mobility measures: a working paper*, The University of Sheffield. https://www.sheffield.ac.uk/polopoly_fs/1.867756!/file/MetricsWorkingPaper.pdf and Oman, S (2019b) *Measuring social mobility in the creative and cultural industries – the importance of working in partnership to improve data practices and address inequality*, The University of Sheffield. https://www.sheffield.ac.uk/polopoly_fs/1.867754!/file/MetricsPolicyBriefing.pdf

²⁵ RSA / The Royal Society for the Encouragement of Arts, Manufactures and Commerce Royal Society of Arts - Renate Samson, Kayshani Gibbon & Anna Scott (2019) *About data about us*. <https://www.thersa.org/discover/publications-and-articles/reports/data-about-us>, p. 19

²⁶ For example Citizens Advice - Illuminas (2016) *Consumer expectations for personal data management in the digital world*. <https://www.citizensadvice.org.uk/Global/CitizensAdvice/Consumer%20publications/Personal%20data%20consumer%20expectations%20research.docx.pdf>

2.2. People are concerned about data practices

There is extensive evidence that people are concerned about data practices. The Information Commissioner's annual report of 2019 claimed that in that year, 'a record number of people' were 'raising data protection concerns'²⁷. Surveys and polls undertaken in the past five years also paint a picture of public concern about data practices. An earlier report for Big Brother Watch claimed that 79% of people were concerned about their online privacy, with 29% of people very concerned²⁸. In this study, after being asked how concerned they were about their privacy, respondents were then asked whether they thought consumer experiences were being harmed or enhanced by 'big companies gathering large amounts of their personal data for internal use', wording that may have influenced responses. Less than half of respondents (46%) chose harmed, 19% opted for enhanced, and the remainder opted for neither or said they did not know. Later in the questionnaire, respondents were asked whether regulators should have been more strict in a 2012 investigation into Google's privacy policy and 68% responded that they should have been. 72% subsequently responded that national regulators should do more to force Google to comply with privacy rules. This report suggests that people are concerned that their online data is not being used appropriately, and Google, in particular, should be more tightly regulated. The report was based on a ComRes online panel of around 1,000 people.

The Direct Marketing Association (DMA) found similar results, but framed them very differently²⁹. Their key question was 'On a scale from 1 to 10 where 1 is "not at all concerned" and 10 is "very concerned", how do you rate your levels of concern about the issue of online privacy these days?'. 75% of respondents gave a score from 7 to 10, which is comparable to the 79% of people in the Big Brother Watch study who were concerned about their online privacy. This is also consistent with findings from elsewhere in the world, such as polls carried out by the Pew Research Center's Internet and American Life Project³⁰. These are not discussed in detail here because of our focus on grey literature from the UK, but it is worth noting that a 2019 study found that 79% of people in the US were concerned about how companies use their data³¹. The DMA frames its finding in its executive summary with the headline 'Overall privacy concerns decline, as happiness with the level of data shared rises', because 75% of respondents scoring 7/10 is a decrease from 84% in 2012. This report similarly interprets the increase of people who agree with the statement 'The exchange of personal information is essential to the smooth running of modern society', from 38% in 2012 to 51% in 2017, as 'Rising awareness and understanding of the role of data exchange in modern societies.' This report finds smaller fractions of younger than older people reporting concern about their online privacy, based on an online survey of 2,017 people.

The Oxford Internet Institute (OII) have run a tracker survey focusing on internet use, behaviour and attitudes since 2003. In 2015 they concluded that 'Young people care more about their online privacy than older users' on the basis of the fractions of young people who had changed their privacy settings

²⁷ ICO / Information Commissioner's Office (2019) *Information Commissioner's annual report and financial statements 2018-19*. <https://ico.org.uk/media/about-the-ico/documents/2615262/annual-report-201819.pdf> p9

²⁸ Big Brother Watch & ComRes (2015) *UK public research – online privacy*. <https://www.bigbrotherwatch.org.uk/wp-content/uploads/2015/03/Big-Brother-Watch-Polling-Results.pdf>

²⁹ The Direct Marketing Association (2018) *Data privacy: what the consumer really thinks*.

https://dma.org.uk/uploads/misc/5a857c4fdf846-data-privacy---what-the-consumer-really-thinks-final_5a857c4fdf799.pdf

³⁰ The Pew Research Center's Internet & American Life project <https://www.pewresearch.org/internet/>

³¹ Pew Research Center - Brooke Auxier, Lee Rainie, Monica Anderson, Andrew Perrin, Madhu Kumar & Erica Turner *Americans and Privacy: Concerned, Confused and Feeling Lack of Control Over Their Personal Information*.

<https://www.pewresearch.org/internet/2019/11/15/americans-and-privacy-concerned-confused-and-feeling-lack-of-control-over-their-personal-information/>

on social media: around 80% of people aged between 16-24, compared with around 30% of people aged over 65³². Thus young respondents did not say that they care or are concerned about their online privacy; rather, this is assumed from their reported online behaviours. The 2019 wave of data collection asked why some people do not use the internet and found that 10% reported privacy worries. While this is far below the 69% of people who said they weren't interested, this was a rise from just 1% in 2013. These estimates are based on face-to-face interviews with 2,000 people in each wave³³.

The Office of Communications (Ofcom) measured people's attitudes to online data usage in two key ways in a report on a wide range of issues relating to media use³⁴. Firstly, people were asked how confident they were in knowing how to manage who had access to their personal data online. 72% of people stated that they were either very or fairly confident, with larger fractions of younger people reporting confidence, consistent with reports from the OII. However, this question implies that citizens *can* manage who has access to their online personal data, which is not always the case, as platforms and other online services often share data with third parties in ways which are not transparent. Thus answers to the question may refer to knowing how to change privacy settings rather than other aspects of online data management. Respondents were also asked about specific behaviours, for example whether they consider privacy or data security implications when they post photos that they have taken online. 46% said they always do so, and 31% said they sometimes do, with people in the youngest age group the least likely to do so. A much larger fraction of women than men say they always do so: 51% compared with 41%. The patterns are similar for questions about tagging other people in photos uploaded, commenting on other people's posts, and checking into locations. This report is based on a regular tracker survey, with 1,846 people having been interviewed face-to-face in this wave.

In addition to the grey literature, academic research also finds widespread concern about or dissatisfaction with data practices. Bergstrom reports on aspects of a survey carried out in Sweden relating to how privacy concerns are perceived in different online contexts. She concludes that privacy concerns differ across distinct populations, with people who have less formal education being far more concerned than those with more³⁵. Rendina and Mustanski also found different degrees of concern regarding data sharing with a nationwide sample of men who have sex with men³⁶. Concerns were dependent on the kinds of data and platforms. In a qualitative study, Guberek and others explored risk perceptions in relation to technology use amongst undocumented migrants in the US, and found concerns about identity theft, privacy and online harassment, some of which related to uses of their personal data³⁷. Williams and colleagues surveyed 564 UK Twitter users to understand how they felt

³² OII / Oxford Internet Institute (2015) *Internet use, behaviour and attitudes in Great Britain 2003-2015*.

<http://oxis.oii.ox.ac.uk/wp-content/uploads/2015/01/OxIS-Brochure.pdf>

³³ OII / Oxford Internet Institute (2019) *Perceived threats to privacy online: the Internet in Britain*.

<https://oxis.oii.ox.ac.uk/wp-content/uploads/sites/43/2019/09/OxIS-report-2019-final-digital-PDFA.pdf>

³⁴ Ofcom (2019) *Adult's media use and attitudes report: 2019*.

https://www.ofcom.org.uk/data/assets/pdf_file/0021/149124/adults-media-use-and-attitudes-report.pdf

³⁵ Bergstrom, A (2015) 'Online privacy concerns: a broad approach to understanding the concerns of different groups for different uses', *Computers In Human Behavior*, 53:419-426. <https://doi.org/10.1016/j.chb.2015.07.025>

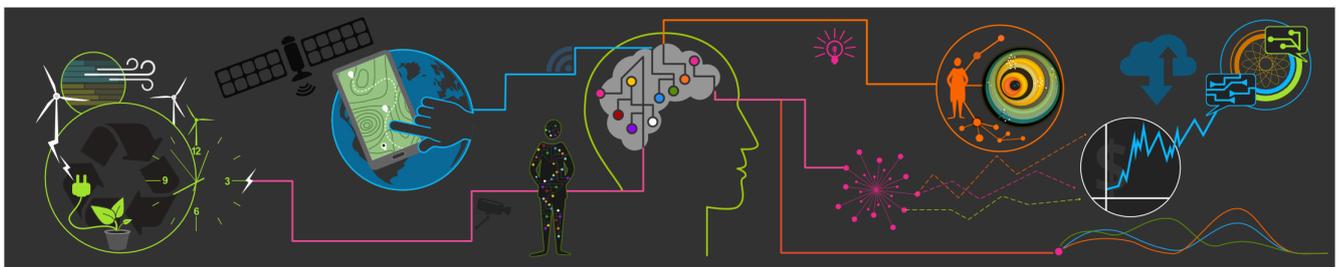
³⁶ Rendina, H J & Mustanski, B (2018) 'Privacy, trust, and data sharing in web-based and mobile research: participant perspectives in a large nationwide sample of men who have sex with men in the united states', *Journal of Medical Internet Research*, 20(7). <https://doi.org/10.2196/jmir.9019>

³⁷ Guberek, T, McDonald, A, Simioni, S, Mhaidli, A, Toyama, K, Schaub, F (2018) 'Keeping a Low Profile?: Technology, Risk and Privacy among Undocumented Immigrants', *CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, April 2018 Paper No.: 114. <https://doi.org/10.1145/3173574.3173688>

about their tweets being used for different kinds of research³⁸. They found that degree of concern depended on context: the authors found that 16% of respondents were 'quite concerned or very concerned' about university research, 49% about government research and 51% about commercial research settings³⁹.

In the examples above, authors set out to study concerns; elsewhere, concern surfaced as a theme in research that did not set out to study it. For example, Lupton and Michael aimed to 'generate discussion about personal digital dataveillance' and to evaluate the usefulness of cultural probes (that is, artefacts like postcards or diaries used to understand cultural contexts of particular practices) as a mechanism for doing so⁴⁰. In the process, they found that focus group participants both saw the value of data mining but were suspicious of those that do it. In another example, Pink and others undertook a programme of research that focused on values in relation to digital technologies, data and design futures⁴¹. As part of this, they identified anxieties relating to data in participants' everyday lives, including at work.

In sum, the evidence presented in this section, from quantitative and qualitative research, and from grey and academic literature, suggests that there is public concern about data practices. Some of the evidence indicates that concern is widespread. This points towards the need to listen to people's views and to address their concerns. But there is also a need to think about how people deal with their concerns and whether they feel they have any agency to do so. We address these issues in the next section.



³⁸ Williams, M L, Burnap, P & Sloan, L (2017) 'Towards an ethical framework for publishing twitter data in social research: taking into account users' views, online context and algorithmic estimation', *Sociology*, 51(11):1149-1168. <https://doi.org/10.1177/0038038517708140>

³⁹ See also Demos - Harry Evans, Steve Ginnis & Jamie Bartlett (2015) *#Socialethics a guide to embedding ethics in social media research*. <https://www.ipsos.com/ipsos-mori/en-uk/ipsos-mori-and-demoscasm-call-better-ethical-standards-social-media-research>

⁴⁰ Lupton, D & Michael, M (2017) "'Depends on who's got the data": public understandings of personal digital dataveillance', *Surveillance and Society*, 15:254-268. <https://doi.org/10.24908/ss.v15i2.6332>

⁴¹ Pink, S, Lanzeni, D & Horst, H (2018) 'Data anxieties: finding trust in everyday digital mess', *Big Data & Society*, 5(1). <https://doi.org/10.1177/2053951718756685>

2.3. People are not *only* concerned: they have some agency, they negotiate or resist data, and sometimes have contradictory views

The *About Data About Us* report cited above states that ‘we must engage with and listen to people about how they feel, and stop writing off “the public” as being complacent or ignorant about data protection issues, as they often are by people in power and in the media’⁴². Doing so often requires qualitative methods, and it often calls into question a simple picture of concern and dissatisfaction. Research in this vein often highlights the strategies that people deploy in order to negotiate, embed or resist data practices in their everyday lives. This research draws attention to people’s agency in relation to data practices, and so serves to complicate the picture of public perceptions of data practices as primarily negative. Such research often explores what people *do* as a result of, or way of dealing with, their concerns, rather than focusing on understanding and perceptions – some of the studies mentioned in the previous section serve as examples. For example, Pink and others found that the design professionals in their study build specific, individual routines for managing their data and this helps them to deal with their concerns⁴³. Guberek and colleagues explored the strategies that undocumented migrants deploy to navigate the relationship between their technology use and risk perceptions, finding fewer negotiation strategies than Pink and her collaborators, thus pointing to inequalities that we discuss further below⁴⁴.

Research into people’s agency in relation to data practices often focuses on self-tracking data. Self-tracking, or the practice of generating data about everyday activities like eating, exercise or sleeping for purposes of self-improvement, puts data in the hands of users, as well as the corporations which produce self-tracking devices and the third parties with which this data is shared. This enables the kinds of negotiated, agentic responses to data that researchers identify. Researchers have found that self-tracking data is made meaningful in the context of the everyday lives of self-trackers, according to their aims and needs, as Ruckenstein and Pantzar argue⁴⁵. Similarly, Lomborg and colleagues argue that self-trackers find meaning in their data flows in specific contexts of practice⁴⁶. Lupton also argues that the self-trackers she interviewed in Australia incorporate self-tracking data into their lives and relationships and that in so doing, they view data in multiple, diverse ways⁴⁷. Pink and colleagues draw similar conclusions from a qualitative study of self-tracking cycling commuters, arguing that their participants produce, experience and engage with their data in contingent and specific ways⁴⁸.

⁴² RSA / The Royal Society for the Encouragement of Arts, Manufactures and Commerce Royal Society of Arts - Renate Samson, Kayshani Gibbon & Anna Scott (2019) *About data about us*. <https://www.thersa.org/discover/publications-and-articles/reports/data-about-us>, p19.

⁴³ Pink, S, Lanzeni, D & Horst, H (2018) ‘Data anxieties: finding trust in everyday digital mess’, *Big Data & Society*, 5(1). <https://doi.org/10.1177/2053951718756685>

⁴⁴ Guberek, T, McDonald, A, Simioni, S, Mhaidli, A, Toyama, K, Schaub, F (2018) ‘Keeping a Low Profile?: Technology, Risk and Privacy among Undocumented Immigrants’, *CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, April 2018 Paper No.: 114. <https://doi.org/10.1145/3173574.3173688>

⁴⁵ Ruckenstein, M & Pantzar, M (2015) ‘Datafied Life: techno-anthropology as a site for exploration and experimentation’, *Techné: Research in Philosophy and Technology*, 19(2):191-210. <https://doi.org/10.13140/RG.2.1.2553.7762>

⁴⁶ Lomborg, S, Thylstrup, N & Schwartz, J (2018) ‘The temporal flows of self-tracking: checking in, moving on, staying hooked’, *New Media & Society*, 20(12): 4590-4607 <https://doi.org/10.1177/1461444818778542>

⁴⁷ Lupton, D (2019) ‘Data mattering and self-tracking: what can personal data do?’, *Continuum*, 34(1):1-13. <http://dx.doi.org/10.1080/10304312.2019.1691149>

⁴⁸ Pink, S & Fors, V (2017) ‘Being in a mediated world: self-tracking and the mind–body–environment’, *Cultural Geographies*, 24(3)375-388. <https://doi.org/10.1177/1474474016684127>. See the following for more examples in this vein: Ajana, B (2017) *Self-tracking: Empirical and philosophical investigations*, Springer International Publishing. <https://doi.org/10.1007/978-3-319-65379-2>;

Lupton, D (2017) ‘“It just gives me a bit of peace of mind”: Australian women’s use of digital media for pregnancy and early motherhood’, *Societies*, 7(3)25. <https://doi.org/10.3390/soc7030025>;

Similar arguments can be found in research on other types of data. Writing about direct-to-consumer genetic testing data, Ruckenstein uses the idea of the 'liveliness of data' to argue that users keep data alive as they talk about it and embed it in their everyday practices⁴⁹. Focusing on data gathered by household Intelligent Personal Assistants (IPAs), Pridmore and Mols similarly argue that users negotiate household IPA use in the context of their everyday lives, making choices to preserve privacy or determine acceptable uses of their data⁵⁰. Likewise, Lomborg and Kapsch argue that people decode the work that algorithms do, valorising and responding to algorithmic activity in daily media use through interpretative work which is interwoven with reflections on other issues⁵¹. All of these studies suggest that people have some agency over these kinds of data and are able to embed them, or make them meaningful, in their daily lives. However, they rarely acknowledge that social inequalities affect the extent to which people can engage in such activities. We discuss inequalities in greater detail below in Section 2.9.

In an early study of perceptions of data mining on social media platforms based on focus groups with 65 participants, Kennedy and others found that participants evaluated data practices in a case-by-case way⁵². They noted that participants in their research weighed up who was gathering data, whose data was being gathered, what data was being gathered, for what purpose and with what effects, in their considerations of each case. In this study, a concern for fairness emerged as a common trope among users. These findings provided an early indication that people's views of data practices are considered; people differentiate amongst data practices when they assess them. We say more about this point in sections 2.5. and 2.9. below.

Some studies draw attention to people's contradictory perspectives on data practices. Lupton and Michael's study, mentioned above, is one example: they found that participants saw the value of data mining but were simultaneously suspicious of those that do it⁵³. A recent study by Ruckenstein and Granroth, which explored perceptions of targeted advertising and the data practices that underpin it, concluded that consumers want contradictory things: they oppose intrusive advertising yet expect relevant real-time analysis⁵⁴. What authors describe as the 'corporate surveillance' of data-driven targeted advertising was seen as disturbing, yet this perception co-existed with the pleasure of feeling seen and recognised through targeting amongst participants in their study.

Lupton, D (2019) 'Data mattering and self-tracking: what can personal data do?', *Continuum*, 34(1):1-13.

<http://dx.doi.org/10.1080/10304312.2019.16911492019>;

Pink, S & Fors, V (2017) 'Being in a mediated world: self-tracking and the mind-body-environment', *Cultural Geographies*, 24(3)375-388. <https://doi.org/10.1177/1474474016684127>;

Weiner, K, Will, C, Henwood, F, Williams, R (2020) 'Everyday curation? Attending to data, records and record keeping in the practices of self-monitoring', *Big Data & Society*, 7(1). <https://doi.org/10.1177/2053951720918275>

⁴⁹ Ruckenstein, M S (2017) 'Keeping data alive: talking DTC genetic testing', *Information, Communication and Society*, 20(7):1024-1039. <https://doi.org/10.1080/1369118X.2016.1203975>

⁵⁰ Pridmore, J & Mols, A (2020) 'Personal choices and situated data: Privacy negotiations and the acceptance of household Intelligent Personal Assistants', *Big Data & Society*, 7(1). <https://doi.org/10.1177/2053951719891748>

⁵¹ Lomborg, S & Kapsch, P (2019) 'Decoding algorithms', *Media, Culture & Society*.

<https://doi.org/10.1177/0163443719855301>

⁵² Kennedy, H, Elgesem, D, & Miguel, C (2015) 'On fairness: user perspectives on social media data mining', *Convergence*, 23(3): 270-288. <https://doi.org/10.1177/1354856515592507>

⁵³ Lupton, D & Michael, M (2017) "'Depends on who's got the data": public understandings of personal digital dataveillance', *Surveillance and Society*, 15:254-268. <https://doi.org/10.24908/ss.v15i2.6332>

⁵⁴ Ruckenstein, M, & Granroth, J (2019) 'Algorithms, advertising and the intimacy of surveillance', *Journal of Cultural Economy*, 13(1):12-24. <https://doi.org/10.1080/17530350.2019.1574866>

A small number of researchers explore how people resist data practices. One example is Gangadharan who draws on findings from Our Data Bodies, a project which explored how data-driven systems impact the lives of people in marginalised neighbourhoods in the US⁵⁵. Gangadharan argues that her participants' self-exclusion from data-driven technologies is a form of technological refusal, an act of agency which acknowledges the negative consequences of data practices for marginalised communities⁵⁶. She argues that 'informed refusal' is political and affirmative, a means for members of historically marginalised groups to collectively self-determine in a technologically mediated world.

Apart from Gangadharan's work, few publications acknowledge how agency in relation to data practices is differentiated across populations. Another exception is Guberek and colleagues' study, already mentioned, which looked at how Latinx undocumented migrants in the US managed self-expression, group privacy and self-censorship related to their immigration status, and how this related to trust in service providers⁵⁷. We note above that Guberek and their colleagues found much less strategic negotiation of and agency in relation to data practices than Pink and her colleagues, perhaps because of the very different socio-economic position of the participants in the two studies⁵⁸. Guberek et al found what they describe as resignation towards government surveillance and strong trust in service providers – we come back to the concept of resignation below.

Pybus and others ran participatory workshops with 20 young coders (aged 14 to 18) to explore what data-making possibilities exist when users have access to the data that they produce on their mobile phones that they are usually unable to access⁵⁹. The authors argue that gaining access to one's own data augments the agency of both the individual and the collective. This is another example which acknowledges how populations are differently situated with regard to their agency in relation to data practices, as the young coders who were the participants in this study already had coding skills and some knowledge of data practices, and as such were able to experiment with data in ways that were meaningful to them.

Most of the research discussed in this section is academic, but some grey literature is also concerned with the everyday and with agency. *About Data About Us*, already mentioned above, focuses on everyday experiences of data. The authors carried out two focus groups and one participatory workshop in London to understand public feelings towards data and the reasoning behind these feelings. The authors conclude that 'most people want to make a choice based on how they feel at a moment in time, and be able to change their minds when they feel differently'. The authors argue that this nuance is often ignored or misinterpreted, particularly in quantitative research. They state that participants in their focus groups and workshops saw the benefits of online data practices, but were also worried about 'how much they understand, how well they are educated in using connected

⁵⁵ The Our Data Bodies project <https://www.odbproject.org/>

⁵⁶ Gangadharan, S P (2021) 'Digital exclusion: a politics of refusal', in H Landemore, R Reich & L Bernholz (eds) *Digital Technology and Democratic Theory*. University of Chicago Press. <http://eprints.lse.ac.uk/103076/>

⁵⁷ Guberek, T, McDonald, A, Simioni, S, Mhaidli, A, Toyama, K, Schaub, F (2018) 'Keeping a Low Profile?: Technology, Risk and Privacy among Undocumented Immigrants', *CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, April 2018 Paper No.: 114. <https://doi.org/10.1145/3173574.3173688>

⁵⁸ Pink, S, Lanzeni, D & Horst, H (2018) 'Data anxieties: finding trust in everyday digital mess', *Big Data & Society*, 5(1). <https://doi.org/10.1177/2053951718756685>

⁵⁹ Pybus, J, Coté, M & Blanke, T (2015) 'Hacking the social life of big data: a data literacy framework', *Big Data & Society*, 2(2). <https://doi.org/10.1177/2053951715616649>

technologies, how safe they are, and their lack of control over how data about them is used' and about organisations making decisions which might not benefit them or society⁶⁰.

In some reports, agency in relation to data is seen as a right, part of a 'data rights framework', for example in the Royal Statistical Society's *The Data Manifesto*⁶¹. There is widespread discussion about data rights in broader literature which is not empirical, and some of the empirical research we review here is also focused on rights – the right to redress, for example⁶². Such literature is concerned with how much control people have over their own data, and efforts to assert control⁶³ or to balance the risks and benefits of particular practices⁶⁴. The *About Data About Us* report concludes that people 'generally feel positive about the benefits brought by the internet and being more connected, but want greater honesty and transparency, agency and control, rights and responsibility, context and fairness, and compliance and enforceability over how data about them is used'⁶⁵. This quote captures some of the contradictions and the desire for agency that we have discussed here, as well as other issues which we discuss below, including the importance of context and a desire for fairness.



