



Data Matters Are Human Matters: final *Living With Data* report on public perceptions of public sector data uses

Helen Kennedy, Susan Oman, Hannah Ditchfield, Mark Taylor, Jo Bates,
Itzelle Medina-Perea, Lulu Pinney, Monika Fratzczak
October 2022



livingwithdata.org

Contents

1.	Executive summary _____	3
2.	<i>Living With Data</i> research in context _____	5
3.	Inequalities and different perceptions of data uses _____	11
4.	Data solidarities _____	13
5.	The role of commercial companies in public sector data uses _____	15
6.	Unpacking concern _____	18
7.	Understanding and imagining data uses _____	20
8.	What our findings tell us about communicating about data uses _____	24
9.	Conclusions and recommendations _____	25
10.	References _____	29
11.	Appendix: information about methods & evidence review findings _____	31
12.	Acknowledgements _____	37

1. Executive summary

A. Different people have different perceptions of data uses

Different people have different perceptions of different data uses. Stakeholders across policy, practice and research should avoid generalising about public opinions of data uses.

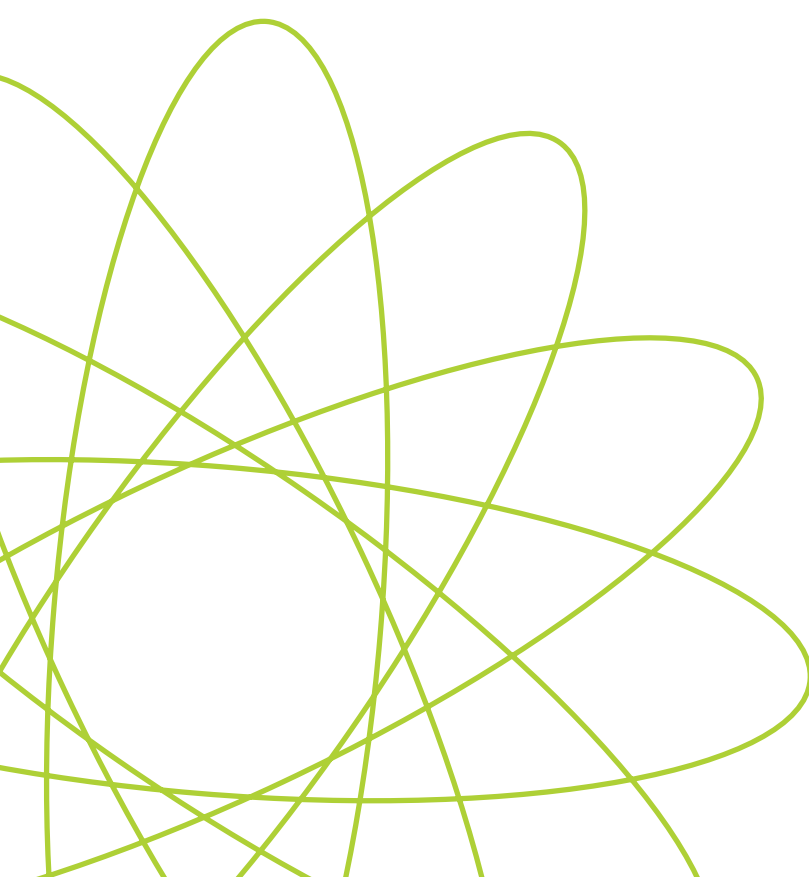
Because of these differences, we need to look beyond headline findings about public opinions of data uses, and regularly consult diverse publics, because data uses and perceptions of them change, and because structural inequalities influence what different people think of different data uses.

B. Data inequalities can lead to data solidarities

As a result of unequal experiences of structures and systems, data-driven or otherwise, some groups are more concerned about some data uses than others.

For example, Black, Asian and other racialised people are more concerned about what the police do with people's data, and LGBTQ+ people are more concerned about data uses in the health context. However, these differences were often small. People from different groups were aware of structural inequalities and concerned about data uses that reproduce or exacerbate them. These findings tell us that people should be put at the centre of data policy-making and data practices:

- in how stakeholders think about public perceptions of data uses (discussed in A);
- in how they communicate about data uses (discussed in C);
- and in data uses themselves, which may need to change or be desisted in response to this finding (discussed in D).



C. Communication should enable genuine understanding & respond to what matters to people

The aim of communicating about data uses and data-driven systems should be to enable genuine understanding in the people whose data feeds such systems. It's not enough just to communicate.

How data uses are communicated is important. Communication about data uses needs to respond to what matters to people and address their concerns. We found that this includes: concerns about what might go wrong in a data-driven system; what might change in the future; and how disadvantaged communities might be more negatively or adversely affected than other groups.

D. Change or desist data uses that are not human-centric, eg that discriminate or from which private companies profit

Good communication alone is not enough. What is needed most is better data uses.

Widespread concern about data uses communicates a strong message to data policy-makers and practitioners about public dissatisfaction with existing data uses. The people who know most about data uses are the most concerned about them, which also suggests that data uses, or aspects of them, are concerning. If data uses continue unchanged, the public will continue to be concerned, regardless of how effectively communication enables understanding.

What constitutes a 'good' data use needs to be defined on a case-by-case basis, in consultation with diverse publics, whose views may change over time. However, sharing data that has been gathered for pro-social or the public good with commercial companies who will make a profit from it is widely disliked. Public sector data practitioners should consider alternative ways of delivering data-driven services. This will not be easy, as global technology companies monopolise the provision of particular technologies and technical infrastructures, but it is not impossible.