⁶⁰ RSA / The Royal Society for the Encouragement of Arts, Manufactures and Commerce Royal Society of Arts - Renate Samson, Kayshani Gibbon & Anna Scott (2019) *About data about us*. <https://www.thersa.org/discover/publications-and-articles/reports/data-about-us>, p35.

⁶¹ RSS / Royal Statistical Society (2019) *The Data Manifesto* <https://www.rss.org.uk/Images/PDF/influencing-change/2019/Data%20Manifesto2019.pdf>

⁶² Doteveryone - Joe Massey, Jacob Ohrvik-Stott & Catherine Miller (2019) *Better redress: building accountability for the digital age: an evidence review from Doteveryone* <https://www.doteveryone.org.uk/wp-content/uploads/2019/12/Better-redress-evidence-review.pdf>

See also: Doteveryone – Catherine Miller (2019) *Engaging the public with responsible technology: four principles and three requirements*. <https://doteveryone.org.uk/download/3225/> and The European Commission (2019) *Special Eurobarometer 487a. Summary - The General Data Protection Regulation* <https://ec.europa.eu/commfrontoffice/publicopinionmobile/index.cfm/Survey/getSurveyDetail/surveyKy/2222>

⁶³ For example: Citizens Advice - Illuminas (2016) *Consumer expectations for personal data management in the digital world*. <https://www.citizensadvice.org.uk/Global/CitizensAdvice/Consumer%20publications/Personal%20data%20consumer%20expectations%20research.docx.pdf>

⁶⁴ Carnegie Trust - Scott, Kaela (2018) *Data for public benefit: balancing the risks and benefits of data sharing*. https://www.involve.org.uk/sites/default/files/field/attachemnt/Data%20for%20Public%20Benefit%20Report_0.pdf.

⁶⁵ RSA / The Royal Society for the Encouragement of Arts, Manufactures and Commerce Royal Society of Arts - Renate Samson, Kayshani Gibbon & Anna Scott (2019) *About data about us*. <https://www.thersa.org/discover/publications-and-articles/reports/data-about-us>, p3

2.4. Feelings play an important role in public understanding and perceptions of data practices

A number of qualitative studies draw attention to the important role that feelings play in shaping people's perceptions of data practices. One quantitative study, by Elhai and others, also foregrounds feeling. Elhai et al explored the factors that influence 'emotional distress' about data hacking and surveillance and found that experience of hacking influenced anxiety, which in turn influenced digital privacy protection behaviour⁶⁶.

On the whole, however, it is in qualitative research where the importance of feelings in perceptions of data practices can be seen. A lot of the literature on perceptions of self-tracking data highlights positive emotional engagements with it. For example, Sumartojo and colleagues depict the experiences of self-tracking cycling commuters as characterised by emotional responses to the data that these practices produce⁶⁷. Expanding the discussion to broader health and well-being data, Ruckenstein identifies the important role of emotions in perceptions of direct-to-consumer genetic testing data⁶⁸.

In studies of other kinds of data, authors also argue that feelings play an important role in responses to datafication. Focusing on social media data, Bucher's small-scale study of how Facebook users think about the platform's algorithms concludes that algorithmic processes and the data on which they draw are experienced emotionally by Facebook users⁶⁹. Colbjørnsen researched what people say about recommender algorithms on social media, finding strong positive and negative feelings about the quality and relevance of recommendations made and input and output mechanisms, and emotional expressions of identification with the algorithms doing the recommending (eg algorithms 'know me so well' or are like a 'smug older brother')^{70,71}. Ruckenstein and Granroth, referenced above, identify pleasurable experiences when living with algorithms, such as feeling understood by advertisers⁷². In a recently published study by Kennedy and others about perceptions of data gathering when signing in to access digital media services, mentioned above, the role of feelings in the formation of opinions about data practices is also highlighted⁷³. Here, the authors also note that particular participants – older participants and younger participants with mild learning disabilities – had strong, largely negative emotional responses to the data practices that were discussed, despite not fully understanding them.

⁶⁶ Elhai, J, Levine, J & Hall, B (2017) 'Anxiety about electronic data hacking: predictors and relations with digital privacy protection behavior', *Internet Research*, 27(3):631-649. <https://doi.org/10.1108/IntR-03-2016-0070>

⁶⁷ Sumartojo, S, Pink, S, Lupton, D & La Bond C H (2016) 'The affective intensities of datafied space', *Emotion, Space and Society*, 21:33-40. <https://doi.org/10.1016/j.emospa.2016.10.004>

⁶⁸ Ruckenstein, M S (2017) 'Keeping data alive: talking DTC genetic testing', *Information, Communication and Society*, 20(7):1024-1039. <https://doi.org/10.1080/1369118X.2016.1203975>

⁶⁹ Bucher, T (2017) 'The algorithmic imaginary: exploring the ordinary affects of Facebook algorithms', *Information, Communication & Society*, 20: 30-44. <https://doi.org/10.1080/1369118X.2016.1154086>

⁷⁰ Colbjørnsen, T (2018) 'My algorithm: user perceptions of algorithmic recommendations in cultural contexts', in Andrea L Guzman (ed) *Human-machine communication: rethinking communication, technology, and ourselves*. Peter Lang.

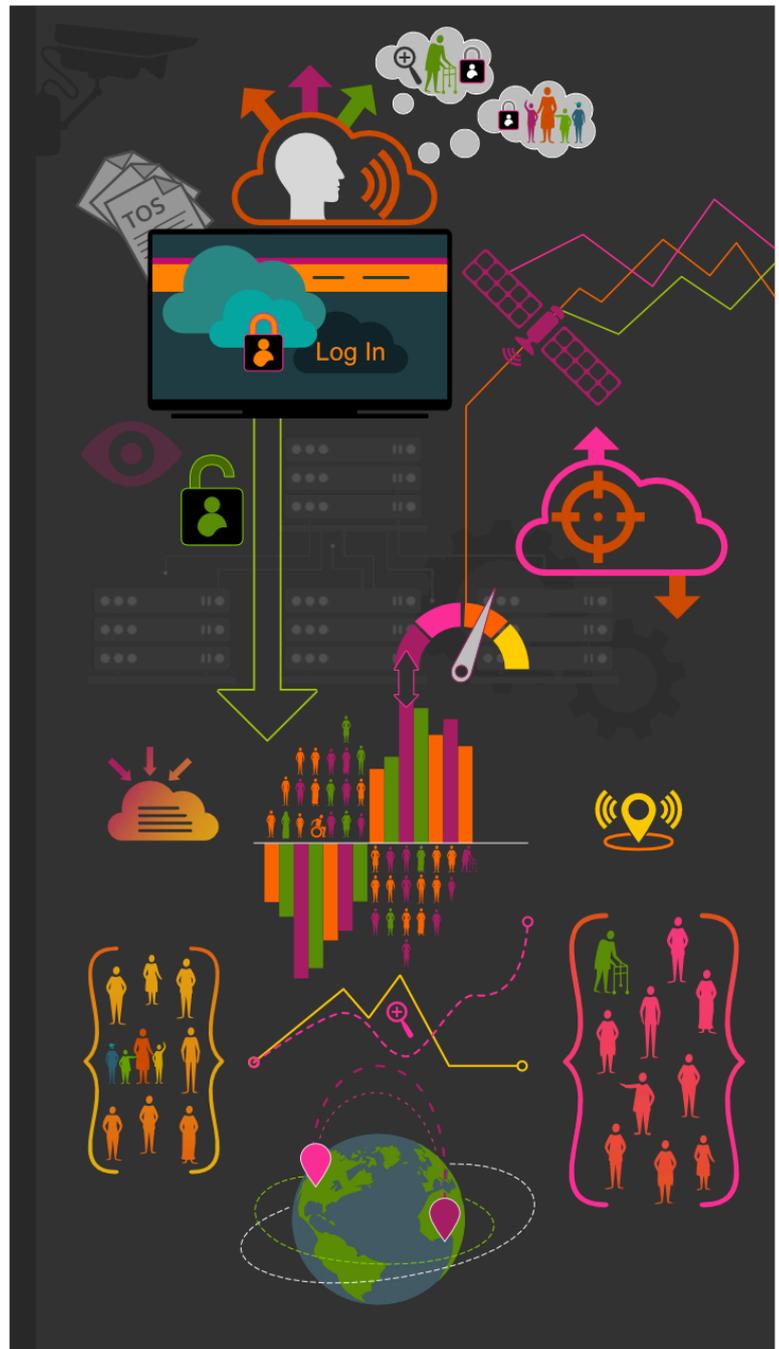
⁷¹ See also Bolin, Göran & Jonas Andersson Schwarz (2015) 'Heuristics of the algorithm. Big Data, user interpretation and institutional translation', *Big Data & Society*, 2(2): 1–12. <https://doi.org/10.1177/2053951715608406> on people's engagements with algorithms.

⁷² Ruckenstein, M, & Granroth, J (2019) 'Algorithms, advertising and the intimacy of surveillance', *Journal of Cultural Economy*, 13(1):12-24. <https://doi.org/10.1080/17530350.2019.1574866>

⁷³ Kennedy, H, Steedman, R & Jones, R (2020a) 'Approaching public perceptions of datafication through the lens of inequality: a case study in public service media', *Information, Communication and Society*. <https://doi.org/10.1080/1369118X.2020.1736122>

Oman's research into how cultural sector workers feel about inequality data gathering indicated that such processes provoke negative feelings. She found that when organisations are not clear about the purposes of their data practices or what happens to data that is collected, this has emotional effects⁷⁴. Despite the apparently mundane nature of equality monitoring data practices compared to more harmful data practices (as described by authors such as Gangandharan, mentioned above), they can still provoke emotional responses. This report concluded that the rationale for such data collection should be made clearer, to reduce its harmful effects and to improve the data that is collected.

These qualitative studies recognise the role feelings play in perceptions of data practices. Based on their research into how people engage with visual representations of data, Kennedy and Hill conclude that emotions play a role in how people make sense of data, and that data are 'as much felt as they are experienced cognitively and rationally'⁷⁵. Therefore emotions need to be seen as an important element in public understanding and perceptions of data practices. They need to be understood as informing and informed by reason and rational thinking, as Coleman has argued⁷⁶. Emotions play a role in the formation of attitudes and the expression of perceptions, in relation to data practices as with other phenomena. This is an important point for policy-makers, practitioners and others involved in initiatives which aim to improve knowledge and understanding of data practices with different populations.



⁷⁴ Oman, S (2019a) *Improving data practices to monitor inequality and introduce social mobility measures: a working paper*. The University of Sheffield. https://www.sheffield.ac.uk/polopoly_fs/1.867756!/file/MetricsWorkingPaper.pdf

⁷⁵ Kennedy, H & Hill, R (2017) 'The feeling of numbers: emotions in everyday engagements with data and their visualisation', *Sociology*, 52(4): 830-848. <https://doi.org/10.1177/0038038516674675>, p831.

⁷⁶ Coleman, S (2013) *How Voters Feel*, Cambridge University Press. <https://doi.org/10.1017/CBO9781139035354>

2.5. Trust or distrust?

A major theme in the research that we reviewed, which brings together the three topics we have highlighted so far – understanding, concern and emotions – is whether people trust data practices. Often this is examined through questions about who people trust with their data. The 2015 Northern Ireland Life and Times Survey asked people whether they trust a range of organisations to use their data, with the results of this module published in a special report. Most trusted were people's own GP surgeries, with 91% of respondents trusting them. This was followed by the NHS more generally (86%), government departments (73%), and academic researchers in universities (71%). At the other end of the scale were charities (51%) and commercial organisations like insurance companies (41%). 42% of respondents reported specific concerns about how these organisations used their data. The most common were that data might be used for purposes other than those for which it was originally shared, and that it might be hacked. The report's key conclusion is that 'Public support for data sharing is linked to trust in organisations'⁷⁷. This survey had a sample size of 1,202, with data collected face-to-face. The questions discussed here were part of a larger survey questionnaire, and the survey itself is part of an ongoing series.

A recent survey for the ODI, comparing the UK and France, asked a similar question but about different institutions. They found that in the UK, healthcare providers were the institutions people were most likely to trust with their personal data, at 64%, followed by banks, building societies and credit card companies, and family and friends, both at 57%. The public sector scored lower: 37% for central government, 41% for local government, and 25% for universities. The lowest scorers were social media organisations and offline retailers at 10% each, and marketing and advertising companies at just 2%. This was a one-off online survey undertaken by YouGov, with a sample of 2,023 in the UK⁷⁸. Digital Catapult presented respondents with a range of sectors and asked which they most trust with their personal data⁷⁹. The public sector (into which healthcare was categorised) was most trusted by 43.5% of respondents, with financial services coming second at 28.6%; other sectors were significantly further behind. This result is based on an online survey of 4,005 people conducted by Experian. Edelman reported that 24% of people trusted social media companies with their data, in contrast with 64% of people trusting traditional media, while 62% of people worried that social media companies would sell their personal data without them knowing⁸⁰. This report, which was about trust in general, is based on a sample of 2,000 people, and data was collected online.

The ICO also asked people how much they trusted different institutions to store and use personal data in an online survey of 2,259 people⁸¹. Once again healthcare was the most trusted sector, with 66% of people having high trust; this was followed by the police, financial services and government, with central government more trusted than local government. Just 15% of people trusted social media platforms. If people said they trusted institutions with their data, this was most commonly because of good previous experience or legislation. Those who were less trusting had either personally

⁷⁷ Robinson, G & Dolk, H (2015) *Research update: public attitudes to data sharing in Northern Ireland*. Administrative Research Data Centre, Northern Ireland <https://www.ark.ac.uk/publications/updates/update108.pdf>, p4.

⁷⁸ ODI / Open Data Institute (2018) *Who do we trust with personal data?* <https://theodi.org/article/who-do-we-trust-with-personal-data-odi-commissioned-survey-reveals-most-and-least-trusted-sectors-across-europe/>

⁷⁹ Digital Catapult (2015) *Trust in personal data: A UK review*. Digital Catapult

⁸⁰ Edelman (2018) *Edelman Trust Barometer 2018, UK Findings*. <https://www.edelman.co.uk/magazine/posts/edelman-trust-barometer-2018/>

⁸¹ ICO / Information Commissioner's Office – Harris Interactive (2019) *Information rights strategic plan: trust and confidence*. <https://ico.org.uk/media/about-the-ico/documents/2615515/ico-trust-and-confidence-report-20190626.pdf>

experienced data loss or attributed their limited trust to concerns about security and data sharing issues.

A report for Deloitte about government included questions about trust in government's uses of people's data⁸². This survey asked how much people trusted government and companies with their personal data. 56% of people trusted government either a great deal or a fair amount, compared with 31% for companies, with younger people and the more highly paid more likely to trust both government and companies. People were asked why they did or did not trust those institutions. Regulation was the most common reason for trust, and feeling a lack of control over how their personal data is used was the most common reason for distrust. Notably, when people distrusted companies, they were likely to refer to the possibility of their data being sold, whereas distrust in government resulted from a concern that it would not be able to keep data secure. This was based on Ipsos Mori's omnibus survey, with 1,071 people interviewed face-to-face⁸³. Similarly, Sopra Steria, an IT consultancy, found that people's main concern in relation to digital government was security, with 48% of respondents stating that they were worried that someone else could access their data⁸⁴. The authors conclude that 'Citizens are unlikely to use digital government services without a guarantee of privacy and security'.

The Higher Education Policy Institute (HEPI)/YouthSight conducted a study of students' trust in their higher education institutions⁸⁵. They found a large fraction of respondents did not trust universities' ability to protect their data, with only 45% responding that they were confident that their institution could keep their personal data secure and private, and 69% responding positively to the question 'Are you concerned by rumours that some universities are facing data security issues?' However, 'Are you concerned by rumours' is a leading framing, and it is possible that respondents had not heard of these rumours before carrying out the survey. This result was based on responses from 1,078 students currently in higher education in the UK, collected online and recruited through UCAS.

The report emerging from the ESRC project looking at how the public feel about private sector data being used in public research concludes that the public need to be informed about data practices in order to trust them. The authors argue that trust can be built with improved communication. They also suggest that a kitemark system could enable citizens to know who to trust⁸⁶. This is echoed in the findings of the Cabinet Office's consultation, Better Use of Data in Government, which found that the majority of respondents supported 'appropriate safeguards, accountability and transparency' as a way

⁸² Ipsos MORI (2018) *The state of the state 2017-2018: austerity, government spending, social care and data*.

<https://www.ipsos.com/sites/default/files/ct/publication/documents/2017-10/the-state-of-the-state-2017-2018.pdf>

⁸³ Ipsos MORI (2018) *The state of the state 2017-2018: austerity, government spending, social care and data*

<https://www.ipsos.com/sites/default/files/ct/publication/documents/2017-10/the-state-of-the-state-2017-2018.pdf>

⁸⁴ Sopra Steria (2017) *The citizen view of the digital transformation of government*.

<https://www.soprasteria.co.uk/docs/librariesprovider41/White-Papers/sopra-steria-ipsos-digitaltransformation-of-govt.pdf?sfvrsn=0>

⁸⁵ HEPI (2019) *Students or data subjects?* <https://www.hepi.ac.uk/wp-content/uploads/2019/12/Students-or-data-subjects-Report-122-Web-FINAL.pdf>

⁸⁶ Hopkins Van Mil: Creating Connections Ltd (2015) *Big data: public views on the use of private sector data for social research - a findings report for the Economic and Social Research Council*. <https://esrc.ukri.org/files/public-engagement/public-dialogues/public-dialogues-on-the-re-use-of-private-sector-data-for-social-research-report/>

of building trust in data usage⁸⁷. A report on Data for Public Benefit argues that improving data practices can also improve trust in government⁸⁸.

Research undertaken for Citizens Advice aimed to understand the meaning of trust (and other issues such as choice, transparency and control) in the context of expectations relating to personal data management⁸⁹. It focused on what consumers think and feel about these terms, and how they expect them to be applied in the context of online data usage. Participants identified changes that they would like to see and that they felt were achievable in order to feel a greater sense of trust towards institutions. These included: better information on what happens to their data and how profit is generated from data; greater accountability for larger companies; good privacy guarantees; ways to track who has access to data and why; the ability to withdraw permission for use if desired.

Some academic studies also focus on trust. Exploring the effects of system transparency on user trust with a convenience sample of 120 students, Kizilcec found that trust was more affected by user expectations, than the actual level of transparency itself⁹⁰. In other words, transparency about data practices only affected trust if user expectations were not met. In another study, Lee measured perceived fairness, trust and emotional response to managerial decisions made by algorithms versus humans with 228 Amazon Mechanical Turk respondents and found that the particular characteristics of tasks affect people's experiences and perceptions of fairness with algorithmic technologies⁹¹.

There is less qualitative research focused on trust in data practices, but Pink and others and Steedman and colleagues are worthy of note⁹². Pink et al's study of how design professionals deal with data-related anxieties focused specifically on the question of trust. The authors assessed the strategies that individuals use to navigate uncertainties relating to data in their everyday lives, including at work. They found that people build specific routines for managing their data and this helped them to feel trust in data practices, which they argue is a feeling that enables people to move on and take action in the future.

Drawing on the same empirical study as Kennedy et al⁹³ mentioned above, Steedman and colleagues draw attention to the complex range of factors that come together to engender, maintain or undermine trust in data practices⁹⁴. These relate to whether people trust the institution that is

⁸⁷ Cabinet Office & Government Digital Service (2016) *Better use of data in government*.

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/535063/better_use_of_data_in_government_response_final.pdf, p1

⁸⁸ Carnegie Trust - Scott, Kaela (2018) *Data for public benefit: balancing the risks and benefits of data sharing*.

https://www.involve.org.uk/sites/default/files/field/attachemnt/Data%20for%20Public%20Benefit%20Report_0.pdf, p20

⁸⁹ Citizens Advice - Illuminas (2016) *Consumer expectations for personal data management in the digital world*.

<https://www.citizensadvice.org.uk/Global/CitizensAdvice/Consumer%20publications/Personal%20data%20consumer%20expectations%20research.docx.pdf>

⁹⁰ Kizilcec, R F (2016) 'How much information? effects of transparency on trust in an algorithmic interface', *CHI '16: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, May 2016.

<https://doi.org/10.1145/2858036.2858402>

⁹¹ Lee, M K (2018) 'Understanding perception of algorithmic decisions: fairness, trust, and emotion in response to algorithmic management', *Big Data & Society*, 5(1). <https://doi.org/10.1177/2053951718756684>

⁹² Pink, S, Lanzeni, D & Horst, H (2018) 'Data anxieties: finding trust in everyday digital mess', *Big Data & Society*, 5(1). <https://doi.org/10.1177/2053951718756685>

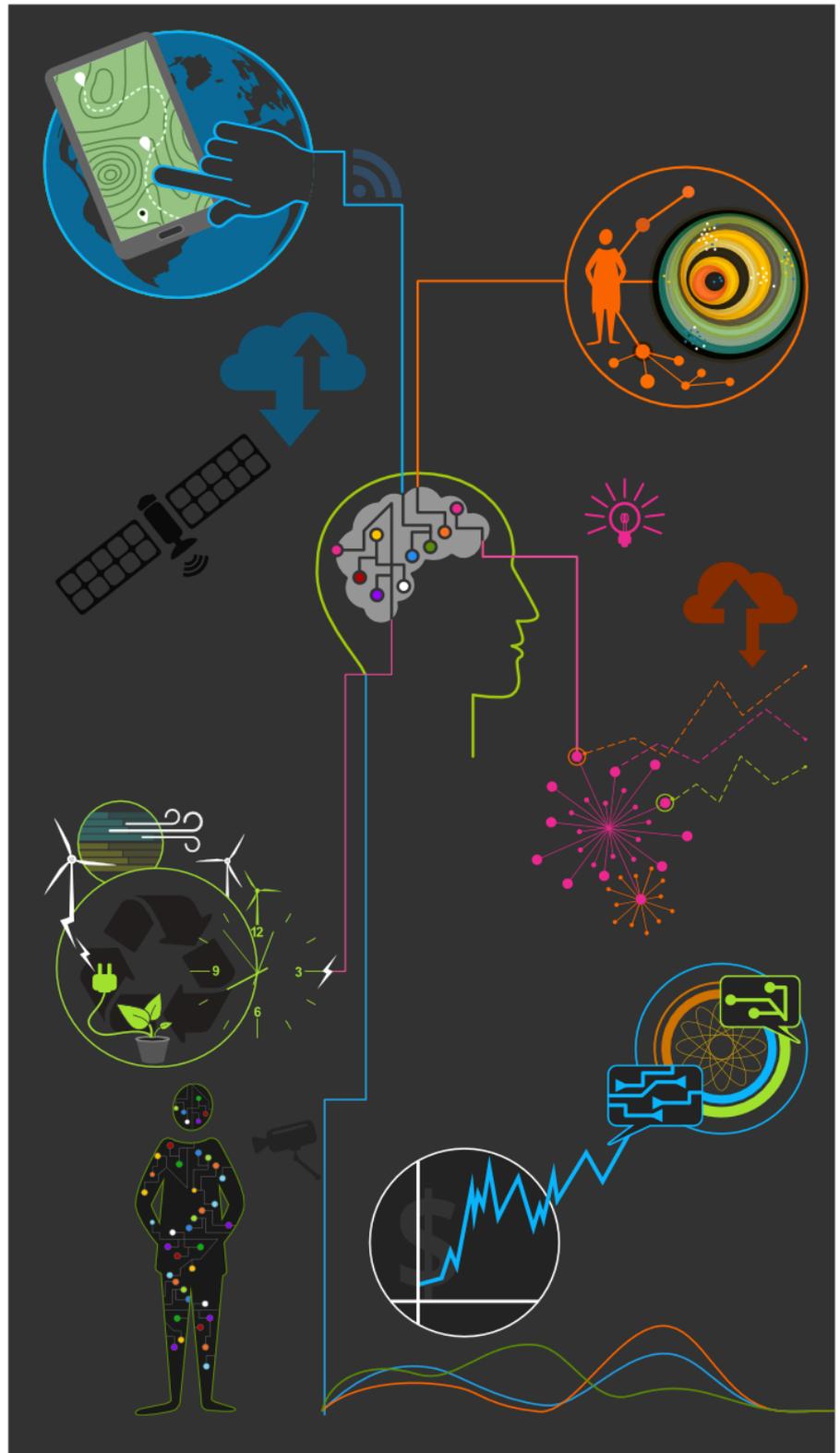
⁹³ Kennedy, H, Steedman, R & Jones, R (2020a) 'Approaching public perceptions of datafication through the lens of inequality: a case study in public service media', *Information, Communication and Society*.

<https://doi.org/10.1080/1369118X.2020.1736122>

⁹⁴ Steedman, R, Kennedy, H & Jones, R (2020) 'Complex ecologies of trust in data practices and data-driven systems', *Information, Communication and Society*. <https://doi.org/10.1080/1369118X.2020.1748090>

gathering data in general, whether they trust it specifically to manage their data securely, degrees of trust in the broader data ecosystem, and even whether they trust themselves to manage their own data carefully and thoughtfully. They use the term ‘complex ecologies of trust’ to capture this and the interplay between trust, scepticism and distrust that were found to co-exist. They point out that distrust is often appropriate, if organisational data practices are not deemed trustworthy, as in the case of scandals about data breaches.

Taken together, surveys indicate that people are most likely to trust healthcare institutions, followed by banks and local government. They are much less likely to trust retailers, with marketing organisations and social media companies the least trusted of all. Overall levels of trust in social media companies are low, and concerns about what happens to people’s online data are high. Variations in findings can partly be explained by differences in questions. Where people do not trust organisations, this is because of concern that organisations will sell or share data without consent (in the case of the private sector) and that organisations are vulnerable (in the case of the public sector). There are inconsistencies around how levels of trust vary by age, with young people being estimated as both the most and least trusting, depending on the measure being used. Like other qualitative studies discussed in this review, qualitative research on trust challenges simple understandings of trust and distrust as clearly distinct and separate. Such research draws attention to the multiple, interrelated, context-dependent layers of trust and distrust that people feel in their interactions with data practices.



2.6. Acceptance or resignation?

The concept of the 'privacy paradox' has been used to explain why people often say they are concerned about online privacy and protecting their personal data and yet they act in ways that do not support this claim. To explain this paradox, it has been argued that people weigh up the risks and benefits of sharing data, and decide whether to 'trade off' the risks against perceived benefits. This leads to an assertion which is sometimes made, that citizens accept data practices. A number of studies have explored whether this is indeed acceptance, or if it is indifference, resignation or other, neither positive nor negative responses. Ellis, for example, carried out 31 semi-structured interviews to investigate 'multiple forms of surveillance' and why people appear indifferent and/or apathetic towards state-corporate forms of surveillance⁹⁵. He explored whether the muted responses that he perceived equate with consent, indifference, apathy, or something else and coins the phrase 'surveillance-apatheia,' arguing that apathy was a significant finding.

Other researchers interpret similar phenomena differently. Dencik and Cable, for example, state that there is widespread resignation to personal data gathering, especially when this is undertaken for surveillant purposes, and they seek to explore this phenomenon through focus groups with members of the public and interviews with activists⁹⁶. As a result of their research, they argue that resignation is a response to 'surveillance realism', or the normalisation of surveillance practices in everyday life. It is not so much acceptance, apathy or indifference, but rather a feeling that there is no choice but to live with data-driven surveillance. In both of these cases, apathy and resignation are assumed starting points for empirical research; they are not empirical research findings.

In the US, Joseph Turow has undertaken a number of surveys since 2003 into Americans' views of the online tracking of their digital data and the things that result from it, such as targeted advertising. These have mostly been published as grey literature⁹⁷. As such, given our geographical criterion to exclude non-UK focused grey literature, we do not discuss these publications in detail here. Nonetheless, we feel it is important to mention them, given that they represent ground-breaking research into public perceptions of data practices which originated long before the current flurry of studies on this topic. These surveys also form the basis of academic publications, international samples of which we do include, as noted in the introduction.

The most significant contribution by Turow and his collaborators in relation to the 'acceptance or resignation?' question is their concept of 'digital resignation'⁹⁸. Draper and Turow use this term to describe what they saw in their empirical research into people's views on personal data use. They argue that people want to control the data that corporations have about them but feel unable to do so, and that these feelings of futility result from and are a rational response to the activities of said

⁹⁵ Ellis, D (2019) 'Techno-securitisation of everyday life and cultures of surveillance-apatheia', *Science as Culture*, 29(1), 11-29. <https://doi.org/10.1080/09505431.2018.1561660>

⁹⁶ Dencik, L & Cable, J (2017) 'Digital Citizenship and Surveillance: The advent of surveillance realism: public opinion and activist responses to the Snowden leaks', *International Journal of Communication*, 11(2017):763-781. <https://ijoc.org/index.php/ijoc/article/view/5524/1939>

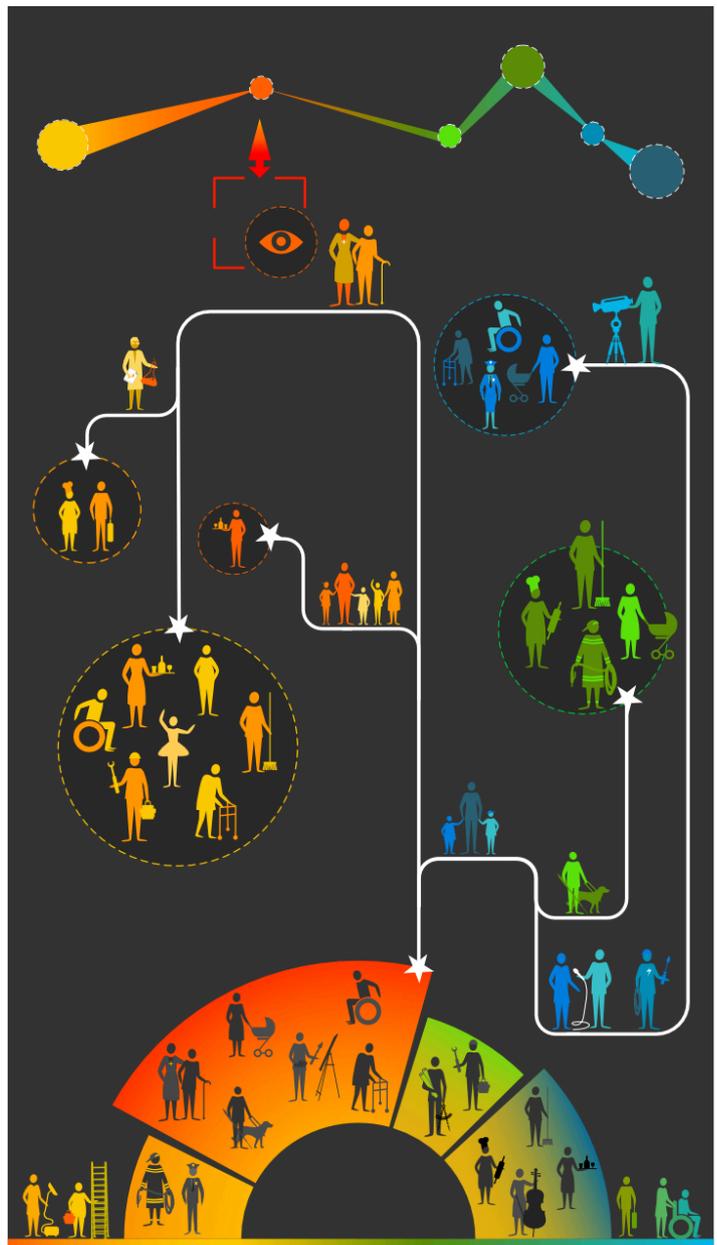
⁹⁷ For example Turow, J, Hennessy, M & Draper, N (2015) *The trade off fallacy*. https://www.asc.upenn.edu/sites/default/files/TradeoffFallacy_1.pdf and Turow, J, Hennessy, M, Draper, N, Akanbi, O & Virgilio, D (2018) *Divided we feel: partisan politics drive American's emotions regarding surveillance of low-income populations*. https://repository.upenn.edu/cgi/viewcontent.cgi?article=1563&context=asc_papers

⁹⁸ Draper, N A, & Turow, J (2019) 'The corporate cultivation of digital resignation', *New Media & Society*, 21(8):1824-1839. <https://doi.org/10.1177/1461444819833331> and Turow, J, Hennessy, M & Draper, N (2015) *The trade off fallacy*. https://www.asc.upenn.edu/sites/default/files/TradeoffFallacy_1.pdf

corporations. They conclude that the uneven power relationships between companies and publics that result in digital resignation are a problem that needs to be addressed.

An early focus group study of perceptions of social media data mining by Kennedy et al also produced findings that challenge the suggestion that people simply accept or are resigned to data practices, even though their study was not ostensibly focused on this issue⁹⁹. As noted above, they found that participants undertook measured, case-by-case evaluations of social media data mining practices that were put to them in discussions, weighing up a range of factors in order to decide whether they considered each case as fair. Examples discussed included Facebook's uses of its users' personal data, commercial companies mining data on behalf of their clients, and academic uses of social media data mining. Participants did not simply accept, reject, or feel indifferent to the data practices in question. Their responses varied, depending on the case in hand.

In summary, it is simplistic to see responses to data practices as acceptance, indifference or apathy. Researchers have pointed out that these responses need to be understood in the context of the ways in which powerful commercial and governmental actors normalise data mining, and that people feel unable to control the flows of their personal data even if they want to. People are able to evaluate data practices in thoughtful ways, and they want to exercise agency in relation to their data, but the conditions do not currently exist that enable them to do so.



⁹⁹ Kennedy, H, Elgesem, D, & Miguel, C (2015) 'On fairness: user perspectives on social media data mining', *Convergence*, 23(3): 270–288. <https://doi.org/10.1177/1354856515592507>

2.7 What fair data practices look like, from public perspectives

The above sections summarise people's understandings of and attitudes towards the use of personal data. This relates to *Living With Data's* first research question: *What do different people know and feel about specific data-related practices in different domains of everyday life?* Some of the literature focuses on how people feel personal data *should* be used. This relates to *Living With Data's* second research question: *What do fair data practices look like, from non-experts' perspectives?*

One example is a report for the RSA which focuses on uses of Artificial Intelligence (AI) in decision-making, rather than data practices more generally¹⁰⁰. The RSA commissioned a YouGov online survey of 2,000 people which assessed people's attitudes towards uses of AI in decision-making in various domains. The report, which also draws on other methods, concludes that people are not supportive of using AI for decision-making, especially in the workplace and in the criminal justice system (60% of people oppose or strongly oppose its use in these domains). People were more comfortable with the use of AI in decision-making in financial services and in advertising. Respondents were asked first about which potential problems they were most concerned with, followed by a question about which potential benefits they were most looking forward to. While 30% of people responded that they weren't looking forward to any potential benefits, the most commonly-selected potential benefits related to improving accuracy, consistency and efficiency. The biggest concern was a lack of empathy inherent in automated decision-making, although significant numbers of people were also concerned about accountability and oversight across all potential uses. Few people (just 19%) believed that using automated decision-making could reduce existing biases or inequalities. There was also little optimism about future uses of AI: just 26% of people were comfortable with more decisions being fully automated in the future as the accuracy and consistency of automated systems improves. This report consistently compares opportunities and concerns and presents balanced questions. Given this balanced approach, it is noteworthy that there are few uses of AI in decision-making that the public would consider to be fair.

A survey by the Ada Lovelace Institute on attitudes to uses of facial recognition technology found that people are broadly supportive of police uses of these technologies, with 55% of people thinking they should be restricted to particular circumstances, and 71% believing the police should have the right to use them in public spaces to reduce crime¹⁰¹. However, 50% of people believe that the private sector should not sell facial recognition technologies to the police. There is far less support for other sectors using facial recognition; 70% of people believe they should not be used in schools, only 22% believe they should be permitted on public transport to identify whether passengers have paid, and just 7% believe they should be permitted in supermarkets to track shopper behaviour. Findings show that people differentiate the contexts in which uses of facial recognition might be fair (in policing) and those in which they are not (in most other contexts). But they also indicate that contradictory views co-exist (the police should have the right to use facial recognition technologies, but the private sector should not sell them to the police).

Other survey reports include sections about what people believe should happen with their personal data, but are not fully focused on this issue. The Big Brother Watch & ComRes report states that 58%

¹⁰⁰ RSA / The Royal Society for the Encouragement of Arts, Manufactures and Commerce Royal Society of Arts (2018) *Artificial intelligence: real public engagement*. <https://www.thersa.org/discover/publications-and-articles/reports/artificial-intelligence-real-public-engagement>

¹⁰¹ Ada Lovelace Institute (2019) *Beyond face value - public attitudes to facial recognition technology*. https://www.adalovelaceinstitute.org/wp-content/uploads/2019/09/Public-attitudes-to-facial-recognition-technology_v.FINAL_.pdf

of their respondents felt that companies should never gather personal data unless they explain why and the people whose data is being collected specifically give permission each time, while 16% felt that companies should never gather personal data under any circumstances¹⁰². Similarly, 43% of people felt that it was never acceptable for data about purchase, browser, search, location, or email history to be collected, and 51% felt that it was never acceptable for this data to be combined with other data. These findings suggest that for respondents in this survey, greater transparency about data gathering and less data gathering in general, especially of particular types of data, might be seen as fair.

Research commissioned by Understanding Patient Data, Involve and Carnegie UK consulted people from public and voluntary organisations on what they think fair data practices might involve. The authors of the resulting report point out that this particular research did not aim to ‘determine absolutes of what is and what is not acceptable’ in relation to data practices¹⁰³. The workshops that they undertook confirmed that this would indeed be a difficult undertaking, because they found significant variation across sectors with regard to how the benefits and risks of data sharing were understood. Other grey literature reporting research which takes a qualitative approach suggests that fair data practices are characterised by: accountability and agency¹⁰⁴, the balancing of risks and benefits¹⁰⁵, attention to issues of social justice¹⁰⁶, or the improvement of public services¹⁰⁷. From this we can conclude that fairness is contingent on context.

The Cabinet Office’s *Better Use of Data in Government* touches on issues of fairness¹⁰⁸. This is a report based on a consultation to which there were 282 responses, 152 classified as from citizens. Questions focusing specifically on fair data uses addressed whether governments should be able to access personal data, such as date of birth, in order to administer child benefit, for example. The majority of respondents were supportive of these proposals, as long as appropriate safeguards, accountability and transparency were in place to build trust with citizens on the usage of their data. Data sharing for the public good was seen as fair, as long as it is accompanied by safeguarding people and data.

Doteveryone’s report *Better redress: building accountability for the digital age: an evidence review* aimed to ‘chart public perception, the changing regulatory systems and the gaps that exist’ by

¹⁰² Big Brother Watch & ComRes (2015) *UK public research – online privacy*. <https://www.bigbrotherwatch.org.uk/wp-content/uploads/2015/03/Big-Brother-Watch-Polling-Results.pdf>

¹⁰³ Carnegie Trust - Scott, Kaela (2018) *Data for public benefit: balancing the risks and benefits of data sharing*. https://www.involve.org.uk/sites/default/files/field/attachemnt/Data%20for%20Public%20Benefit%20Report_0.pdf

¹⁰⁴ Accountability and agency: Hopkins Van Mil: Creating Connections Ltd (2015) *Big data: public views on the use of private sector data for social research - a findings report for the Economic and Social Research Council*. <https://esrc.ukri.org/files/public-engagement/public-dialogues/public-dialogues-on-the-re-use-of-private-sector-data-for-social-research-report/>.

¹⁰⁵ Carnegie Trust - Scott, Kaela (2018) *Data for public benefit: balancing the risks and benefits of data sharing*. https://www.involve.org.uk/sites/default/files/field/attachemnt/Data%20for%20Public%20Benefit%20Report_0.pdf and RSA / The Royal Society for the Encouragement of Arts, Manufactures and Commerce Royal Society of Arts - Renate Samson, Kayshani Gibbon & Anna Scott (2019) *About data about us*. <https://www.thersa.org/discover/publications-and-articles/reports/data-about-us>

¹⁰⁶ Oman, S (2019a) *Improving data practices to monitor inequality and introduce social mobility measures: a working paper*. The University of Sheffield. https://www.sheffield.ac.uk/polopoly_fs/1.867756!/file/MetricsWorkingPaper.pdf

¹⁰⁷ Cabinet Office & Government Digital Service (2016) *Better use of data in government*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/535063/better_use_of_data_in_government_response_final.pdf and Carnegie Trust - Scott, Kaela (2018) *Data for public benefit: balancing the risks and benefits of data sharing*. https://www.involve.org.uk/sites/default/files/field/attachemnt/Data%20for%20Public%20Benefit%20Report_0.pdf

¹⁰⁸ Cabinet Office & Government Digital Service (2016) *Better use of data in government*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/535063/better_use_of_data_in_government_response_final.pdf

analysing previous research and carrying out an open public consultation. The report argues that redress is important, in that it enables ‘individuals and communities to seek remedies for unfair treatment and can help tip the scales when a power imbalance becomes too great’¹⁰⁹. This implies that redress is an important component of a fair data ecosystem, because it is a mechanism for dealing with perceived unfairness. The RSA’s report on ethical AI states that their research findings support honesty and transparency as elements of fair and ethical uses of AI (alongside compliance and enforceability)¹¹⁰. Others who have researched how people feel about data practices also advocate that transparency about data use should be built into institutional practice through open and approachable communications with people about what their data is for¹¹¹.

In part, building on learning from preliminary analysis of some of the research reviewed here, Kennedy et al carried out a survey on public views of eight different data management models as a way of addressing the question of what fair data practices look like, from public perspectives¹¹². Their findings are based on an online sample of 2,169 people recruited through the Qualtrics platform. Their survey assessed a wide range of different models for managing personal data, including: the predominant current model which gives services control over user data; a personal data store which gives individuals control over their data; several collective, trust-like models; and the option of opting out of data gathering. They also asked knowledge questions and explored attitudes towards uses of personal data more broadly. Consistent with other reports and as noted above, they found that respondents also felt strongly that their personal data *should* be handled in particular ways: for example, 89% wanted more control over their personal data, and 84% felt that data should be gathered, analysed and managed in ethical ways. Using a series of conjoint analyses, they found that the current model was the least well-liked option of all, and that personal data stores and a regulatory public body were the most widely-preferred models of data management. The authors conclude that their findings suggest that new approaches to data management are urgently needed, because there is a strong desire from the public for an alternative to the status quo – that is, the current model.

In the particular field of health data, Understanding Patient Data (UPD) produced a summary of existing research into public attitudes to patient data use in September 2018¹¹³. As noted in the introduction, we do not focus on the health domain in this report more is known about public perceptions of datafication in the health sector than in other domains, but we mention this summary here because it provides valuable insight into what people might consider to be fair uses of health data. The summary concludes that even though people have limited understanding of the ways patient data is used in health, most people support sharing patient data for individual care and many support sharing patient data for research where there is public benefit. The report notes that a key factor in

¹⁰⁹ Doteveryone - Joe Massey, Jacob Ohrvik-Stott & Catherine Miller (2019) *Better redress: building accountability for the digital age: an evidence review from Doteveryone* <https://www.doteveryone.org.uk/wp-content/uploads/2019/12/Better-redress-evidence-review.pdf>, p1

¹¹⁰ RSA / The Royal Society for the Encouragement of Arts, Manufactures and Commerce Royal Society of Arts - Renate Samson, Kayshani Gibbon & Anna Scott (2019) *About data about us*. <https://www.thersa.org/discover/publications-and-articles/reports/data-about-us>

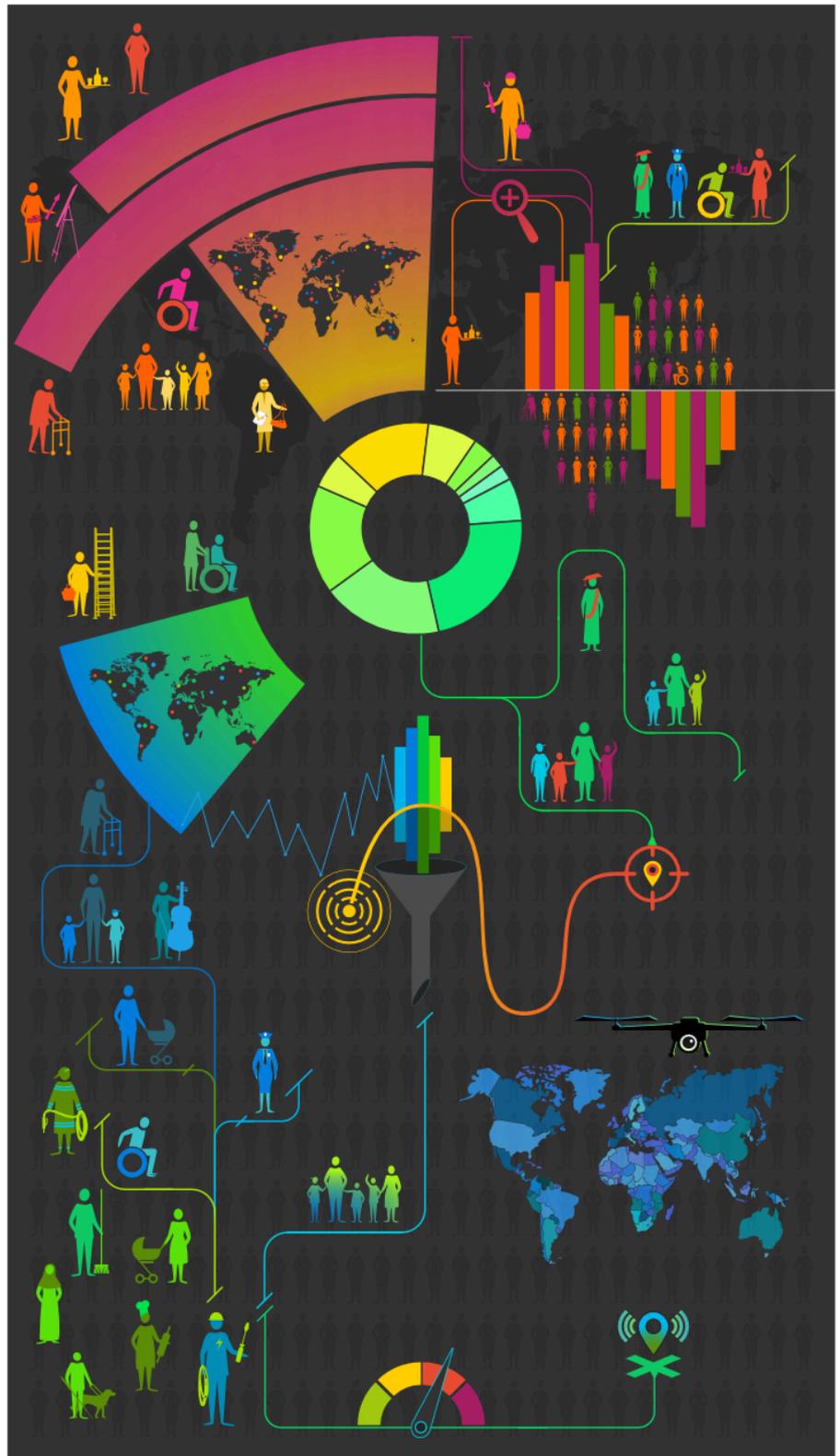
¹¹¹ For example: Oman, S (2019b) Measuring social mobility in the creative and cultural industries – the importance of working in partnership to improve data practices and address inequality. The University of Sheffield. https://www.sheffield.ac.uk/polopoly_fs/1.867754!/file/MetricsPolicyBriefing.pdf and Popham, J Lavoie, J & Coomber, N (2020) ‘Constructing a public narrative of regulations for big data and analytics: results from a community-driven discussion’, *Social Science Computer Review*, 38(1)75-90. <https://doi.org/10.1177/0894439318788619>

¹¹² Kennedy, H, Hartman, T, Steedman, R & Jones, R (2020b) *UK public unhappy with the ways their data is managed*. <https://livingwithdata.org/project/wp-content/uploads/2020/03/Views-on-Data-Management-Full-Report.pdf>

¹¹³ Understanding Patient Data (2018) Public attitudes to patient data use: a summary of existing research. http://understandingpatientdata.org.uk/sites/default/files/2019-01/Public%20attitudes%20key%20themes_0.pdf

gathering support is knowing who has access to patient data and how it is used. Support exists even though there are concerns, in part because the NHS is highly trusted, compared to other organisations, as noted above. Using real-world examples, being transparent and encouraging genuine dialogue (rather than purely transmitting information) are ways in which organisations can increase trust in their data practices, according to the UPD summary. This might be seen as a recipe for fair data practices.

While the research discussed in this section adopts different approaches, it consistently suggests that people have strong normative opinions about uses of personal data and what fair data practices might look like. Research consistently finds dissatisfaction with current ways in which data is used and managed, and a desire for change. More honesty, transparency and genuine dialogue with the public; better regulation, enforcing compliance, the existence of safeguards and accountability, and the right to redress; and greater personal control are some of the changes that are proposed. The widespread view that data practices could be fairer than they currently are inevitably leads to consideration of what needs to change. We address this topic in the next section.



2.8. State and industry need to enact change, according to the research

A lot of the literature discussed in this review makes suggestions about the changes that research findings point towards. Grey literature is especially solutions-oriented, as it generally aims to influence policy or practice in some way. The literature commonly identifies one or more of the following as responsible for implementing change: the state (through changes to policy); industry (through changes to practice); people themselves (through better understanding of data-driven systems which will enable them to manage their personal data appropriately – this is often dependent on state and industry changes such as better communication or greater transparency). Often, proposed changes involve bringing two or more of these actors together (for example, greater transparency about data practices could involve industry and the state). Sometimes, change is proposed, but it is not clear who should implement it.

The way in which a topic is approached often determines whether changes of some kind are recommended, and what kinds of change are proposed. What we describe as **everyday life-oriented literature** focuses on people's experiences of data practices and seeks to understand these experiences and the embeddedness of data in everyday life. This literature tends to be interested in exploring the extent to which people have agency in the face of data practices, responding to critical literature which suggests they do not. It is often qualitative, academic literature, concerned to advance understanding and contribute to theory, or ways of thinking, rather than making suggestions about what should change. As such, it often does not contemplate what kinds of change might be needed. This literature often emerges from media and communication studies, sociology, anthropology or cognate disciplines, and tends to be less solutions-oriented than literature taking the other approaches discussed below. Many examples of this kind of literature are discussed in 2.3. People are not only concerned¹¹⁴. Some grey literature also focuses on everyday life¹¹⁵.

¹¹⁴ For example: Bucher, T (2017) 'The algorithmic imaginary: exploring the ordinary affects of Facebook algorithms', *Information, Communication & Society*. 20: 30-44. <https://doi.org/10.1080/1369118X.2016.1154086>;
Couldry, N, Fotopoulou, A & Dickens, L (2016) 'Real social analytics: a contribution towards a phenomenology of a digital world', *The British Journal of Sociology* 67(1): 118-137. <https://doi.org/10.1111/1468-4446.12183>;
Lupton, D (2017) "'It just gives me a bit of peace of mind': Australian women's use of digital media for pregnancy and early motherhood", *Societies* 7(3)25. <https://doi.org/10.3390/soc7030025>; Lupton, D (2019) 'Data mattering and self-tracking: what can personal data do?', *Continuum*, 34(1):1-13. <http://dx.doi.org/10.1080/10304312.2019.1691149>;
Pink, S & Fors, V (2017) 'Being in a mediated world: self-tracking and the mind–body–environment', *Cultural Geographies*. 24(3)375-388. <https://doi.org/10.1177/1474474016684127>; Pink, S, Sumartojo, S, Lupton, D & La Bond, C H (2017) 'Mundane data: the routines, contingencies and accomplishments of digital living', *Big Data & Society*, 4(1). <https://doi.org/10.1177/2053951717700924>; Pridmore, J & Mols, A (2020) 'Personal choices and situated data: Privacy negotiations and the acceptance of household Intelligent Personal Assistants', *Big Data & Society*, 7(1). <https://doi.org/10.1177/2053951719891748>; Ruckenstein, M S (2017) 'Keeping data alive: talking DTC genetic testing', *Information, Communication and Society*, 20(7):1024-1039. <https://doi.org/10.1080/1369118X.2016.1203975>; Ruckenstein, M & Pantzar, M (2015) 'Datafied Life: techno-anthropology as a site for exploration and experimentation', *Techné: Research in Philosophy and Technology*, 19(2):191-210. <https://doi.org/10.13140/RG.2.1.2553.7762>; Weiner, K, Will, C, Henwood, F, Williams, R (2020) 'Everyday curation? Attending to data, records and record keeping in the practices of self-monitoring', *Big Data & Society* 7(1). <https://doi.org/10.1177/2053951720918275>; and Wilmott, C (2016) 'Small moments in spatial big data: calculability, authority and interoperability in everyday mobile mapping', *Big Data & Society*, 3(2). <https://doi.org/10.1177/2053951716661364>.

¹¹⁵ Such as: Citizens Advice - Illuminas (2016) Consumer expectations for personal data management in the digital world. <https://www.citizensadvice.org.uk/Global/CitizensAdvice/Consumer%20publications/Personal%20data%20consumer%20expectations%20research.docx.pdf> and RSA / The Royal Society for the Encouragement of Arts, Manufactures and Commerce Royal Society of Arts - Renate Samson, Kayshani Gibbon & Anna Scott (2019) *About data about us*. <https://www.thersa.org/discover/publications-and-articles/reports/data-about-us>

In the rest of the literature we have reviewed, we identify three broad approaches which tend to result in particular recommendations for change:

1. Systems-oriented literature, such as that which emerges from the field of Human Computer Interaction (HCI), which aims to understand experiences of system use and often suggests changes to systems, design or user experience, to be implemented by industry;
2. Critical academic literature, which focuses on actual and potential harms and negative consequences of widespread data practices, and often suggests regulatory changes, such as stricter governance of industry;
3. Policy and practice-oriented grey literature, which aims to influence policy and practice in some way, and makes diverse recommendations for change.

We discuss these approaches and their proposed solutions in more detail below, referring back to literature discussed in previous sections.

2.8.1. Systems-oriented literature recommends design and systems change



Systems-oriented literature often focuses on systems design or user experience (UX). Within this field, researchers sometimes experiment with approaches to HCI or UX design which might improve the effectiveness of data-driven systems. For example, Eslami et al carried out a study with 40 Facebook users in which they explored awareness of the ways in which Facebook algorithms curate the content that users see in their newsfeeds¹¹⁶. They developed a tool called FeedVis, which revealed the difference between algorithmically curated content and content that has not been algorithmically curated. They found that, unaware of algorithmic curation, people interpreted the absence of friends' content as exclusion from these friends' lives. Greater algorithmic awareness improved understanding and led to feelings of greater control, which in turn led to more active engagement.

¹¹⁶ Eslami, M, Rickman, A, Vaccaro, K, Aleyasen, A, Vuong, A, Karahalios, K, Hamilton, K & Sandvig, C (2015) 'I always assumed that I wasn't really that close to [her]: reasoning about invisible algorithms in news feeds', *CHI '15: Proceedings of the 2015 CHI Conference on Human Factors in Computing Systems*, April 2015. <https://doi.org/10.1145/2702123.2702556>

In another study by Eslami and others, the focus was on how people perceive algorithmically suggested advertisements¹¹⁷. The authors conclude that clear and neutral communication about how algorithms shape visible content improve the experience of platform usage. Elsewhere, Brown et al carried out a qualitative study of the perspectives of those affected by algorithmic decision-making in child welfare services¹¹⁸. Through this, they identify strategies to make vulnerable communities more comfortable with algorithmic processing, proposing more accountable algorithm design in child welfare applications and transparency as design solutions to improve trust for vulnerable communities. In another example, Fiesler and Hallinan analysed public comments on news articles about online data sharing and privacy controversies, in order to understand reactions to these controversies and what they describe as 'pervasive privacy attitudes'¹¹⁹. Like other authors mentioned here, they also conclude by identifying platform communication and design strategies that would benefit users and therefore the platforms themselves.

Horne and Przepiorka examined the relationship between technology provider behaviour, norms and the perceived trustworthiness of data-driven systems¹²⁰. They found that the more common usage of particular systems becomes, the weaker privacy expectations become. Being informed of privacy violations by systems providers has less impact on people choosing to use these technologies than might be expected, as people take other people's use of systems as evidence of good privacy practices therein. Thus, widespread use of technologies by peers is likely to further increase usage, even where there are privacy concerns. The authors propose that technology providers need to take these complex privacy dynamics into account. (See also Machuletz et al whose study of webcam covering leads them to conclude that developers should design privacy-enhancing technologies which align with users' privacy behaviours¹²¹.) Elsewhere, Sannon et al explored how, when and why people lie to protect their privacy in different online contexts¹²². This research also acknowledges the complex contextual factors and motivations in individuals' data-related behaviours, and concludes by proposing the use of a statistical model to predict privacy lies.

A small number of HCI studies focus on social issues and make recommendations for addressing them. As noted above, Guberek et al explored risk perceptions and technology use amongst undocumented migrants in the US¹²³. Recommendations include developing community appropriate educational

¹¹⁷ Eslami, M, Sneha, R, Kumaran, K, Sandvig, C & Karahalios, K (2018) 'Communicating algorithmic process in online behavioral advertising', *CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, April 2018 Paper No.: 432. <https://doi.org/10.1145/3173574.3174006>

¹¹⁸ Brown, A, Chouldechova, A, Putnam-Hornstein, E, Tobin, A & Vaithianathan, R (2019) 'Toward algorithmic accountability in public services: a qualitative study of affected community perspectives on algorithmic decision-making in child welfare services', *CHI '19: Proceedings of the 2019 Conference on Human Factors in Computing Systems*, May 2019. Paper No.: 41. <https://doi.org/10.1145/3290605.3300271>

¹¹⁹ Fiesler, C & Hallinan, B (2018) 'We are the product: public reactions to online data sharing and privacy controversies in the media', *CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, April 2018 Paper No.: 53. <https://doi.org/10.1145/3173574.3173627>

¹²⁰ Horne, C & Przepiorka, W (2019) 'Technology use and norm change in online privacy: experimental evidence from vignette studies', *Information, Communication & Society*, (0):1-17. <https://doi.org/10.1080/1369118X.2019.1684542>

¹²¹ Machuletz, D, Laube, S, Böhme, R (2018) 'Webcam covering as planned behavior', *CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, April 2018, Paper No.: 180. <https://doi.org/10.1145/3173574.3173753>

¹²² Sannon, S, Bazarova, N N & Cosley, D (2018) 'Privacy lies: understanding how, when, and why people lie to protect their privacy in multiple online contexts', *CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, April 2018 Paper No.: 52 Pages 1–13. <https://doi.org/10.1145/3173574.3173626>

¹²³ Guberek, T, McDonald, A, Simioni, S, Mhaidli, A, Toyama, K, Schaub, F (2018) 'Keeping a Low Profile?: Technology, Risk and Privacy among Undocumented Immigrants', *CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, April 2018 Paper No.: 114. <https://doi.org/10.1145/3173574.3173688>

resources; being careful with organisational communication; supporting on-demand information hiding; limiting exposure of identifying information and providing virtual sanctuary for undocumented migrants. In some cases, these recommendations are directed at specific actors, including the major social media platforms; in others, it is not stated who might be responsible for implementing recommendations. Similarly, Alvarado and Waern studied experiences of algorithms on social media, and they conclude that more user control would enhance the experience of usage, but they do not specify how this might be achieved¹²⁴.

Bowyer et al explored family perspectives on the storage, sharing and handling of family civic data, through ethnographic interviews with four families in the north east of England¹²⁵. The authors deliver their findings as design guidelines, which suggests that they can be implemented by developers working in industry, and yet they propose that people should have more control over their own data, something that needs regulatory support. Families would feel most empowered through an ongoing co-operative relationship with the state on data use, the authors claim. It must be noted that this was a very small-scale study, and findings are not generalisable.

Apart from this example by Bowyer et al, most of the literature discussed in this section proposes solutions at the systems level. Eslami et al suggest that clear and neutral communication improves system effectiveness, while Fiesler and Hallinan also identify communication and design strategies as solutions¹²⁶. Horne and Przepiorka argue that systems design needs to take account of complex privacy dynamics and Sannon et al also recognise that complex contextual factors play a role in data privacy behaviours, suggesting that statistical modelling can help to predict such behaviours¹²⁷. Guberek et al make a number of recommendations relating to technology and systems design¹²⁸. On the whole, HCI literature does not engage in discussion of the range of social and political factors that shape the contexts within which data-driven systems are deployed, nor differentiate between user populations and their diverse experiences of data practices.

¹²⁴ Alvarado, O & Waern, A (2018) 'Towards algorithmic experience: initial efforts for social media contexts', *CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, April 2018. Paper No.: 286. <https://doi.org/10.1145/3173574.3173860>

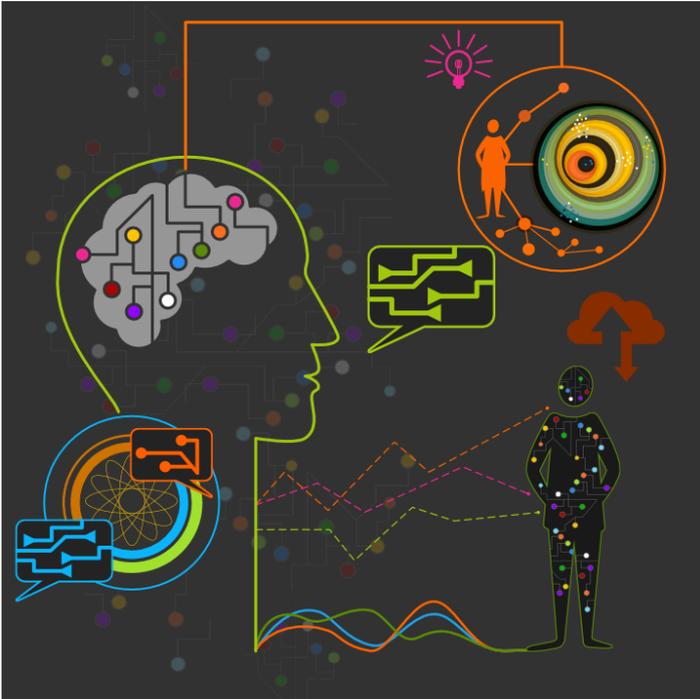
¹²⁵ Bowyer, A, Montague, K, Wheeler, S, McGovern, S, Lingam, R & Balaam, M (2018) 'Understanding the family perspective on the storage, sharing and handling of family civic data', *CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, April 2018. Paper No.: 136. <https://doi.org/10.1145/3173574.3173710>

¹²⁶ Eslami, M, Rickman, A, Vaccaro, K, Aleyasen, A, Vuong, A, Karahalios, K, Hamilton, K & Sandvig, C (2015) 'I always assumed that I wasn't really that close to [her]: reasoning about invisible algorithms in news feeds', *CHI '15: Proceedings of the 2015 CHI Conference on Human Factors in Computing Systems*, April 2015. <https://doi.org/10.1145/2702123.2702556> and Fiesler, C & Hallinan, B (2018) 'We are the product: public reactions to online data sharing and privacy controversies in the media', *CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, April 2018 Paper No.: 53. <https://doi.org/10.1145/3173574.3173627>

¹²⁷ Horne, C & Przepiorka, W (2019) 'Technology use and norm change in online privacy: experimental evidence from vignette studies', *Information, Communication & Society*, (0):1-17. <https://doi.org/10.1080/1369118X.2019.1684542> and Sannon, S, Bazarova, N N & Cosley, D (2018) 'Privacy lies: understanding how, when, and why people lie to protect their privacy in multiple online contexts', *CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, April 2018 Paper No.: 52 Pages 1–13. <https://doi.org/10.1145/3173574.3173626>

¹²⁸ Guberek, T, McDonald, A, Simioni, S, Mhaidli, A, Toyama, K, Schaub, F (2018) 'Keeping a Low Profile?: Technology, Risk and Privacy among Undocumented Immigrants', *CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, April 2018 Paper No.: 114. <https://doi.org/10.1145/3173574.3173688>

2.8.2. Critical academic literature recommends changes to regulation of markets and industry



We see much more recognition of these matters in **critical academic literature**, which focuses on the harms and negative consequences of widespread data practices (there is also a significant critical grey literature, which we discuss in the next section). This literature often aims to reveal the power dynamics that underlie data usage. A lot of critical literature is not empirical, and so we do not review it here. Where critical empirical literature does exist, it often focuses on experiences of data practices, rather than understandings and perceptions of them. However, as noted above, empirical research about experiences can often tell us something, explicitly or implicitly, about understandings and perceptions.

For example, in ‘The downside of digital inclusion: expectations and experiences of privacy and surveillance among marginal internet users’, Seeta Peña Gangadharan draws on mixed methods research into digital inclusion programmes in the US, to demonstrate tensions between the promise of broadband opportunities and the asymmetrical ‘privacy-poor, surveillance-rich’ flows of information that broadband enables¹²⁹. In this article, she argues that there is a policy opportunity to guard against privacy and data profiling problems faced by marginalised individuals with low levels of technical knowledge. Thus she identifies the state and regulatory actors as responsible for implementing change. Likewise, also from the US so not discussed in detail here but worthy of note, a report on the collaborative Our Data Bodies project¹³⁰ argues that it is at the level of governance and regulatory systems that change is needed¹³¹. ‘Just’ modes of governance of data practices which guarantee human rights are required, the report concludes.

Virginia Eubanks’ book *Automating inequality: how high-tech tools profile, punish and police the poor* investigates the impacts of automated decision-making, data mining and predictive risk models on poor and working-class people in America, highlighting the impacts of data-driven discrimination on people living in poverty in the US¹³². The conclusion of the book focuses on what should change, with most proposed changes focusing on how to address poverty in US society, for example changing how poverty is conceived, implementing a universal basic income, and offering more generous public assistance. Like many critical scholars, she focuses on governance and regulation as the most significant domain in which change should happen. She also argues that a Hippocratic oath of non-harms for data scientists would be beneficial, thus also suggesting change to industry.

¹²⁹ Gangadharan, S P (2017) ‘The downside of digital inclusion: expectations and experiences of privacy and surveillance among marginal internet users,’ *New Media and Society*, 19(4):597-615. <https://doi.org/10.1177/1461444815614053>

¹³⁰ The Our Data Bodies project <https://www.odbproject.org/>

¹³¹ Petty, T, Saba, M, Lewis, T, Gangadharan, S P & Eubanks, V (2018) *Reclaiming our data: interim report*. https://www.odbproject.org/wp-content/uploads/2016/12/ODB.InterimReport.FINAL_7.16.2018.pdf

¹³² Eubanks, V (2018) *Automating inequality: how high-tech tools profile, punish and police the poor*. St Martin’s Press.

Although they have not carried out extensive research into perceptions and understandings of data practices, researchers at the Data Justice Lab at the University of Cardiff deserve a mention for their work in calling for change at a governmental level to address the unequal effects of datafication on marginalised communities. These proposed changes emerge from empirical research into data harms (that is, experiences rather than perceptions and understandings of data practices¹³³), research with social justice activists and analyses of increasingly widespread surveillance practices¹³⁴ and the use of data scoring in government¹³⁵. From these diverse projects, a set of recommendations for governmental and legislative change emerge, such as those outlined in the Lab's submission to the UN Special Rapporteur on extreme poverty and human rights¹³⁶. Recommendations in relation to data uses in government, include: transparency about uses of automated systems and data sharing; consultation with stakeholder groups; better oversight and regulation pertaining to data protection, discrimination, human rights; the possibility of opting out of data collection; data literacy training for people using automated systems focusing on issues like data quality, bias, errors and false positives.

2.8.3. Policy and practice-oriented grey literature recommends a range of changes, often involving state or governmental actors



As noted, much of the grey literature advocates for change, as its very purpose is to influence policy and practice, and much of it is also critical. Like the critical academic literature discussed in the previous section, some grey literature makes recommendations for changes to be enacted by states or government actors. The Big Brother Watch report, referenced above, argues that the ICO should do more to protect our personal data¹³⁷. The report claims that the public wants 'moves for more opting out of data sharing' and 'stronger enforcement of data protection laws by privacy regulators', as well as privacy by design as standard. Sopra Steria makes similar, government-focused suggestions for change, despite occupying a very different political position to Big Brother Watch¹³⁸.

Their report concludes that government needs to develop both policies and technical solutions relating to security, authentication and data storage, to ensure that citizens' data remains private. The report continues that without these measures, support for digital government services will be undermined.

¹³³ For example: Redden, J (2018) 'The harm that data do', *Scientific American*, 319(5).

<https://www.scientificamerican.com/article/the-harm-that-data-do/>

¹³⁴ For example: Dencik, L & Cable, J (2017) 'Digital Citizenship and Surveillance: The advent of surveillance realism: public opinion and activist responses to the Snowden leaks', *International Journal of Communication*, 11(2017):763-781.

<https://ijoc.org/index.php/ijoc/article/view/5524/1939>

¹³⁵ Dencik, L, Hintz, A, Redden, J & Warne, H (2018) *Data Scores as Governance: Investigating uses of citizen scoring in public services*, Cardiff University. <https://datajustice.files.wordpress.com/2018/12/data-scores-as-governance-project-report2.pdf>

¹³⁶ United Nations Special Rapporteur on extreme poverty and human rights to the United Kingdom of Great Britain and Northern Ireland from 5 to 16 November 2018 – Submissions:

<https://www.ohchr.org/EN/Issues/Poverty/Pages/UKVisitSubmissions.aspx>

¹³⁷ Big Brother Watch & ComRes (2015) *UK public research – online privacy*. <https://www.bigbrotherwatch.org.uk/wp-content/uploads/2015/03/Big-Brother-Watch-Polling-Results.pdf>

¹³⁸ Sopra Steria (2017) *The citizen view of the digital transformation of government*.

<https://www.soprasteria.co.uk/docs/librariesprovider41/White-Papers/sopra-steria-ipsos-digitaltransformation-of-govt.pdf?sfvrsn=0>

On the whole, the changes proposed in grey literature are diverse, rarely focusing on one particular actor. For example, the Ada Lovelace Institute's survey of public attitudes to facial recognition technology concludes that governments should introduce regulation to limit its use by the police and in schools and that facial recognition companies should not sell their products to these sectors, thus identifying change that should happen at the levels of the state and industry¹³⁹. The RSA's report on views on uses of AI in decision-making also identifies changes that these same actors should make, concluding that automated decision-making should not be used in public services or in the workplace¹⁴⁰.

Doteveryone have produced a number of reports in which they identify a number of changes¹⁴¹. As noted above, they conclude their 2018 report on digital attitudes by calling for investment in new forms of public engagement and education across the entire age range and nation. Their 2019 report *Engaging the public with responsible technology*, authored by Miller, advocates for a 'social contract for the digital age'. The report states that 'For a fair, inclusive and sustainable democratic digital society, people must be empowered to be resilient to technological change and enabled to play an active role in shaping a digital society'. To achieve this, the report makes three recommendations: 1) an independent technology regulator should implement a system of redress so the public can hold industry to account; 2) research should be commissioned into gaps in public advocacy; and 3) the government's media literacy strategy should build on the notion of public empowerment. Kennedy et al's report on public attitudes to data management models similarly makes three recommendations that cut across government and other actors¹⁴². The authors conclude that: 1) approaches to data management which give individuals control over their personal data and include oversight from regulatory bodies are needed; 2) investment of resources is required, to support technical development and public consultation; and 3) more research is needed to further advance understanding of public views about data management models.

We can see from the literature discussed in this section that a significant volume of work acknowledges that change is needed to the current landscape of data practices, although some literature does not focus on change. Systems-oriented literature, focusing on the interaction between the human and the computer, often identifies changes to be enacted in data-driven systems, such as clear communication, attending to complex privacy and context dynamics, and changes to design. On the whole, this research does not discuss the broader socio-economic and political contexts in which these systems are deployed and experienced, although Guberek et al and Bowyer et al are exceptions¹⁴³.

¹³⁹ Ada Lovelace Institute (2019) *Beyond face value - public attitudes to facial recognition technology*.

https://www.adalovelaceinstitute.org/wp-content/uploads/2019/09/Public-attitudes-to-facial-recognition-technology_v.FINAL_.pdf

¹⁴⁰ RSA / The Royal Society for the Encouragement of Arts, Manufactures and Commerce Royal Society of Arts (2018) *Artificial intelligence: real public engagement*. <https://www.thersa.org/discover/publications-and-articles/reports/artificial-intelligence-real-public-engagement>

¹⁴¹ The reports from Doteveryone are: Doteveryone (2018) *People, power, and technology: the 2018 digital understanding report*. http://understanding.doteveryone.org.uk/files/Doteveryone_PeoplePowerTechDigitalUnderstanding2018.pdf; Doteveryone (2018) *People, power and technology: the 2018 digital attitudes report*.

<https://www.doteveryone.org.uk/report/digital-attitudes/> and Doteveryone – Catherine Miller (2019) *Engaging the public with responsible technology: four principles and three requirements*. <https://doteveryone.org.uk/download/3225/>.

¹⁴² Kennedy, H, Hartman, T, Steedman, R & Jones, R (2020b) *UK public unhappy with the ways their data is managed*. <https://livingwithdata.org/project/wp-content/uploads/2020/03/Views-on-Data-Management-Full-Report.pdf>

¹⁴³ Guberek, T, McDonald, A, Simioni, S, Mhaidli, A, Toyama, K, Schaub, F (2018) 'Keeping a Low Profile?: Technology, Risk and Privacy among Undocumented Immigrants', *CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, April 2018 Paper No.: 114. <https://doi.org/10.1145/3173574.3173688> and

2.9. Attention to differences is growing, but more is needed

Not all data practices are the same, and people experience them from different social positions. Differences matter, something which literature on public understanding and perceptions of data practices is only just beginning to acknowledge. We outline a number of important differences in this section. First, **the effects of the widespread use of data-driven systems are not experienced equally by all**. Already socially unequal populations may feel the effects of data-related practices more negatively than others, and this in turn may affect how these systems and practices are perceived and understood. Literature on public understanding and perceptions of data practices has been slow to acknowledge the important role that social inequalities play. There are some exceptions to this, which we discuss below. Second, **differences in types of data or contexts of data use are also important**, and yet much of the literature does not address the implications of these kinds of differences. Again, we discuss exceptions to this claim. Third, **exploring how different degrees of knowledge and understanding influence perceptions** is another gap in the literature which is not often addressed. Fourth, some of the literature discussed here finds that people are concerned about data practices, but **how does concern about data practices differ in significance from other concerns that people have?** We discuss these four issues below.

2.9.1. Social inequalities and unequal experiences of the effects of data practices

Some research has highlighted that datafication impacts disadvantaged and marginalised groups in disproportionate, negative ways¹⁴⁴. Much of this research comes from the US, and it is often focused on experience rather than understanding and perception. Eubanks discusses the negative impacts of data-driven discrimination on people living in poverty, highlighting intersections between class and race¹⁴⁵. Research by Gangadharan, some of it in collaboration with Eubanks, also highlights the negative experiences of racially and socio-economically marginal populations in relation to datafication¹⁴⁶. As noted above, Gangadharan argues that this can lead these communities to exclude themselves from exposure to data-driven systems¹⁴⁷.

Kennedy et al's paper 'Approaching public perceptions of datafication through the lens of inequality', referenced above, focuses on perceptions and inequalities, as the title suggests¹⁴⁸. The authors found that inequalities relating to age, dis/ability, poverty and their intersections played a role in shaping perceptions of BBC data practices relating to users accessing digital services, in complex and diverse ways. Oman explored how creative industries workers feel about answering questions that seek to

¹⁴⁴ For example, Gangadharan, S P & Niklas, J (2019) 'Decentering technology in discourse on discrimination', *Information, Communication and Society*, 22(7), 882–899. <https://doi.org/10.1080/1369118X.2019.1593484> and Redden, J (2018) 'The harm that data do'. *Scientific American*, 319(5). <https://www.scientificamerican.com/article/the-harm-that-data-do/>

¹⁴⁵ Eubanks, V (2018) *Automating inequality: how high-tech tools profile, punish and police the poor*. St Martin's Press.

¹⁴⁶ For example: Gangadharan, S P (2017) 'The downside of digital inclusion: expectations and experiences of privacy and surveillance among marginal internet users', *New Media and Society*, 19(4):597-615.

<https://doi.org/10.1177/1461444815614053> and Petty, T, Saba, M, Lewis, T, Gangadharan, S P & Eubanks, V (2018) *Reclaiming our data: interim report*.

https://www.odbproject.org/wpcontent/uploads/2016/12/ODB.InterimReport.FINAL_7.16.2018.pdf

¹⁴⁷ Gangadharan, S P (2021) 'Digital exclusion: a politics of refusal', in H Landemore, R Reich & L Bernholz (eds) *Digital Technology and Democratic Theory*. University of Chicago Press. <http://eprints.lse.ac.uk/103076/>

¹⁴⁸ Kennedy, H, Steedman, R & Jones, R (2020a) 'Approaching public perceptions of datafication through the lens of inequality: a case study in public service media', *Information, Communication and Society*. <https://doi.org/10.1080/1369118X.2020.1736122>

identify their social class of origin alongside protected characteristics such as race and gender¹⁴⁹. She found that the recommended way of doing this – asking about parental or carer occupation when growing up – was perceived negatively by research participants and that there were differences in findings across age and geographical location.

Some other research also looks at demographic difference and social inequalities as they relate to perceptions of data-driven systems. In HCI literature, Guberek et al found concerns about identity theft, privacy and online harassment amongst undocumented migrants in the US, some of which related to uses of their personal data¹⁵⁰. Brown et al carried out a qualitative study of the perspectives of those affected by algorithmic decision-making in child welfare services, who are often economically disadvantaged¹⁵¹. Woodruff et al found that their participants from race- and class-based marginal communities in the US were not aware of ‘algorithmic unfairness’, but that participation in the project increased awareness and encouraged some participants to recognise they had experienced it¹⁵². On the whole, HCI research often uses researchers’ own students as a convenience sample, rarely acknowledges difference and inequality, and so produces findings that cannot be generalised to wider, diverse populations.

Some secondary analyses of large-scale datasets has focused on demographic differences. In secondary analysis of a pan-European dataset, Potoglou et al found that privacy-preferences vary across cultural and national settings, as well as across age, gender and education¹⁵³. Similarly Bergstrom (2015) looked at how online privacy concerns were distributed across groups in Sweden, finding that socio-economic differences have a bearing on privacy concerns.

Two desk-based reviews of existing research attend to demographic differences. The first, *Online Data Privacy from Attitudes to Action: an evidence review* covers both quantitative and qualitative empirical studies published by public, private and third-sector organisations, as well as academic papers that explore what people think and what they do in relation to online data privacy¹⁵⁴. The authors note that different groups have different attitudes to data privacy issues, yet most research looking at difference focuses on age, with much less examining gender (about which there are mixed findings) and socio-economic status. However, the different measures used to define socio-economic status across the different studies made it difficult for the authors to appraise and compare the limited evidence available.

¹⁴⁹ Oman, S (2019a) *Improving data practices to monitor inequality and introduce social mobility measures: a working paper*. The University of Sheffield. https://www.sheffield.ac.uk/polopoly_fs/1.867756!/file/MetricsWorkingPaper.pdf

¹⁵⁰ Guberek, T, McDonald, A, Simioni, S, Mhaidli, A, Toyama, K, Schaub, F (2018) ‘Keeping a Low Profile?: Technology, Risk and Privacy among Undocumented Immigrants’, *CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, April 2018 Paper No.: 114. <https://doi.org/10.1145/3173574.3173688>

¹⁵¹ Brown, A, Chouldechova, A, Putnam-Hornstein, E, Tobin, A & Vaithianathan, R (2019) ‘Toward algorithmic accountability in public services: a qualitative study of affected community perspectives on algorithmic decision-making in child welfare services’, *CHI '19: Proceedings of the 2019 Conference on Human Factors in Computing Systems*, May 2019. Paper No.: 41. <https://doi.org/10.1145/3290605.3300271>

¹⁵² Woodruff, A & Fox, S E, Rousso-Schindler S, & Warshaw, J (2018) ‘A qualitative exploration of perceptions of algorithmic fairness’, *CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, April 2018 Paper No.: 656. <https://doi.org/10.1145/3173574.3174230>

¹⁵³ Potoglou, D, Dunkerley, F, Patil, S & Robinson, N (2017) ‘Public preferences for internet surveillance, data retention and privacy enhancing services: evidence from a pan-European study’, *Computers in Human Behavior*, 75: 811-825. <https://doi.org/10.1016/j.chb.2017.06.007>

¹⁵⁴ Carnegie Trust - Carolyn Black, Lucy Setterfield & Rachel Warren (2019) *Online data privacy from attitudes to action: an evidence review*. https://d1ssu070pg2v9i.cloudfront.net/pex/carnegie_uk_trust/2018/08/03110116/Online-Data-Privacy-from-Attitudes-to-Action-CUKT.pdf

The second review, of public attitudes to health data by Understanding Patient Data, referenced above, also offers some analysis which acknowledges demographic differences¹⁵⁵. The review notes that the research suggests that younger people are generally more knowledgeable about and supportive of data practices, that people in lower socio-economic groups are less likely to see the benefits of data practices and more likely to feel powerless to address data-related harms, and that ethnic minority groups are slightly less likely than ethnic majority groups to trust that their data will remain secure.

Not all studies that address social inequalities discuss what they found about how these relate to understandings and perceptions of data practices. For example Molina et al explored how location, terms and conditions and VPN influence the extent to which people disclose private information in public places, considering gender, age, ethnicity and educational differences¹⁵⁶. However, they do not comment analytically on what they found in relation to these differences. Nor do all studies that address inequalities handle them sensitively. For example Weinberger et al studied 'sex differences' in attitudes towards online privacy and anonymity amongst a group of Israeli students¹⁵⁷. The conclusions they draw about gendered differences (which are not the same as sex differences) in technical knowledge are not situated within the context of sexism and inequality which may explain these differences.

In summary, there has been a gap in literature focusing on how social inequalities shape understanding and perceptions of data practices. This is beginning to change, but more research is needed to fully understand the role that inequality plays. When inequalities are addressed, this needs to be done sensitively, situating them in appropriate political and historical contexts.



¹⁵⁵ Understanding Patient Data (2018) *Public attitudes to patient data use: a summary of existing research*.

http://understandingpatientdata.org.uk/sites/default/files/2019-01/Public%20attitudes%20key%20themes_0.pdf

¹⁵⁶ Molina, M D, Gambino, A & Sundar, S S (2019) 'Online privacy in public places: how do location, terms and conditions and VPN influence disclosure?', *CHI '19: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, May 2019. Paper No.: LBW2616. <https://doi.org/10.1145/3290607.3312932>

¹⁵⁷ Weinberger, M, Zhitomirsky-Geffet, M & Bouhnik, D (2017) 'Sex differences in attitudes towards online privacy and anonymity among Israeli students with different technical backgrounds', *Information Research: An International Electronic Journal*, 22(4). <https://eric.ed.gov/?id=EJ1164311>

2.9.2. Differences in types of data and context of data use matter

Surveys exploring public trust in data practices, as well as other research discussed here, indicates that different contexts of data usage and different types of data play a role in shaping perceptions of data practices¹⁵⁸. Some surveys, however, do not differentiate between types of data and contexts of use in the questions they ask about public perceptions. For example, Digital Catapult's 2019 survey asks general questions about uses of personal data, and the ICO addresses people's concern about companies and organisations storing and using their personal information without giving specific examples¹⁵⁹. This reflects the particular focus of the surveys rather than inadequacies in survey design, but it makes it difficult to interpret findings across different types of data and different ways in which data might be used.

To address this concern, in a study already mentioned, Kennedy et al carried out a survey experiment about public attitudes to data management models, to explore the importance of different characteristics that may influence preferences with regard to data use and management¹⁶⁰. Respondents were asked to express preferences for profiles of data scenarios generated from a combination of factors: type of data (eg medical, financial, media consumption); who has control (eg individual, trustee, commercial organisation); what management arrangements mean for the individual (eg full control over what happens to data, know what data is held about them, by whom and what they do with it); uses and beneficiaries of the data (eg personal insights, generate profit, benefit society). The study found that the single most important factor influencing data management preferences is the locus of control over the data rather than types of data, uses and beneficiaries. This finding suggests that differences in types of data and context may be less important than other issues.

Qualitative studies usually focus on specific types of data or contexts of use, often social media, health or self-tracking. There are plenty of examples of such studies cited in this document¹⁶¹. The evidence

¹⁵⁸ For example: ICO / Information Commissioner's Office – Harris Interactive (2019) *Information rights strategic plan: trust and confidence*. <https://ico.org.uk/media/about-the-ico/documents/2615515/ico-trust-and-confidence-report-20190626.pdf>; ODI / Open Data Institute (2018) *Who do we trust with personal data?* <https://theodi.org/article/who-do-we-trust-with-personal-data-odi-commissioned-survey-reveals-most-and-least-trusted-sectors-across-europe/>; and Robinson, G & Dolk, H (2015) *Research update: public attitudes to data sharing in Northern Ireland*. Administrative Research Data Centre, Northern Ireland <https://www.ark.ac.uk/publications/updates/update108.pdf>.

¹⁵⁹ Information Commissioner's Office – Harris Interactive (2019) *Information rights strategic plan: trust and confidence*. <https://ico.org.uk/media/about-the-ico/documents/2615515/ico-trust-and-confidence-report-20190626.pdf> and Digital Catapult (2015) *Trust in personal data: A UK review*

¹⁶⁰ Kennedy, H, Hartman, T, Steedman, R & Jones, R (2020b) *UK public unhappy with the ways their data is managed*. <https://livingwithdata.org/project/wp-content/uploads/2020/03/Views-on-Data-Management-Full-Report.pdf>

¹⁶¹ Such as: Bowyer, A, Montague, K, Wheeler, S, McGovern, S, Lingam, R & Balaam, M (2018) 'Understanding the family perspective on the storage, sharing and handling of family civic data', *CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, April 2018. Paper No.: 136. <https://doi.org/10.1145/3173574.3173710>; Brown, A, Chouldechova, A, Putnam-Hornstein, E, Tobin, A & Vaithianathan, R (2019) 'Toward algorithmic accountability in public services: a qualitative study of affected community perspectives on algorithmic decision-making in child welfare services', *CHI '19: Proceedings of the 2019 Conference on Human Factors in Computing Systems*, May 2019. Paper No.: 41. <https://doi.org/10.1145/3290605.3300271>; Bucher, T (2017) 'The algorithmic imaginary: exploring the ordinary affects of Facebook algorithms', *Information, Communication & Society*. 20: 30-44. <https://doi.org/10.1080/1369118X.2016.1154086>; Kennedy, H, Steedman, R & Jones, R (2020a) 'Approaching public perceptions of datafication through the lens of inequality: a case study in public service media', *Information, Communication and Society*; Lupton, D (2017) "'It just gives me a bit of peace of mind': Australian women's use of digital media for pregnancy and early motherhood", *Societies* 7(3)25. <https://doi.org/10.3390/soc7030025>; Lupton, D (2019) 'Data mattering and self-tracking: what can personal data do?', *Continuum*, 34(1):1-13. <http://dx.doi.org/10.1080/10304312.2019.1691149>; Molina, M D, Gambino, A & Sundar, S S (2019) 'Online privacy in public places: how do location, terms and conditions and VPN influence disclosure?', *CHI '19: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, May 2019. Paper No.: LBW2616.

would seem to suggest that different uses of different types of data in different contexts are perceived in different ways by different people. This in turn suggests that research should differentiate along these lines, to avoid over-generalised conclusions.

2.9.3. Different degrees of knowledge and understanding affect perceptions

What is the relationship between different degrees of knowledge and understanding of data practices, and people's perceptions of them? This is an important question, for a number of reasons. Above, we indicate that some of the changes that are proposed as a result of empirical research relate to clearer communication about data practices or improved transparency about them. This implies a belief that better understanding of data practices might lead to changes in perceptions of them, and that users of data-driven systems may become more trusting of them if they understand the data that are gathered and the uses to which data are put. Data literacy initiatives also assume that better understanding will lead to more positive outcomes.

To date, there has been very little research exploring this relationship. A report for Doteveryone by Miller is one exception¹⁶². The author found that participants felt that there was a shared responsibility across government, regulators, industry, society as a whole and themselves as individuals to address data-related problems. She found that those who 'felt most confident about their own digital understanding [...] were most likely to highlight individual responsibility'. Focus group research by Kennedy and others is another exception¹⁶³. They found that people do not need to fully understand data practices in order to have strong feelings about them, drawing attention to the importance of feelings in the formation of views about data practices. A survey about views on data management models found that existing knowledge about issues relating to data was a significant predictor of model preferences¹⁶⁴. More knowledgeable respondents preferred approaches that offered more control and/or oversight over personal data by a regulatory public body than less knowledgeable

<https://doi.org/10.1145/3290607.3312932>; Pink, S & Fors, V (2017) 'Being in a mediated world: self-tracking and the mind-body-environment', *Cultural Geographies*, 24(3)375-388. <https://doi.org/10.1177/1474474016684127>; Pink, S, Sumartojo, S, Lupton, D & La Bond, C H (2017) 'Mundane data: the routines, contingencies and accomplishments of digital living', *Big Data & Society*, 4(1). <https://doi.org/10.1177/2053951717700924>; Pridmore, J & Mols, A (2020) 'Personal choices and situated data: Privacy negotiations and the acceptance of household Intelligent Personal Assistants', *Big Data & Society*, 7(1). <https://doi.org/10.1177/2053951719891748>; Rendina, H J & Mustanski, B (2018) 'Privacy, trust, and data sharing in web-based and mobile research: participant perspectives in a large nationwide sample of men who have sex with men in the united states', *Journal of Medical Internet Research*, 20(7). <https://doi.org/10.2196/jmir.9019>; Ruckenstein, M S (2017) 'Keeping data alive: talking DTC genetic testing', *Information, Communication and Society*, 20(7):1024-1039. <https://doi.org/10.1080/1369118X.2016.1203975>; Ruckenstein, M & Pantzar, M (2015) 'Datafied Life: techno-anthropology as a site for exploration and experimentation', *Techné: Research in Philosophy and Technology*, 19(2):191-210. <https://doi.org/10.13140/RG.2.1.2553.7762>; Weiner, K, Will, C, Henwood, F, Williams, R (2020) 'Everyday curation? Attending to data, records and record keeping in the practices of self-monitoring', *Big Data & Society* 7(1). <https://doi.org/10.1177/2053951720918275>; and Yamashita, N Kuzuoka, H, Kudo, T, Hirata, K, Aramaki, E & Hattori, K (2018) 'How information sharing about care recipients by family caregivers impacts family communication', *CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, April 2018, Paper No.: 222. <https://doi.org/10.1145/3173574.3173796>

¹⁶² Doteveryone – Catherine Miller (2019) Engaging the public with responsible technology: four principles and three requirements. <https://doteveryone.org.uk/download/3225/>

¹⁶³ Kennedy, H, Steedman, R & Jones, R (2020a) 'Approaching public perceptions of datafication through the lens of inequality: a case study in public service media', *Information, Communication and Society*. <https://doi.org/10.1080/1369118X.2020.1736122> and Steedman, R, Kennedy, H & Jones, R (2020) 'Complex ecologies of trust in data practices and data-driven systems' *Information, Communication and Society*. <https://doi.org/10.1080/1369118X.2020.1748090>

¹⁶⁴ Kennedy, H, Hartman, T, Steedman, R & Jones, R (2020b) *UK public unhappy with the ways their data is managed*. <https://livingwithdata.org/project/wp-content/uploads/2020/03/Views-on-Data-Management-Full-Report.pdf>

We do not know how concerns about data practices differ in significance from other concerns that research participants and respondents may have, because there is little research which undertakes such comparison. One exception is the literature on trust, some of which explores how trust compares across sectors or institutions¹⁶⁶. Some researchers suggest that to understand experiences of and attitudes towards data practices, we should decentre data – in other words, we should situate data practices in relation to other social practices, to acknowledge that experiences of and attitudes towards data practice depend on other things. Gangadharan and Niklas propose that decentring data is necessary in order to understand the relationship between datafication and inequalities, and how they both shape each other¹⁶⁷. Similarly, Dencik argues that situating data practices in relation to other social practices makes it possible to uncover why they matter in the ways that they do¹⁶⁸. In short, asking questions about data may not be the best way to understand people’s perceptions, thoughts and feelings about data; our research should, perhaps, start elsewhere. Given the opacity and black boxing of many data practices and the fact that the role of datafication in lives and experiences is not always transparent, it is possible that important questions for citizens and publics are not ‘what happens to my personal data’ but rather ‘do I get access to the services to which I am entitled’.

These points raise a number of methodological questions for future research. Can we talk about data practices in isolation from other issues, or do we need always to contextualise them? Our research, *Living With Data*, is based on an assumption that the latter is necessary. But, as noted above, once we do, how generalisable are our findings, and how useful for policy-making? Should we ask respondents and participants to rank or otherwise compare their concerns, thus situating concern about what happens to their data in the context of other concerns? Should we research public perceptions of data practices in a way that does not start with data practices, but rather which starts elsewhere? What methods enable us to explore the relationship between what people say and what they do? We need research which attends to these issues, and which makes it possible to access how people feel about data practices, something they may not have previously thought about or might feel they do not care about.

In summary, differences matter when it comes to public perceptions of data practices. This is particularly important in relation to: social inequalities; differences in types of data or contexts of data use; how degrees of knowledge and understanding influence perceptions; how concerns about data practices differ in significance from other concerns. Research needs to attend to differences, being mindful of producing useful evidence for policy and informing practice.

¹⁶⁶ For example Edelman (2018) *Edelman Trust Barometer 2018, UK Findings*. <https://www.edelman.co.uk/magazine/posts/edelman-trust-barometer-2018/>; ICO / Information Commissioner’s Office – Harris Interactive (2019) *Information rights strategic plan: trust and confidence*. <https://ico.org.uk/media/about-the-ico/documents/2615515/ico-trust-and-confidence-report-20190626.pdf>; ODI / Open Data Institute (2018) *Who do we trust with personal data?* <https://theodi.org/article/who-do-we-trust-with-personal-data-odi-commissioned-survey-reveals-most-and-least-trusted-sectors-across-europe/>; and Robinson, G & Dolk, H (2015) *Research update: public attitudes to data sharing in Northern Ireland*. Administrative Research Data Centre, Northern Ireland <https://www.ark.ac.uk/publications/updates/update108.pdf>

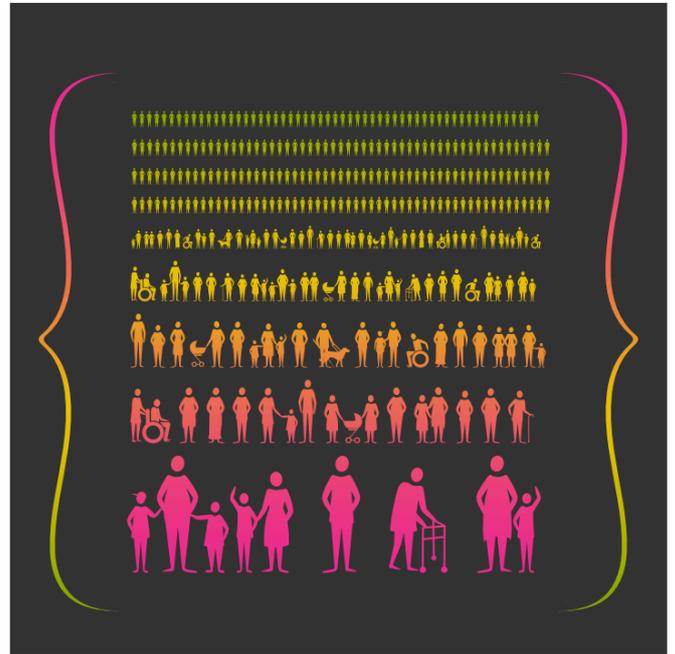
¹⁶⁷ Gangadharan, S P & Niklas, J (2019) ‘Decentering technology in discourse on discrimination’, *Information, Communication and Society*, 22(7), 882–899. <https://doi.org/10.1080/1369118X.2019.1593484>

¹⁶⁸ Dencik, L (2019) ‘Situating practices in datafication — from above and below’, in H Stephansen and E Treré (eds) *Citizen media and practice*. Routledge.

2.10 Methods and framing shape findings

2.10.1. Methodological challenges

Research investigating how people perceive something and what they know and understand about it can struggle to disaggregate these aspects of human experience. What people say they think in research contexts is unlikely to be consistent (people's thoughts, opinions and behaviours can change regularly) or may not reflect what they actually do in their everyday lives. The wording of a survey question, the effect of interviewer presence, the framing of an issue and the impact of others in a focus group setting can all affect responses to research questions. While these are well-known issues in social research, they are not widely acknowledged in the research discussed here. Methodological challenges like these have implications for what is claimed about public understanding and perception of data practices.



One particular issue that we found in our review was that methods were not always described well. In some cases, what was described as a survey approach did not, in fact, use a survey or draw from survey data. In other cases, limitations of samples, relating either to sample size or sampling design, were not acknowledged. Often, research does not draw attention to the limits of the sample in relation to the claims made. This makes it difficult to establish whether research was carried out rigorously and how reliable resulting claims are.

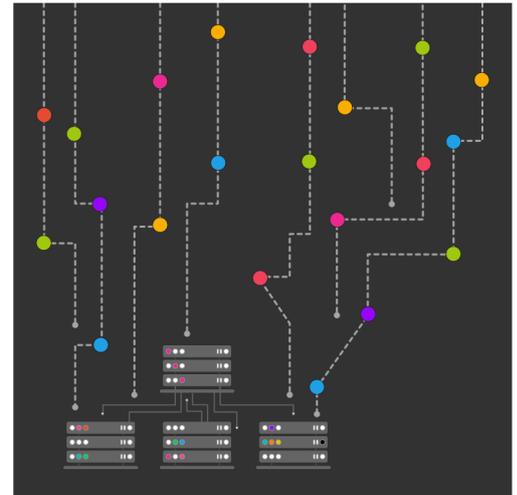
Another issue, especially with surveys, is that questions are sometimes leading. One example is the survey on which the HEPI report is based, which asked 'Are you concerned by rumours that some universities are facing data security issues?' A question articulated like this will produce particular findings¹⁶⁹. Some qualitative, participatory methods also have effects on participants and on findings. These are often positive and intended, but nonetheless, reflection is needed on what this means for findings. Examples include Woodruff et al's workshops and interviews with people from race- and class-based marginal communities in the US, participation in which led participants to realise that they had experienced 'algorithmic unfairness'¹⁷⁰.

¹⁶⁹ HEPI (2019) *Students or data subjects?* <https://www.hepi.ac.uk/wp-content/uploads/2019/12/Students-or-data-subjects-Report-122-Web-FINAL.pdf>

¹⁷⁰ Woodruff, A & Fox, S E, Rousso-Schindler, S & Warshaw, J (2018) 'A qualitative exploration of perceptions of algorithmic fairness', *CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, April 2018 Paper No.: 656. <https://doi.org/10.1145/3173574.3174230>. See also Pybus, J, Coté, M & Blanke, T (2015) 'Hacking the social life of big data: a data literacy framework', *Big Data & Society*, 2(2). <https://doi.org/10.1177/2053951715616649>.

2.10.2. Different entry points and framings shape findings

Not all of the studies we discuss here set out to find out specifically about public perceptions of data practices, of what happens to their personal data, or other data-specific issues. Some focus on other, related issues. (Exceptions include Kennedy (various), Lupton (various), Pink (various), Ruckenstein (various), who, in collaboration with others, generally put thoughts and feelings about data and related matter at the centre of the research¹⁷¹. Bowyer et al (2018), Dolin et al (2018) and Fiore-Gartland and Neff (2015) also do¹⁷²). Some research draws conclusions not directly related to what was initially researched. In some cases, research undertaken to explore one phenomenon also found things of interest about how people experience data and related practices. Not all literature made it easy to establish whether this was the case, or to identify the analytical work that had been undertaken that meant that empirical research on one issue led to findings or conclusions about perceptions of data practices. What's more, people might respond to a question about a particular data practice with their thoughts about a related phenomenon, such as a recent high profile data breach or similar story in the media. This issue could be more explicitly acknowledged and addressed in the literature.



Historically, a lot of research relevant to this review has been framed in terms of privacy and surveillance. Examples of research included in our review that focuses primarily on privacy includes Bergstrom, who surveyed online privacy concerns, and Fiesler and Hallinan and Rendina and Mustanski, mentioned above¹⁷³. Examples that focus on surveillance include Dencik and Cable, who

¹⁷¹ See: Kennedy, H, Elgesem, D, & Miguel, C (2015) 'On fairness: user perspectives on social media data mining', *Convergence*, 23(3): 270–288. <https://doi.org/10.1177/1354856515592507>; Kennedy, H & Hill, R (2017) 'The feeling of numbers: emotions in everyday engagements with data and their visualisation', *Sociology*, 52(4): 830-848. <https://doi.org/10.1177/0038038516674675>; Kennedy, H, Steedman, R & Jones, R (2020a) 'Approaching public perceptions of datafication through the lens of inequality: a case study in public service media', *Information, Communication and Society*. <https://doi.org/10.1080/1369118X.2020.1736122>; Lupton, D (2017) "'It just gives me a bit of peace of mind': Australian women's use of digital media for pregnancy and early motherhood", *Societies* 7(3)25. <https://doi.org/10.3390/soc7030025>; Lupton, D (2019) 'Data mattering and self-tracking: what can personal data do?', *Continuum*, 34(1):1-13. <http://dx.doi.org/10.1080/10304312.2019.1691149>; Pink, S & Fors, V (2017) 'Being in a mediated world: self-tracking and the mind–body–environment', *Cultural Geographies*. 24(3)375-388. <https://doi.org/10.1177/1474474016684127>; Pink, S, Sumartojo, S, Lupton, D & La Bond, C H (2017) 'Mundane data: the routines, contingencies and accomplishments of digital living', *Big Data & Society*, 4(1). <https://doi.org/10.1177/2053951717700924>; Ruckenstein, M S (2017) 'Keeping data alive: talking DTC genetic testing', *Information, Communication and Society*, 20(7):1024-1039. <https://doi.org/10.1080/1369118X.2016.1203975>; Ruckenstein, M & Pantzar, M (2015) 'Datafied Life: techno-anthropology as a site for exploration and experimentation', *Techné: Research in Philosophy and Technology*, 19(2):191-210. <https://doi.org/10.13140/RG.2.1.2553.7762>

¹⁷² Bowyer, A, Montague, K, Wheeler, S, McGovern, S, Lingam, R & Balaam, M (2018) 'Understanding the family perspective on the storage, sharing and handling of family civic data', *CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, April 2018. Paper No.: 136. <https://doi.org/10.1145/3173574.3173710>; Dolin, C, Weinshel, B, Shan, S, Hahn, C, Choi, E, Mazurek, M & Blase, U (2018) 'Unpacking perceptions of data-driven inferences underlying online targeting and personalization', *CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, April 2018. Paper No.: 493. <https://doi.org/10.1145/3173574.3174067>; Fiore-Gartland, B & Neff, G (2015) 'Communication, mediation and the expectations of data: data valences across health and wellness communities', *International Journal of Communication*, 9. <http://ijoc.org/index.php/ijoc/article/view/2830>

¹⁷³ Bergstrom, A (2015) 'Online privacy concerns: a broad approach to understanding the concerns of different groups for different uses', *Computers In Human Behavior*, 53:419-426. <https://doi.org/10.1016/j.chb.2015.07.025>; Fiesler, C &

use the term surveillance realism to understand public opinion and activist responses to the Snowden leaks; Ellis's project on everyday surveillance and Potoglou et al's secondary analysis of a European dataset on internet surveillance and data retention¹⁷⁴. Some researchers examine perceptions of data practices alongside these phenomena. For example, Fiesler and Hallinan explore public reactions to online data sharing and privacy controversies, Rendina and Mustanski also research data sharing and privacy, and Turow (various, with others) sometimes focuses on data sharing, sometimes on privacy, sometimes on surveillance, and sometimes on a combination of these things¹⁷⁵.

More recently, algorithms have captured the imagination of the public and therefore researchers¹⁷⁶. Even more recently, there have been reports on public perceptions of AI, such as the work of the RSA in the UK, and the AI NOW research institute in the US¹⁷⁷. Research on perceptions of particular types of AI, such as machine learning¹⁷⁸, or household Intelligent Personal Assistants¹⁷⁹, is also emerging.

Hallinan, B (2018) 'We are the product: public reactions to online data sharing and privacy controversies in the media', *CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, April 2018 Paper No.: 53.

<https://doi.org/10.1145/3173574.3173627> and Dencik, L & Cable, J (2017) 'Digital Citizenship and Surveillance: The advent of surveillance realism: public opinion and activist responses to the Snowden leaks', *International Journal of Communication*, 11(2017):763-781. <https://ijoc.org/index.php/ijoc/article/view/5524/1939>

¹⁷⁴ Dencik, L & Cable, J (2017) 'Digital Citizenship and Surveillance: The advent of surveillance realism: public opinion and activist responses to the Snowden leaks', *International Journal of Communication*, 11(2017):763-781.

<https://ijoc.org/index.php/ijoc/article/view/5524/1939>; Ellis, D (2019) 'Techno-securitisation of everyday life and cultures of surveillance-apatheia', *Science as Culture*, 29(1), 11-29. <https://doi.org/10.1080/09505431.2018.1561660>; Potoglou, D, Dunkerley, F, Patil, S & Robinson, N (2017) 'Public preferences for internet surveillance, data retention and privacy enhancing services: evidence from a pan-European study', *Computers in Human Behavior*, 75: 811-825. <https://doi.org/10.1016/j.chb.2017.06.007>.

¹⁷⁵ Fiesler, C & Hallinan, B (2018) 'We are the product: public reactions to online data sharing and privacy controversies in the media', *CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, April 2018 Paper No.: 53.

<https://doi.org/10.1145/3173574.3173627>; Rendina, H J & Mustanski, B (2018) 'Privacy, trust, and data sharing in web-based and mobile research: participant perspectives in a large nationwide sample of men who have sex with men in the united states', *Journal of Medical Internet Research*, 20(7). <https://doi.org/10.2196/jmir.9019>; Turow, J, Hennessy, M, Draper, N, Akanbi, O & Virgilio, D (2018) *Divided we feel: partisan politics drive American's emotions regarding surveillance of low-income populations*. https://repository.upenn.edu/cgi/viewcontent.cgi?article=1563&context=asc_papers; Turow, J, Hennessy, M & Draper, N (2015) *The trade off fallacy*. https://www.asc.upenn.edu/sites/default/files/TradeoffFallacy_1.pdf

¹⁷⁶ For example; Alvarado, O & Waern, A (2018) 'Towards algorithmic experience: initial efforts for social media contexts', *CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, April 2018. Paper No.: 286.

<https://doi.org/10.1145/3173574.3173860>; Bucher, T (2017) 'The algorithmic imaginary: exploring the ordinary affects of Facebook algorithms', *Information, Communication & Society*. 20: 30-44. <https://doi.org/10.1080/1369118X.2016.1154086>;

Colbjørnsen, T (2018) 'My algorithm: user perceptions of algorithmic recommendations in cultural contexts', in Andrea L Guzman (ed) *Human-machine communication: rethinking communication, technology, and ourselves*. Peter Lang; Eslami, M, Rickman, A, Vaccaro, K, Aleyasen, A, Vuong, A, Karahalios, K, Hamilton, K & Sandvig, C (2015) 'I always assumed that I wasn't really that close to [her]: reasoning about invisible algorithms in news feeds', *CHI '15: Proceedings of the 2015 CHI Conference on Human Factors in Computing Systems*, April 2015.

<https://doi.org/10.1145/2702123.2702556>; Kizilcec, R F (2016) 'How much information? effects of transparency on trust in an algorithmic interface', *CHI '16: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, May 2016.

<https://doi.org/10.1145/2858036.2858402>; Magalhães, J C (2018) 'Do algorithms shape character? considering algorithmic ethical subjectivation', *Social Media + Society*, 4(2).

<https://doi.org/10.1177/2056305118768301> and Woodruff, A & Fox, S E, Rousso-Schindler S, & Warshaw, J (2018) 'A qualitative exploration of perceptions of algorithmic fairness', *CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, April 2018 Paper No.: 656.

<https://doi.org/10.1145/3173574.3174230>

¹⁷⁷ RSA / The Royal Society for the Encouragement of Arts, Manufactures and Commerce Royal Society of Arts - The Forum for Ethical AI (2019) *Democratising decisions about technology: a toolkit*. <https://www.thersa.org/discover/publications-and-articles/reports/democratising-decisions-technology-toolkit>;

RSA / The Royal Society for the Encouragement of Arts, Manufactures and Commerce Royal Society of Arts (2018) *Artificial intelligence: real public engagement*.

<https://www.thersa.org/discover/publications-and-articles/reports/artificial-intelligence-real-public-engagement>.

The AI NOW research institute in the US <https://ainowinstitute.org/>

¹⁷⁸ Harrison, G, Hanson, J, Jacinto, C, Ramirez, J & Blase, U (2020) 'An empirical study on the perceived fairness of realistic, imperfect machine learning models', *ACM FAT* conference 2020*. <https://dl.acm.org/doi/abs/10.1145/3351095.3372831>

These different entry points into researching public perceptions of data-related matters are important, because they have implications for the findings of the research that is undertaken. Framing research as being about privacy or surveillance, for example, means that the research will find things about privacy and surveillance, even if these are not primary concerns for participants. Kennedy et al attempted to address this challenge by approaching their early research into attitudes to social media data mining with no framing, inviting participants to say what they thought about the topic and so enabling themes to emerge from participants' own words¹⁸⁰. Diverse entry points can also make it hard to be clear about whether findings relate to perceptions of data practices or other issues.

2.10.3. Context shapes findings

The context in which research is undertaken affects how research participants respond to questions about their perceptions of data practices. Some research acknowledges the impact of events such as the introduction of GDPR or the Facebook/Cambridge Analytica scandal on what people feel and say about data issues¹⁸¹. However, some research does not. Our report does not reflect on this point in relation to all of the research we have reviewed, in the interests of brevity, but we note here that this is important. We finished writing this report during the Covid-19 pandemic in 2020, a context in which new data practices were introduced, from contact tracing to symptom tracking and the sharing of health data with organisations like supermarkets, things considered not in the public interest prior to the global health crisis. As noted in the introduction, this context will affect peoples understanding and perceptions of data practices, something that will need to be captured in future research.

Another contextual factor that matters is the perspective of the researcher or organisation commissioning the research. This may lead to particular research priorities, to asking particular questions, or to analysing results in particular ways. This methodological challenge is not easily resolved, and it is something we have been aware of in relation to the methods we used to produce this report, which we discuss in detail in section 3.



¹⁷⁹ Pridmore, J & Mols, A (2020) 'Personal choices and situated data: Privacy negotiations and the acceptance of household Intelligent Personal Assistants', *Big Data & Society*, 7(1). <https://doi.org/10.1177/2053951719891748>

¹⁸⁰ Kennedy, H, Elgesem, D, & Miguel, C (2015) 'On fairness: user perspectives on social media data mining', *Convergence*, 23(3): 270–288. <https://doi.org/10.1177/1354856515592507>

¹⁸¹ For example: Oman, S (2019a) *Improving data practices to monitor inequality and introduce social mobility measures: a working paper*, The University of Sheffield. https://www.sheffield.ac.uk/polopoly_fs/1.867756!/file/MetricsWorkingPaper.pdf

3. Conclusions

3.1. **People have some knowledge and understanding of data practices.** Findings from both quantitative and qualitative studies suggest that people's knowledge about what happens to their data is mixed. Knowledge and understanding of data practices are varied. Some people understand some data practices, and they interpret their understanding in different ways which, in turn, leads to different levels of concern.

3.2. There is extensive evidence across all literature that **people are concerned about data practices.** This is an important finding that emerges from a lot of diverse research. Policy-makers and data practitioners need to be willing to address these concerns.

3.3. But this is not the whole picture: **people are not *only* concerned.** They find ways to negotiate, embed or resist data practices in their everyday lives. People often hold contradictory views about data practices, recognising their benefits and feeling concerned about potential harms at the same time. In some contexts, people feel they have some agency around their data, especially personal data that they can easily access, such as health self-tracking data.

3.4. **Emotions play an important role in understandings and perceptions of data practices.** Qualitative studies recognise the significant role that feelings play in perceptions of data practices. How emotions matter varies across demographic groups. Emotions inform and are informed by reason and rational thinking, so they need to be understood as an important element in public understanding and perceptions of data practices.

3.5. People **trust** some sectors with their data more than others. The relationship between trust in institutions in general and trust in institutions' data practices is complicated, and findings are contradictory. For example, people trust the police with their data but they do not trust automated, data-driven decision-making in criminal justice practices. Qualitative research suggests that trust and distrust in data practices are not experienced separately. Trust and distrust are context-dependent, and sometimes trust and distrust co-exist. Sometimes, distrust is appropriate, because trust needs to be earned. Research suggests that people believe that better communication and the existence of safeguards, accountability and transparency would make organisations more trustworthy.

3.6. Some responses to data practices are seen as **apathy or acceptance.** But responses need to be understood in a context in which people feel unable to control the flows of their personal data, even if they want to. Some researchers see this as **digital resignation**, not apathy or acceptance. Using data-driven services does not mean that people accept data practices – we also saw that people are concerned about them and that people hold contradictory views about them. In addition, sometimes people **resist** data practices, and there are various ways in which they do this.

3.7. Most research finds dissatisfaction with the current ways in which data is used and managed, and a desire for this to change. A number of characteristics of **changed, fairer data practices** have been identified. These include:

- Honesty, transparency and genuine dialogue with the public;
- Regulation, enforcing compliance, the existence of safeguards and accountability, and the right to redress;
- Personal control.

As with people's concerns and degrees of trust, contexts of data use influence people's thinking about whether they are fair or not.

3.8. Because people are concerned about existing data practices, because trust in them is limited, and because people have views about how data practices could be fairer, clearly, change is needed. Views about **what needs to change** are influenced by research discipline and researcher perspective.

Systems-focused literature recommends changes in system design, foregrounding actions that could be taken by technology providers, such as clear communication, privacy by design, and attending to complex privacy and contextual dynamics. **Critical academic literature** and **some policy and practice-oriented grey literature** identify governments as the most important enablers of change. Changes proposed include the introduction and enforcement of just policies to guard against harms, especially to marginalised populations; greater transparency about existing data practices, state commitment to public consultation and education as a means to citizen empowerment. The possibility of opting out, training for users of data-driven systems, data security and privacy-by design are proposed changes which involve both state and industry.

Some of these proposals are already provided for under GDPR, which is still in force in the UK at the time of writing. This raises questions for future research, about whether existing arrangements are perceived as fair but in need of better enforcement, or whether fairer regulation is seen as needed.

3.9. **Differences matter** when it comes to public perceptions of data practices. Not all data practices are the same, and people experience them from different social positions. Research is beginning to pay attention to differences, but more understanding of them is needed. Differences matter in relation to types of data or contexts of data use, how degrees of knowledge and understanding influence perceptions and how concern about data practices differs in significance from other concerns that people have. Most importantly, social inequalities play a major role in shaping people's experiences of data practices, and therefore their understanding and perceptions of them.

3.10. **How research is conducted makes a difference to what it finds.** Research methods, the questions asked, how findings are interpreted and presented, the disciplinary background and the political orientation of researchers all play a role in shaping findings that emerge and claims that are made in the research we reviewed. Decision-making based on the evidence we have reviewed should be alert to this fact.

The above conclusions are drawn from the different sections of this report. In addition, three **overarching conclusions** also emerge. First, data matters are **human matters**. This means that data-related governance and decision-making needs to be human-centric. It needs to start with the experiences and perceptions of the people who are affected by data practices. Second, in public understanding and perceptions of data practices, **context matters**. Who gathers data, what and whose data is gathered, for what purpose and with what effects, influences people's attitudes. Third, as stated above, **inequalities matter**. Social inequalities influence knowledge and understanding, concerns, degree of trust and feelings about data practices.

4. Recommendations

Our review of existing research suggests that the following should happen:

4.1. For policy-makers

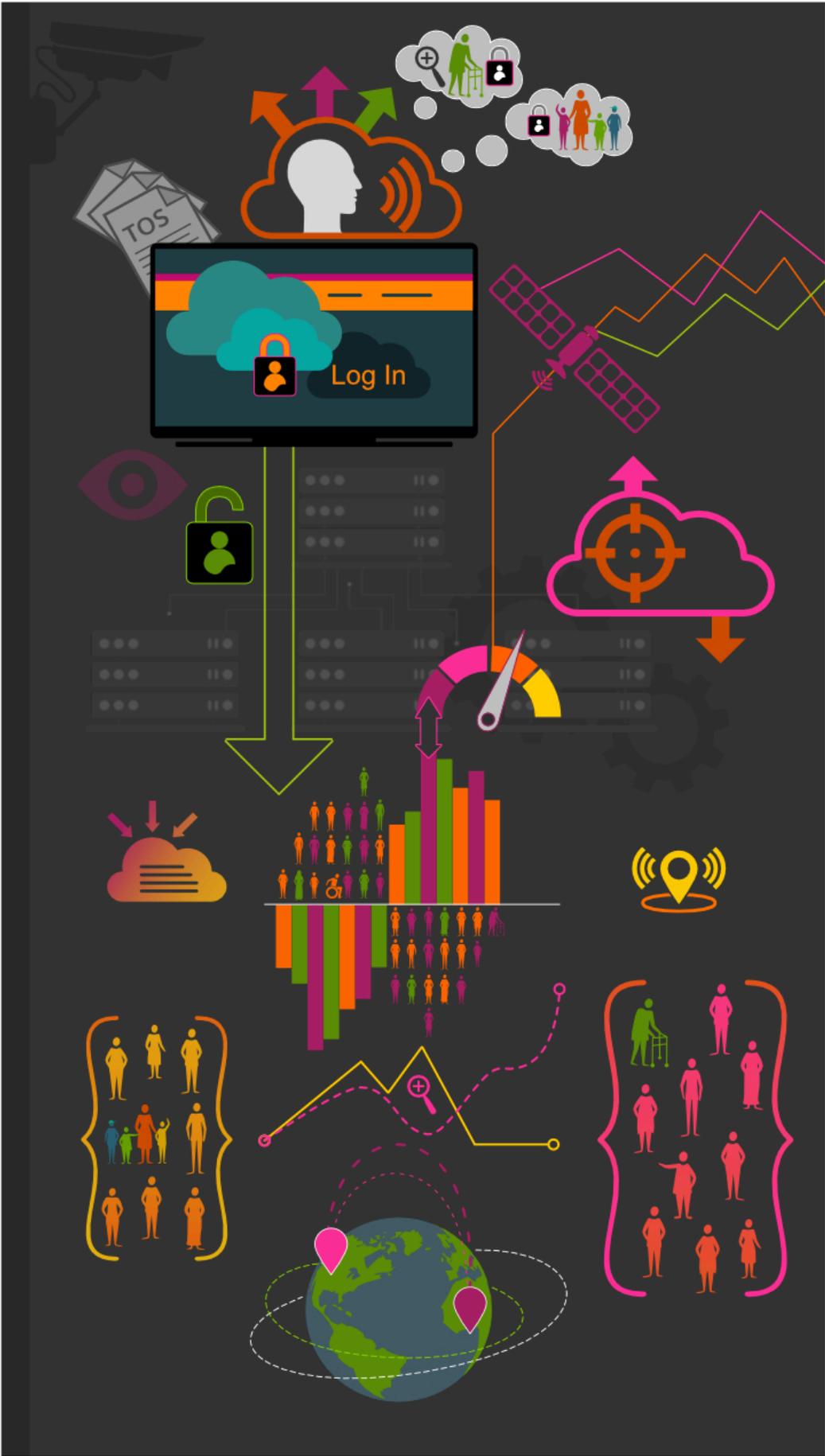
- Look beyond headline findings about public perceptions of data practices. Policy-makers need to engage with the breadth of evidence available across academic disciplines, policy domains and the third sector, to understand that methods and framing shape research findings and to be wary of taking findings on face value.
- Because public experiences of and attitudes towards data practices relate to issues such as social inequality and expectations about fairness, data-related policy-making needs to be 'joined-up' with areas such as equalities and government transparency.
- Focus not only on the individual responsibilities of citizens to protect themselves from data-related harms, but also on what government and practitioners can do to improve data practices, given the important role that the state and industry play in a fairer data future.
- Clear, accessible communication that facilitates lay understanding and transparency about data practices is needed.

4.2. For practitioners

- Be transparent, clear and honest about data practices.
- Enter into a genuine dialogue with the public about what they consider to be fair data practices, earning public trust and listening to the voices of minority / marginalised groups.
- Educate data practitioners to develop understanding of the potential negative consequences of data-driven decisions and broader data practices, especially for minority / marginalised groups.
- Understand that people using data-driven systems are not all the same, and that social inequalities mean that data practices impact more negatively on minority / marginalised groups. Develop expertise and understanding of the possible negative consequences of data-driven systems, including issues relating to inequality

4.3. For researchers

- Carry out further research to advance understanding of public perceptions of what fair data practices might look like and what might need to change to make them fairer, to balance expert views with public views.
- Carry out further research to advance understanding of how social inequalities influence knowledge and understanding, concerns, degree of trust and feelings about data practices.
- Recognise the importance of factors that are not easily captured by quantitative methods. These include: the co-existence of contradictory views; the importance of emotions in shaping responses; people's everyday, mundane experiences of data; and how people negotiate or resist data practices.
- Be detailed and transparent about context, methods and framings, making it clear how these have shaped findings.
- Avoid categorical claims based on headline findings about public perceptions of data practices.



5. Review methodology

5.1. Introduction to review methodology

We undertook a detailed review of previous literature that attempted to address aspects related to our two research questions: *What do different people know and feel about data-related practices in different domains of everyday life?* and *What do fair data practices look like, from non-experts' perspectives?* We undertook a number of searches to establish lists of relevant literature to enable us to meet our aims. There were two approaches: a) a systematic search of online citation databases, using multiple keywords relating to what people think and how they feel about data practices and what happens to their personal data; and b) a manual, snowball search that began with grey and academic literature with which we were already familiar and which then snowballed out (for example by searching bibliographies and relevant websites, observing Twitter discussions and building on word-of-mouth recommendations). The literature we identified through these processes was reviewed iteratively according to one principle inclusion criterion: **does it report empirical research about what people think and how they feel about data practices and what happens to their personal data?** We outline our various approaches in more detail below.

5.2. The systematic search of online databases

Starting with our two research questions, we identified a series of key terms under three notional headings: who; how; and what (see Table 1).

Who	How	What
People	Experiences	Data
Ordinary people	Perception	Big data
Citizens	Attitudes	Personal data
The public	Opinions	Datafication
Publics	Views	Data mining
General public	Perspectives	Data analytics
Users	Feelings	Data uses
Non-experts	Emotions	Data gathering / collection / harnessing
	Affect	AI
	Think	Automation
	Thought	Automated-decision making
	Accept	Machine Learning
		Automated
		Data-driven
		Algorithms
		Algorithmic (bias, culture)
		Privacy
		Surveillance
		Trust

Table 1: Key concepts for initial literature searching

We undertook initial searches using these terms, which aimed to identify useful ways to narrow down our parameters (for example by date, location, methodology, discipline or publication type), rather than starting with tighter parameters and excluding potentially relevant literature. This was important given that systematic reviews can overlook particular kinds of literature. We carried out two waves of searches, which we discuss below.

5.2.1. Wave 1: trialling key terms and databases

A full systematic search was conducted using Boolean strings made up of the terms listed in Table 1. Three search portals were identified for searching, and initial scoping searches looked for all literature available, irrespective of date, location, methodology, discipline or publication type. Bramer et al suggest 'optimal searches in systematic reviews should search at least Embase, MEDLINE, Web of Science, and Google Scholar as a minimum requirement to guarantee adequate and efficient coverage'¹⁸². Taking this as our starting point, our systematic search process began with the Web of Science Core Collection and Google Scholar, excluding Embase and MEDLINE because of their medical focus, and replacing them with the International Bibliography of the Social Sciences (IBSS). The results from these trials were collated and shared with the research team. Evaluating the value and limits of different searches for different purposes the results of each trial and the user experience of the search portals, we then decided to focus on Web of Science¹⁸³.

5.2.2. Wave 2: fine-tuning key terms and building the search

Our search terms were refined to reduce the number of results and improve their fit with our research questions. First, we removed terms relating to people and citizens – this involved removing the whole of the first 'who' column. We did this because these terms produced many results that were unrelated to our focus. Furthermore, terms like 'attitudes' and 'opinions' incorporate people without needing to include these terms in our searches. Second, we truncated words to allow for variations of words – for example, *experienc** includes suffixes such as *ing / ed / es*. There was no way of filtering for only empirical research, so this would have to be applied subsequent to the search.

Using Boolean search strings, such as *AND/OR*, we built a series of smaller searches. For example:

- **EXAMPLE 1:** (Data NEAR/1 (Big OR Analytics OR Mining))
- **EXAMPLE 2:** ((*experience* OR perception* OR attitude* OR opinion* OR view* OR perspective* OR feeling* OR thought* OR emotion* OR affect* OR think* OR accept**) NEAR/2 (of OR about OR toward* OR on))

We conducted 45 small searches like these, combining outcomes into a 46th search. This search yielded around 350 results, which we concluded offered an acceptable balance of coverage/recall (number of items to review) versus accuracy/precision (the relevance of items shown to our research questions). The abstract of each result was scanned to assess whether it fulfilled our principle inclusion criterion, and coded to show recommendations for exclusion or inclusion. The bibliographic list and codes were then shared with the project team for discussion about inclusion.

¹⁸² Bramer, W M, Rethlefsen, M L, Kleijnen, J et al. (2017) 'Optimal database combinations for literature searches in systematic reviews: a prospective exploratory study'. *Syst Rev* 6(245). <https://doi.org/10.1186/s13643-017-0644-y>

¹⁸³ Martín-Martín, A, Orduna-Malea, E, Thelwall, M & Delgado-López-Cózar, E (2019) 'Google Scholar, Web of Science and Scopus: which is best for me?' *LSE Impact Blog*, <https://blogs.lse.ac.uk/impactofsocialsciences/2019/12/03/google-scholar-web-of-science-and-scopus-which-is-best-for-me/>

5.3. The snowball search (of academic and grey literature)

We also carried out a snowball search. In addition to the search described above, a repository was established at the beginning of the project, designed to capture different kinds of evidence. This included academic and grey, as well as quantitative and qualitative research. These were sourced from existing knowledge, together with searching of key bibliographies, word of mouth and Twitter. A process of coding took place using the same technique as on the systematic search list, assessing whether a reference met our inclusion criteria. This kind of descriptive coding enables discussion and deliberation as to why a particular item was or was not shortlisted for review. The two academic lists were combined with grey literatures and shared with the research team, together with coding and notes relating to the search criteria.

5.4. The review of survey reports

We reviewed surveys and survey reports by first looking at the technical details of surveys, making notes on the sample sizes, fieldwork dates, and survey modes, and identifying the key findings from the survey research. In almost all cases, these key findings consisted of percentages of respondents having given particular answers to questions, rather than being derived from statistical models. We then compared what we identified to be key findings with concluding or key claims made in reports about the surveys. In some cases, one survey would generate two reports.

5.5. Finalising bibliographic lists

All literature was assessed against our inclusion criterion: does it report empirical research about what people think and how they feel about data practices and what happens to their personal data? In some cases, project team members were already familiar with some literature. Abstracts or executive summaries of literature that emerged from the searches described above was rapidly reviewed. Where methods were not indicated in abstracts / executive summaries, the literature itself was rapidly reviewed. Full texts were retrieved for literature that met the inclusion criteria. These in turn were assessed against inclusion criteria. Some literature was then excluded, and some previous excluded items were reintroduced at this stage.

At this stage we introduced more detailed inclusion criteria. Because of the proliferation of research on certain aspects of this topic in recent years, we limited the research that we reviewed in the following ways:

1. **By dates:** we reviewed literature published between 2015 and 2019, and we included some literature published in early 2020 which we became aware of whilst undertaking our review.
2. **By geography for grey literature:** because of the proliferation of surveys and polls in our field of enquiry in recent years, and because grey literature often aims to have a national impact, the grey literature we reviewed is UK-focused (either UK-only or based on international research which included the UK). Relevant research has been undertaken elsewhere in the world (for example by the Pew Research Center's Internet & American Life project) but we did not include it in our analysis¹⁸⁴. There is less academic research that is UK-focused and there are a number of international collaborations in this field. Because of this, and because academic research contributes to an international conversation, we included selected international studies in the academic literature we discuss.

¹⁸⁴ The Pew Research Center's Internet & American Life project <https://www.pewresearch.org/internet/>

3. **By populations researched/research subjects:** we excluded literature about children's understandings and perceptions of data practices because the study of children (and digital media) is a specialist field outside the remit of our own research. Studies of adults' perceptions of data practices relating to children were included.
4. **By domain, especially with regard to research about public perceptions of health data:** research into public perceptions of uses of health data and the ethics of health data re-use is more advanced than in any other domain. As a result, more is known about public perceptions of datafication in health, and high quality syntheses have already been undertaken, for example by *Understanding Patient Data*¹⁸⁵. We therefore focused our review primarily on domains other than health.
5. **By subject matter/focus:** a large proportion of the research into public attitudes and perceptions that has been undertaken focuses on privacy, surveillance and security. We largely excluded this from our review, except where there was an obvious focus on attitudes to data practices included in the research.
6. **Existing evidence syntheses and reviews:** A small number of evidence syntheses and reviews have been published in this field. Some of these cover publications outside our timeframe¹⁸⁶. Others cover domains that are not our focus¹⁸⁷. For these reasons and in order not to reproduce work already undertaken, we did not carry out analysis of the literature covered by these syntheses in this document.

This rationale and the bibliographic lists agreed by the research team were shared with the *Living With Data* Advisory Group, on our project website, via Twitter and on a number of relevant academic lists. We invited people to suggest further inclusions, which we added to our lists if they met our criteria. As we wrote our review, we became aware of a small number of relevant new publications via social media, email lists and our networks, and we added these to our lists. These then became our final lists: those derived from searches and reviewed by the team + additions suggested to us + items found whilst writing the review.

No literature search is exhaustive. Each database search has its limitations, as Martín-Martín et al point out¹⁸⁸. Pragmatic decisions therefore must be made, 'sacrificing a little accuracy for a lot more comprehensive coverage' or vice-versa¹⁸⁹. It has been noted that Web of Science is limited when it comes to profiling the social sciences¹⁹⁰. Furthermore, systematic searches privilege certain types of literature, such as peer-reviewed articles in certain journals, over other literature that is equally relevant, such as grey literature like reports. This can lead to the exclusion of some relevant research from evidence reviews. Snowball searches that start with the expertise of the reviewing team also

¹⁸⁵ *Understanding Patient Data* (2018) *Public attitudes to patient data use: a summary of existing research*.

http://understandingpatientdata.org.uk/sites/default/files/2019-01/Public%20attitudes%20key%20themes_0.pdf

¹⁸⁶ For example Bakir, V, Cable, J, Dencik, L, Hintz, A & McStay, A (2015) *Public Feeling on Privacy, Security and Surveillance, DATA-PSST and DCSS Project Report*. <https://orca.cf.ac.uk/87335/1/Public-Feeling-on-Privacy-Security-Surveillance-DATAPSST-DCSS-Nov2015.pdf> draws on publications from before 2015.

¹⁸⁷ For example *Understanding Patient Data* (2018) *Public attitudes to patient data use: a summary of existing research*. http://understandingpatientdata.org.uk/sites/default/files/2019-01/Public%20attitudes%20key%20themes_0.pdf on health data.

¹⁸⁸ Martín-Martín, A, Orduna-Malea, E, Thelwall, M & Delgado-López-Cózar, E (2019) 'Google Scholar, Web of Science and Scopus: which is best for me?' *LSE Impact Blog*, <https://blogs.lse.ac.uk/impactofsocialsciences/2019/12/03/google-scholar-web-of-science-and-scopus-which-is-best-for-me/>

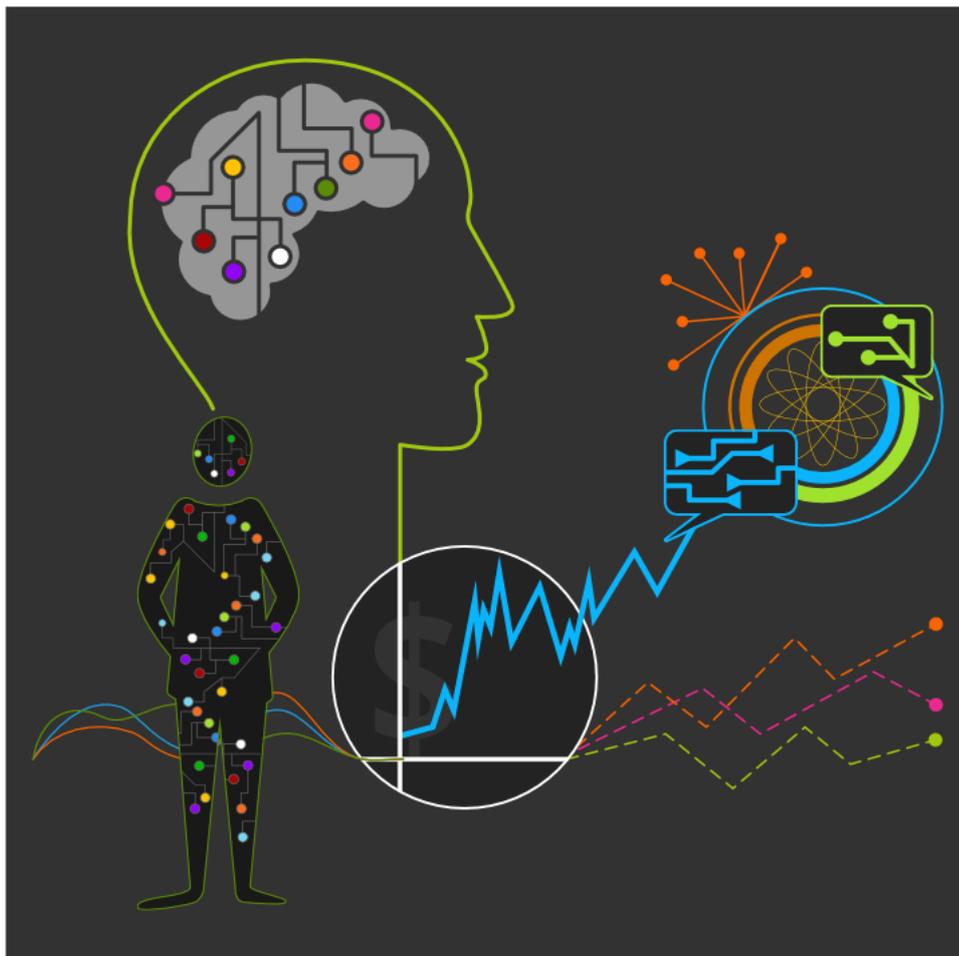
¹⁸⁹ Harzing, A (2016) *Sacrifice a little accuracy for a lot more comprehensive coverage* <https://harzing.com/publications/white-papers/gsbook-prologue>

¹⁹⁰ It has been noted in Harzing, A (2013) 'Document categories in the ISI Web of Knowledge: Misunderstanding the Social Sciences?', *Scientometrics* 94:23–34 <https://doi.org/10.1007/s11192-012-0738-1>, for example.

have their limitations and biases, shaped by the disciplinary expertise and experiential knowledge of team members. We tried to minimise these limitations by trialling and refining search terms, combining a systematic search with a snowball search and inviting input into our search processes from Advisory Group members and other interested parties.

5.6. Analysing the literature

Through inductive analyses, we identified key themes, similarities, differences and issues of concern, rather than systematically rating the literature we identified as relevant to our interests. We have structured this review according to what emerged from this process. We noted that how research was framed, what its aims were and the methods that were deployed all played a role in shaping findings and claims made about them, and we discuss this issue above. In section 2.10 above, we note that researchers' perspectives may lead to particular research priorities, to asking particular questions, or to analysing results in particular ways. In recognition of this point, we acknowledge here that we, the authors of this report, have our own (sometimes different) views on existing data practices, whether and how they should change, and what fair data practices might look like and that these views will have informed our analysis of the research that we discuss in this report.



6. References

6.1. Academic literature

1. Ajana, B (2017) *Self-tracking: Empirical and philosophical investigations*, Springer International Publishing. <https://doi.org/10.1007/978-3-319-65379-2>
2. Alvarado, O & Waern, A (2018) 'Towards algorithmic experience: initial efforts for social media contexts', *CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, April 2018. Paper No.: 286. <https://doi.org/10.1145/3173574.3173860>
3. Bergstrom, A (2015) 'Online privacy concerns: a broad approach to understanding the concerns of different groups for different uses', *Computers In Human Behavior*, 53:419-426. <https://doi.org/10.1016/j.chb.2015.07.025>
4. Bolin, Göran & Jonas Andersson Schwarz (2015) 'Heuristics of the algorithm. Big Data, user interpretation and institutional translation', *Big Data & Society*, 2(2): 1–12. <https://doi.org/10.1177/2053951715608406>
5. Bowyer, A, Montague, K, Wheeler, S, McGovern, S, Lingam, R & Balaam, M (2018) 'Understanding the family perspective on the storage, sharing and handling of family civic data', *CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, April 2018. Paper No.: 136. <https://doi.org/10.1145/3173574.3173710>
6. Brown, A, Chouldechova, A, Putnam-Hornstein, E, Tobin, A & Vaithianathan, R (2019) 'Toward algorithmic accountability in public services: a qualitative study of affected community perspectives on algorithmic decision-making in child welfare services', *CHI '19: Proceedings of the 2019 Conference on Human Factors in Computing Systems*, May 2019. Paper No.: 41. <https://doi.org/10.1145/3290605.3300271>
7. Bucher, T (2017) 'The algorithmic imaginary: exploring the ordinary affects of Facebook algorithms', *Information, Communication & Society*, 20: 30-44. <https://doi.org/10.1080/1369118X.2016.1154086>
8. Colbjørnsen, T (2018) 'My algorithm: user perceptions of algorithmic recommendations in cultural contexts', in Andrea L Guzman (ed) *Human-machine communication: rethinking communication, technology, and ourselves*. Peter Lang.
9. Couldry, N, Fotopoulou, A & Dickens, L (2016) 'Real social analytics: a contribution towards a phenomenology of a digital world', *The British Journal of Sociology* 67(1): 118-137. <https://doi.org/10.1111/1468-4446.12183>
10. Dencik, L & Cable, J (2017) 'Digital Citizenship and Surveillance: The advent of surveillance realism: public opinion and activist responses to the Snowden leaks', *International Journal of Communication*, 11(2017):763-781. <https://ijoc.org/index.php/ijoc/article/view/5524/1939>
11. Dencik, L (2019) 'Situating practices in datafication — from above and below', in H Stephansen and E Treré (eds) *Citizen media and practice*.
12. Dolin, C, Weinschel, B, Shan, S, Hahn, C, Choi, E, Mazurek, M & Blase, U (2018) 'Unpacking perceptions of data-driven inferences underlying online targeting and personalization', *CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, April 2018. Paper No.: 493. <https://doi.org/10.1145/3173574.3174067>
13. Draper, N A, & Turow, J (2019) 'The corporate cultivation of digital resignation', *New Media & Society*, 21(8):1824–1839. <https://doi.org/10.1177/1461444819833331>
14. Elhai, J, Levine, J & Hall, B (2017) 'Anxiety about electronic data hacking: predictors and relations with digital privacy protection behavior', *Internet Research*, 27(3):631-649. <https://doi.org/10.1108/IntR-03-2016-0070>

15. Ellis, D (2019) 'Techno-securitisation of everyday life and cultures of surveillance-apatheia', *Science as Culture*, 29(1), 11-29. <https://doi.org/10.1080/09505431.2018.1561660>
16. Eslami, M, Rickman, A, Vaccaro, K, Aleyasen, A, Vuong, A, Karahalios, K, Hamilton, K & Sandvig, C (2015) 'I always assumed that I wasn't really that close to [her]: reasoning about invisible algorithms in news feeds', *CHI '15: Proceedings of the 2015 CHI Conference on Human Factors in Computing Systems*, April 2015. <https://doi.org/10.1145/2702123.2702556>
17. Eslami, M, Sneha, R, Kumaran, K, Sandvig, C & Karahalios, K (2018) 'Communicating algorithmic process in online behavioral advertising', *CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, April 2018 Paper No.: 432. <https://doi.org/10.1145/3173574.3174006>
18. Eubanks, V (2018) *Automating inequality: how high-tech tools profile, punish and police the poor*. St Martin's Press.
19. Fiesler, C & Hallinan, B (2018) 'We are the product: public reactions to online data sharing and privacy controversies in the media', *CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, April 2018 Paper No.: 53. <https://doi.org/10.1145/3173574.3173627>
20. Fiore-Gartland, B & Neff, G (2015) 'Communication, mediation and the expectations of data: data valences across health and wellness communities', *International Journal of Communication*, 9. <http://ijoc.org/index.php/ijoc/article/view/2830>.
21. Gangadharan, S P (2017) 'The downside of digital inclusion: expectations and experiences of privacy and surveillance among marginal internet users,' *New Media and Society*, 19(4):597-615. <https://doi.org/10.1177/1461444815614053>
22. Gangadharan, S P (2021) 'Digital exclusion: a politics of refusal', in H Landemore, R Reich & L Bernholz (eds) *Digital Technology and Democratic Theory*. University of Chicago Press. <http://eprints.lse.ac.uk/103076/>
23. Gangadharan, S P & Niklas, J (2019) 'Decentering technology in discourse on discrimination', *Information, Communication and Society*, 22(7), 882–899. <https://doi.org/10.1080/1369118X.2019.1593484>
24. Guberek, T, McDonald, A, Simioni, S, Mhaidli, A, Toyama, K, Schaub, F (2018) 'Keeping a Low Profile?: Technology, Risk and Privacy among Undocumented Immigrants', *CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, April 2018 Paper No.: 114. <https://doi.org/10.1145/3173574.3173688>
25. Harrison, G, Hanson, J, Jacinto, C, Ramirez, J & Blase, U (2020) 'An empirical study on the perceived fairness of realistic, imperfect machine learning models', *ACM FAT* conference 2020*. <https://dl.acm.org/doi/abs/10.1145/3351095.3372831>
26. Horne, C & Przepiorka, W (2019) 'Technology use and norm change in online privacy: experimental evidence from vignette studies', *Information, Communication & Society*, (0):1-17. <https://doi.org/10.1080/1369118X.2019.1684542>
27. Kennedy, H, Elgesem, D, & Miguel, C (2015) 'On fairness: user perspectives on social media data mining', *Convergence*, 23(3): 270–288. <https://doi.org/10.1177/1354856515592507>
28. Kennedy, H & Hill, R (2017) 'The feeling of numbers: emotions in everyday engagements with data and their visualisation', *Sociology*, 52(4): 830-848. <https://doi.org/10.1177/0038038516674675>
29. Kennedy, H, Steedman, R & Jones, R (2020a) 'Approaching public perceptions of datafication through the lens of inequality: a case study in public service media', *Information, Communication and Society*. <https://doi.org/10.1080/1369118X.2020.1736122>
30. Kizilcec, R F (2016) 'How much information? effects of transparency on trust in an algorithmic interface', *CHI '16: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, May 2016. <https://doi.org/10.1145/2858036.2858402>

31. Lee, M K (2018) 'Understanding perception of algorithmic decisions: fairness, trust, and emotion in response to algorithmic management', *Big Data & Society*, 5(1).
<https://doi.org/10.1177/2053951718756684>
32. Lomborg, S & Kapsch, P (2019) 'Decoding algorithms', *Media, Culture & Society*. <https://doi.org/10.1177/0163443719855301>
33. Lomborg, S, Thylstrup, N & Schwartz, J (2018) 'The temporal flows of self-tracking: checking in, moving on, staying hooked', *New Media & Society*, 20(12): 4590-4607 <https://doi.org/10.1177/1461444818778542>
34. Lupton, D (2017) "'It just gives me a bit of peace of mind': Australian women's use of digital media for pregnancy and early motherhood", *Societies* 7(3)25.
<https://doi.org/10.3390/soc7030025>
35. Lupton, D (2019) 'Data mattering and self-tracking: what can personal data do?', *Continuum*, 34(1):1-13. <http://dx.doi.org/10.1080/10304312.2019.1691149>
36. Lupton, D & Michael, M (2017) "'Depends on who's got the data'": public understandings of personal digital dataveillance', *Surveillance and Society*, 15:254-268.
<https://doi.org/10.24908/ss.v15i2.6332>
37. Machuletz, D, Laube, S, Böhme, R (2018) 'Webcam covering as planned behavior', *CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, April 2018*, Paper No.: 180. <https://doi.org/10.1145/3173574.3173753>
38. Magalhães, J C (2018) 'Do algorithms shape character? considering algorithmic ethical subjectivation', *Social Media + Society*, 4(2). <https://doi.org/10.1177/2056305118768301>
39. Marwick, A, Fontaine, C & Boyd, D (2017) "'Nobody sees it, nobody gets mad": social media, privacy and personal responsibility among low-SES youth' *Social Media + Society*, 3(2).
<https://doi.org/10.1177/2056305117710455>
40. Molina, M D, Gambino, A & Sundar, S S (2019) 'Online privacy in public places: how do location, terms and conditions and VPN influence disclosure?', *CHI '19: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, May 2019. Paper No.: LBW2616.
<https://doi.org/10.1145/3290607.3312932>
41. Pink, S & Fors, V (2017) 'Being in a mediated world: self-tracking and the mind-body-environment', *Cultural Geographies*. 24(3)375-388.
<https://doi.org/10.1177/1474474016684127>
42. Pink, S, Lanzeni, D & Horst, H (2018) 'Data anxieties: finding trust in everyday digital mess', *Big Data & Society* 5(1). <https://doi.org/10.1177/2053951718756685>
43. Pink, S, Sumartojo, S, Lupton, D & La Bond, C H (2017) 'Mundane data: the routines, contingencies and accomplishments of digital living', *Big Data & Society*, 4(1).
<https://doi.org/10.1177/2053951717700924>
44. Popham, J Lavoie, J & Coomber, N (2020) 'Constructing a public narrative of regulations for big data and analytics: results from a community-driven discussion', *Social Science Computer Review*, 38(1)75-90. <https://doi.org/10.1177/0894439318788619>
45. Potoglou, D, Dunkerley, F, Patil, S & Robinson, N (2017) 'Public preferences for internet surveillance, data retention and privacy enhancing services: evidence from a pan-European study', *Computers in Human Behavior*, 75: 811-825. <https://doi.org/10.1016/j.chb.2017.06.007>
46. Pridmore, J & Mols, A (2020) 'Personal choices and situated data: Privacy negotiations and the acceptance of household Intelligent Personal Assistants', *Big Data & Society*, 7(1).
<https://doi.org/10.1177/2053951719891748>
47. Pybus, J, Coté, M & Blanke, T (2015) 'Hacking the social life of big data: a data literacy framework', *Big Data & Society*, 2(2). <https://doi.org/10.1177/2053951715616649>

48. Rader, E & Gray, R (2015) 'Understanding User Beliefs About Algorithmic Curation in the Facebook News Feed', In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*: 173–182. <https://doi.org/10.1145/2702123.2702174>
49. Rendina, H J & Mustanski, B (2018) 'Privacy, trust, and data sharing in web-based and mobile research: participant perspectives in a large nationwide sample of men who have sex with men in the united states', *Journal of Medical Internet Research*, 20(7). <https://doi.org/10.2196/jmir.9019>
50. Ruckenstein, M & Pantzar, M (2015) 'Datafied Life: techno-anthropology as a site for exploration and experimentation', *Techné: Research in Philosophy and Technology*, 19(2):191-210. <https://doi.org/10.13140/RG.2.1.2553.7762>
51. Ruckenstein, M S (2017) 'Keeping data alive: talking DTC genetic testing', *Information, Communication and Society*, 20(7):1024-1039. <https://doi.org/10.1080/1369118X.2016.1203975>
52. Ruckenstein, M, & Granroth, J (2019) 'Algorithms, advertising and the intimacy of surveillance', *Journal of Cultural Economy*, 13(1):12-24. <https://doi.org/10.1080/17530350.2019.1574866>
53. Sannon, S, Bazarova, N N & Cosley, D (2018) 'Privacy lies: understanding how, when, and why people lie to protect their privacy in multiple online contexts', *CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, April 2018 Paper No.: 52 Pages 1–13. <https://doi.org/10.1145/3173574.3173626>
54. Steedman, R, Kennedy, H & Jones, R (2020) 'Complex ecologies of trust in data practices and data-driven systems', *Information, Communication and Society*. <https://doi.org/10.1080/1369118X.2020.1748090>
55. Sumartojo, S, Pink, S, Lupton, D & La Bond C H (2016) 'The affective intensities of datafied space', *Emotion, Space and Society*, 21:33-40. <https://doi.org/10.1016/j.emospa.2016.10.004>
56. Weinberger, M, Zhitomirsky-Geffet, M & Bouhnik, D (2017) 'Sex differences in attitudes towards online privacy and anonymity among Israeli students with different technical backgrounds', *Information Research: An International Electronic Journal*, 22(4). <https://eric.ed.gov/?id=EJ1164311>
57. Weiner, K, Will, C, Henwood, F, Williams, R (2020) 'Everyday curation? Attending to data, records and record keeping in the practices of self-monitoring', *Big Data & Society* 7(1). <https://doi.org/10.1177/2053951720918275>
58. Williams, M L, Burnap, P & Sloan, L (2017) 'Towards an ethical framework for publishing twitter data in social research: taking into account users' views, online context and algorithmic estimation', *Sociology*, 51(11):1149-1168. <https://doi.org/10.1177/0038038517708140>
59. Wilmott, C (2016) 'Small moments in spatial big data: calculability, authority and interoperability in everyday mobile mapping', *Big Data & Society*, 3(2). <https://doi.org/10.1177/2053951716661364>
60. Woodruff, A & Fox, S E, Rousso-Schindler, S & Warshaw, J (2018) 'A qualitative exploration of perceptions of algorithmic fairness', *CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, April 2018 Paper No.: 656. <https://doi.org/10.1145/3173574.3174230>
61. Yamashita, N Kuzuoka, H, Kudo, T, Hirata, K, Aramaki, E & Hattori, K (2018) 'How information sharing about care recipients by family caregivers impacts family communication', *CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, April 2018, Paper No.: 222. <https://doi.org/10.1145/3173574.3173796>

6.2. Grey Literature

1. Ada Lovelace Institute (2019) *Beyond face value - public attitudes to facial recognition technology*. https://www.adalovelaceinstitute.org/wp-content/uploads/2019/09/Public-attitudes-to-facial-recognition-technology_v.FINAL_.pdf
2. Robinson, G & Dolk, H (2015) *Research update: public attitudes to data sharing in Northern Ireland*. Administrative Research Data Centre, Northern Ireland <https://www.ark.ac.uk/publications/updates/update108.pdf>
3. Big Brother Watch & ComRes (2015) *UK public research – online privacy*. <https://www.bigbrotherwatch.org.uk/wp-content/uploads/2015/03/Big-Brother-Watch-Polling-Results.pdf>
4. Cabinet Office & Government Digital Service (2016) *Better use of data in government*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/535063/better_use_of_data_in_government_response_final.pdf
5. Carnegie Trust - Scott, Kaela (2018) *Data for public benefit: balancing the risks and benefits of data sharing*. https://www.involve.org.uk/sites/default/files/field/attachemnt/Data%20for%20Public%20Benefit%20Report_0.pdf
6. Carnegie Trust - Carolyn Black, Lucy Setterfield & Rachel Warren (2019) *Online data privacy from attitudes to action: an evidence review*. https://d1ssu070pg2v9i.cloudfront.net/pex/carnegie_uk_trust/2018/08/03110116/Online-Data-Privacy-from-Attitudes-to-Action-CUKT.pdf
7. Citizens Advice - Illuminas (2016) *Consumer expectations for personal data management in the digital world*. <https://www.citizensadvice.org.uk/Global/CitizensAdvice/Consumer%20publications/Personal%20data%20consumer%20expectations%20research.docx.pdf>
8. Digital Catapult (2015) *Trust in personal data: A UK review*. Digital Catapult
9. Dencik, L, Hintz, A, Redden, J & Warne, H (2018) *Data Scores as Governance: Investigating uses of citizen scoring in public services*, Cardiff University. <https://datajustice.files.wordpress.com/2018/12/data-scores-as-governance-project-report2.pdf>
10. The Direct Marketing Association (2018) *Data privacy: what the consumer really thinks*. https://dma.org.uk/uploads/misc/5a857c4fdf846-data-privacy--what-the-consumer-really-thinks-final_5a857c4fdf799.pdf
11. Doteveryone (2018) *People, power, and technology: the 2018 digital understanding report*. http://understanding.doteveryone.org.uk/files/Doteveryone_PeoplePowerTechDigitalUnderstanding2018.pdf
12. Doteveryone (2018) *People, power and technology: the 2018 digital attitudes report*. <https://www.doteveryone.org.uk/report/digital-attitudes/>
13. Doteveryone – Catherine Miller (2019) *Engaging the public with responsible technology: four principles and three requirements*. <https://doteveryone.org.uk/download/3225/>
14. Doteveryone - Joe Massey, Jacob Ohrvik-Stott & Catherine Miller (2019) *Better redress: building accountability for the digital age: an evidence review from Doteveryone* <https://www.doteveryone.org.uk/wp-content/uploads/2019/12/Better-redress-evidence-review.pdf>
15. Edelman (2018) *Edelman Trust Barometer 2018, UK Findings*. <https://www.edelman.co.uk/magazine/posts/edelman-trust-barometer-2018/>

16. The European Commission (2019) *Special Eurobarometer 487a. Summary - The General Data Protection Regulation*
<https://ec.europa.eu/commfrontoffice/publicopinionmobile/index.cfm/Survey/getSurveyDetail/surveyKy/2222>
17. Demos - Harry Evans, Steve Ginnis & Jamie Bartlett (2015) *#Socialethics a guide to embedding ethics in social media research*. <https://www.ipsos.com/ipsos-mori/en-uk/ipsos-mori-and-demoscasm-call-better-ethical-standards-social-media-research>
18. HEPI (2019) *Students or data subjects?* <https://www.hepi.ac.uk/wp-content/uploads/2019/12/Students-or-data-subjects-Report-122-Web-FINAL.pdf>
19. Hinz, A & Brand, J (nd) *Data policies: regulatory approaches for data-driven platforms in the UK and EU*. <https://datajustice.files.wordpress.com/2020/01/data-policies-research-report-revised.pdf>
20. Hopkins Van Mil: Creating Connections Ltd (2015) *Big data: public views on the use of private sector data for social research - a findings report for the Economic and Social Research Council*. <https://esrc.ukri.org/files/public-engagement/public-dialogues/public-dialogues-on-the-re-use-of-private-sector-data-for-social-research-report/>
21. Information Commissioner's Office – Harris Interactive (2019) *Information rights strategic plan: trust and confidence*. <https://ico.org.uk/media/about-the-ico/documents/2615515/ico-trust-and-confidence-report-20190626.pdf>
22. Information Commissioner's Office (2019) *Information Commissioner's annual report and financial statements 2018-19*. <https://ico.org.uk/media/about-the-ico/documents/2615262/annual-report-201819.pdf>
23. Ipsos MORI (2018) *The state of the state 2017-2018: austerity, government spending, social care and data* <https://www.ipsos.com/sites/default/files/ct/publication/documents/2017-10/the-state-of-the-state-2017-2018.pdf>
24. Kennedy, H, Hartman, T, Steedman, R & Jones, R (2020b) *UK public unhappy with the ways their data is managed*. <https://livingwithdata.org/project/wp-content/uploads/2020/03/Views-on-Data-Management-Full-Report.pdf>.
25. Oman, S (2019a) *Improving data practices to monitor inequality and introduce social mobility measures: a working paper*. The University of Sheffield.
https://www.sheffield.ac.uk/polopoly_fs/1.867756!/file/MetricsWorkingPaper.pdf
26. Oman, S (2019b) *Measuring social mobility in the creative and cultural industries – the importance of working in partnership to improve data practices and address inequality*. The University of Sheffield.
https://www.sheffield.ac.uk/polopoly_fs/1.867754!/file/MetricsPolicyBriefing.pdf
27. ODI / Open Data Institute (2018) *Who do we trust with personal data?*
<https://theodi.org/article/who-do-we-trust-with-personal-data-odi-commissioned-survey-reveals-most-and-least-trusted-sectors-across-europe/>
28. Ofcom (2019) *Adult's media use and attitudes report: 2019*.
https://www.ofcom.org.uk/data/assets/pdf_file/0021/149124/adults-media-use-and-attitudes-report.pdf
29. OII / Oxford Internet Institute (2015) *Internet use, behaviour and attitudes in Great Britain 2003-2015*. <http://oxis.oii.ox.ac.uk/wp-content/uploads/2015/01/OxIS-Brochure.pdf>
30. OII / Oxford Internet Institute (2019) *Perceived threats to privacy online: the Internet in Britain*. <https://oxis.oii.ox.ac.uk/wp-content/uploads/sites/43/2019/09/OxIS-report-2019-final-digital-PDFA.pdf>
31. PEGA (2019) *GDPR: Show me the data survey reveals EU consumers poised to act on legislation*. <https://www.pega.com/system/files/resources/2019-07/GDPR-Show-Me-The-Data-eBook.pdf>

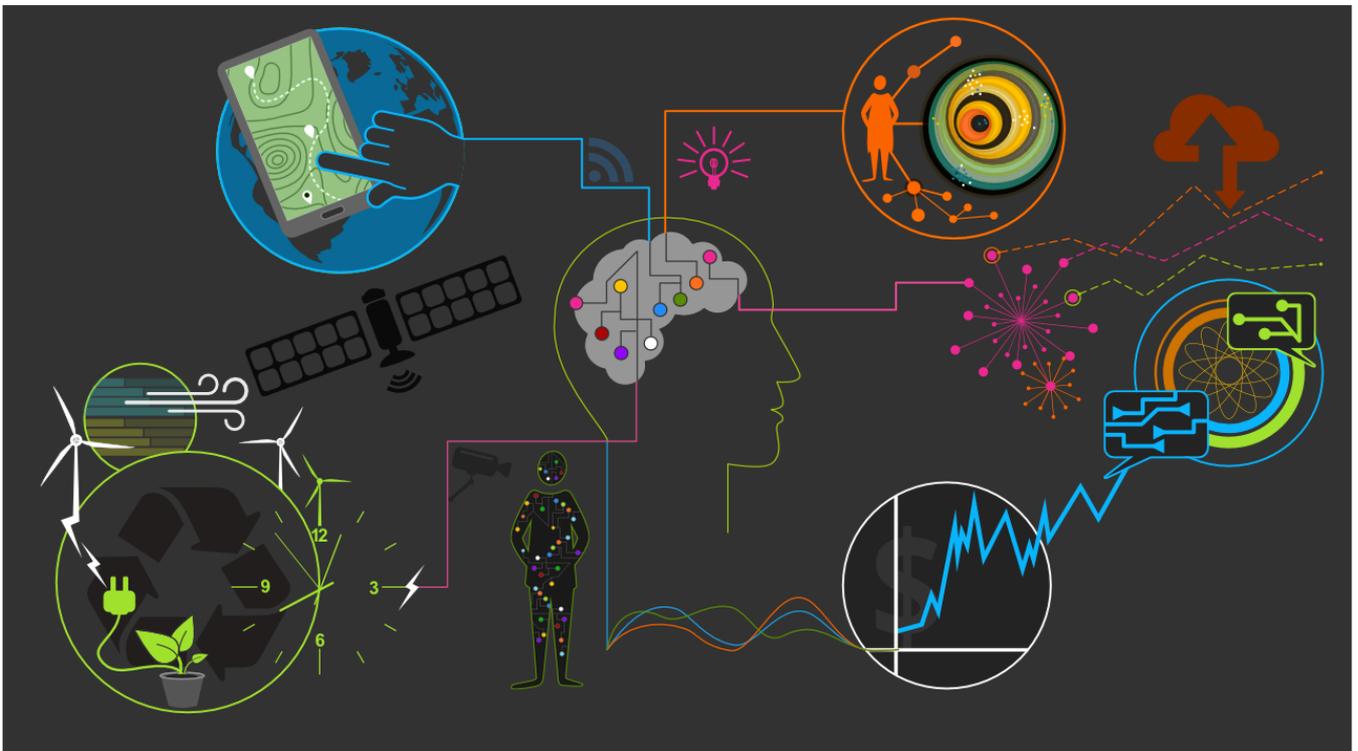
32. RSA / The Royal Society for the Encouragement of Arts, Manufactures and Commerce Royal Society of Arts - The Forum for Ethical AI (2019) *Democratising decisions about technology: a toolkit*. <https://www.thersa.org/discover/publications-and-articles/reports/democratising-decisions-technology-toolkit>
33. RSA / The Royal Society for the Encouragement of Arts, Manufactures and Commerce Royal Society of Arts (2018) *Artificial intelligence: real public engagement*. <https://www.thersa.org/discover/publications-and-articles/reports/artificial-intelligence-real-public-engagement>
34. RSA / The Royal Society for the Encouragement of Arts, Manufactures and Commerce Royal Society of Arts - Renate Samson, Kayshani Gibbon & Anna Scott (2019) *About data about us*. <https://www.thersa.org/discover/publications-and-articles/reports/data-about-us>
35. Sopra Steria (2017) *The citizen view of the digital transformation of government*. <https://www.soprasteria.co.uk/docs/librariesprovider41/White-Papers/sopra-steria-ipsos-digitaltransformation-of-govt.pdf?sfvrsn=0>
36. Turow, J, Hennessy, M, Draper, N, Akanbi, O & Virgilio, D (2018) *Divided we feel: partisan politics drive American's emotions regarding surveillance of low-income populations*. https://repository.upenn.edu/cgi/viewcontent.cgi?article=1563&context=asc_papers
37. Turow, J, Hennessy, M & Draper, N (2015) *The trade off fallacy*. https://www.asc.upenn.edu/sites/default/files/TradeoffFallacy_1.pdf

6.3. Other literature referenced (reviews and syntheses, international reports and non-empirical references)

1. The British Academy & The Royal Society - Franck Fourniol & Fiona McLaughlin (2017) *Data governance: public engagement review*. <https://www.thebritishacademy.ac.uk/sites/default/files/Data-Governance-public-engagement-review.pdf>
2. Coleman, S (2013) *How Voters Feel*, Cambridge University Press. <https://doi.org/10.1017/CBO9781139035354>
3. European Union (2016) *General data protection regulation*, Off J Eur Union 49: L119. <https://gdpr-info.eu>
4. Petty, T, Saba, M, Lewis, T, Gangadharan, S P & Eubanks, V (2018) *Reclaiming our data: interim report*. https://www.odbproject.org/wp-content/uploads/2016/12/ODB.InterimReport.FINAL_7.16.2018.pdf
5. Pew Research Center - Brooke Auxier, Lee Rainie, Monica Anderson, Andrew Perrin, Madhu Kumar & Erica Turner *Americans and Privacy: Concerned, Confused and Feeling Lack of Control Over Their Personal Information*. <https://www.pewresearch.org/internet/2019/11/15/americans-and-privacy-concerned-confused-and-feeling-lack-of-control-over-their-personal-information/>
6. Redden, J (2018) 'The harm that data do'. *Scientific American*, 319(5). <https://www.scientificamerican.com/article/the-harm-that-data-do/>
7. RSS / Royal Statistical Society (2019) *The Data Manifesto* <https://www.rss.org.uk/Images/PDF/influencing-change/2019/Data%20Manifesto2019.pdf>
8. Understanding Patient Data (2018) *Public attitudes to patient data use: a summary of existing research*. http://understandingpatientdata.org.uk/sites/default/files/2019-01/Public%20attitudes%20key%20themes_0.pdf

6.4. References from Introduction and Review methodology sections

1. Bakir, V, Cable, J, Dencik, L, Hintz, A & McStay, A (2015) *Public Feeling on Privacy, Security and Surveillance*, DATA-PSST and DCSS Project Report. <https://orca.cf.ac.uk/87335/1/Public-Feeling-on-Privacy-Security-Surveillance-DATAPSST-DCSS-Nov2015.pdf>
2. Bramer, W M, Rethlefsen, M L , Kleijnen, J *et al.* (2017) 'Optimal database combinations for literature searches in systematic reviews: a prospective exploratory study'. *Syst Rev* 6(245). <https://doi.org/10.1186/s13643-017-0644-y>
3. Harzing, A (2013) 'Document categories in the ISI Web of Knowledge: Misunderstanding the Social Sciences?', *Scientometrics* 94:23–34 <https://doi.org/10.1007/s11192-012-0738-1>
4. [Harzing, A \(2016\)](https://doi.org/10.1007/s11192-012-0738-1) *Sacrifice a little accuracy for a lot more comprehensive coverage* <https://harzing.com/publications/white-papers/gobook-prologue>
5. Martín-Martín, A, Orduna-Malea, E, Thelwall, M & Delgado-López-Cózar, E (2019) 'Google Scholar, Web of Science and Scopus: which is best for me?' *LSE Impact Blog*, <https://blogs.lse.ac.uk/impactofsocialsciences/2019/12/03/google-scholar-web-of-science-and-scopus-which-is-best-for-me/>
6. Mayer-Schoenberger, V & Cukier, K (2013) *Big Data: a revolution that will transform how we live, work and think*, John Murray Publishing.



7. About this report and its authors

Helen Kennedy is Professor of Digital Society at the University of Sheffield. Over 20+ years, she has researched how digital developments are experienced by citizens/publics/'ordinary people' and how these experiences can inform the work of digital media practitioners.

Susan Oman is Lecturer in Data, AI and Inequality at the University of Sheffield. Susan researches how data works in context and in relation to particular policy issues, including well-being, loneliness, inequality and class.

Mark Taylor is Senior Lecturer in Quantitative Methods at the University of Sheffield, and AHRC Leadership Fellow (Creative Economy) until 2021. His research focuses on understanding the relationship between culture and inequality.

Jo Bates is Senior Lecturer in Information Politics and Policy at the University of Sheffield. Her research is situated in the field of critical data studies, and she works collaboratively with scholars across other fields, including computer and information scientists.

Robin Steedman is a Postdoctoral Researcher in Creative and Cultural Industries in Africa at Copenhagen Business School. She is interested in global creative and cultural industries, and in questions of diversity and inequality in media production, distribution, and viewership.

Preferred citation: Helen Kennedy, Susan Oman, Mark Taylor, Jo Bates & Robin Steedman (2020) *Public understanding and perceptions of data practices: a review of existing research*. *Living With Data*, University of Sheffield. <http://livingwithdata.org/current-research/publications/>

Living With Data is a programme of research which aims to understand the new roles of data in society and how they affect the lives of ordinary people. <https://livingwithdata.org/>

Living With Data is funded by The Nuffield Foundation (grant number OSP/43959). 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. www.nuffieldfoundation.org