IN SHORT:

Data uses need to change, so they eliminate harms and are in the public or social interest. Sometimes, in order to do these things, specific data uses need to stop. Much more is needed than simply acknowledging differences and getting the communication right. If data uses overcome inequalities and are stripped of aspects which concern people, then public perceptions of them will probably improve.

2. *Living With Data* research in context

CONTEXT & DEBATES

The increasingly widespread collection and use of digital data has wide-ranging effects on people's lives. The benefits of 'datafication', as data-driven processes are sometimes called, include more efficient and effective service provision, in a broad range of areas.

But these benefits are accompanied by real and potential risks and harms, from loss of privacy to new forms of discrimination, inequality and injustice.

Committed to using data responsibly and minimising harms, policy and civil society actors have responded to these risks. For example, the Centre for Data Ethics and Innovation (CDEI) was established in 2018, a government expert body which aims to enable the trustworthy use of data and AI. The independent Ada Lovelace Institute (Ada) was also established in the same year, the mission of which is to ensure that data works 'for people and society'. There have also been a large number of parliamentary inquiries, for example into data ethics, responsible uses of data and AI, and digital government. More recently, reforms to the UK Data Act, the UK's Digital Strategy, which focuses on 'unlocking' the economic potential of data, and strategy and guidance on specific types of data, from health to location data, reflect continued efforts to get data uses and governance right.

A central aspect of good, responsible data and AI governance is ensuring that it is attuned to what the public wants. As a result, interest in how the public perceives data uses has also grown, amongst academic researchers focusing on public views of the new role of data in society and policy-makers and practitioners keen to ensure that their data

uses and processes are perceived positively. Understanding public views of data uses is at the heart of initiatives like CDEI and Ada. Research into public understanding and perceptions of data uses has therefore flourished in recent years, as evidenced in the extensive [Living With Data literature review](#) that we published in 2020 (Kennedy et al 2020).

However, there are some gaps in research and understanding, which we aimed to fill with our *Living With Data* research. These are:

- Research has found that social inequalities play a major role in shaping people's experiences of data uses (eg Eubanks 2018, Noble 2018). There is less research into whether and how various social inequalities shape perceptions of data uses. Understanding the relationship between social inequalities and perceptions of data uses will help us identify whether and how data uses can be improved. We write about what we found about this relationship in Sections 3 and 4 of this report.
- Some commentators ask whether particular data uses are fair, but others (eg Kalluri 2020) believe that concepts like equity or justice are more helpful, because they recognise the role that structural inequalities play in shaping data uses. We make two contributions to this debate:

- a. We describe the kinds of data uses that diverse members of the public consider to be fair, so that improvements to data uses can be informed by public views.
 - b. We show that fairness, equity and justice overlap for the people we spoke to, in what we describe as ‘data solidarities’, and suggest that these concepts are not as distinct as is sometimes proposed.
- Data uses can be difficult to understand. Greater transparency (eg Gebru et al 2021, Mitchell et al 2019) or improved data literacy (eg Yates et al n.d., ODI 2022) are often proposed as solutions to this problem, a proposal which assumes that more information and better skills will improve understanding. This is linked to another assumption, that the more people understand data uses, the more positive their attitudes will be. Our research challenges these assumptions. We show that we need a better grasp of what it means to understand data uses, how people come to understand them and the roles that imagining plays in people’s understandings of data uses in Section 7. We describe what our findings suggest with regard to communicating about data uses in Section 8.

This discussion can be found in Section 4.

- A lot of research has found that people are concerned about data uses (eg ICO 2019 and Pew 2019). There is less research into the factors that inform people’s concerns, or lack of concern, when it comes to data uses. We have found that these factors include negative impacts on people from disadvantaged and minority groups, discussed in Section 4, the involvement of commercial companies, discussed in Section 5, the context in which a data process takes place, discussed in Section 6, and demographic differences, discussed in Sections 3 and 6.
- This document provides a summary of *Living With Data’s* main findings from across a range of research activities, and it provides links to full reports on all aspects of our work, which can be found in the [Resources & Publications](#) section of the *Living With Data* website. This document also discusses the implications of our findings, and makes recommendations for policy-makers, practitioners and researchers interested in how to ensure that data and AI work for people and societies.

What we did

Living With Data (LWD) was a research project funded by the Nuffield Foundation which ran from September 2019 to September 2022. LWD aimed to understand people’s perceptions of how data about them is collected, analysed, shared and used, and how these processes could be improved.

Readers will note that above, we use the term ‘data uses.’ With this term, we refer to data collection, processing, analysis and sharing and what happens as a result of these practices. We recognise that this simple term may seem misleading to data experts, but we needed a clear and accessible phrase to encourage people to talk to us about this topic. In what follows, we use ‘data uses’ as a shorthand for all of these activities across the data pipeline. The data at the centre of such processes is often personal data, 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 rights).

Aware that much of the previous research into public attitudes to data uses had focused on attitudes to high-profile data practices (eg Dencik and Cable 2017), we felt that it was important to understand people's views of everyday data uses, such as those that take place in the public sector, because of the role that they play in shaping everyday lives. We identified welfare, media and health as three public sector domains on which to focus our research because they are core aspects of everyday life. Focusing on these domains, we asked:

A. What do different people know and feel about specific data-related practices in different domains of everyday life?

B. What do fair data practices look like, from non-experts' perspectives?

To address our research aims and questions, we undertook the following activities:

- i. Evidence Review:** We conducted a review of original empirical research, published in grey and academic literature, into public perceptions of, attitudes towards and feelings about data uses, which was published between 2015-2020. We provide more details about the evidence review and some of our findings in the appendix and in our [full evidence review report](#) which can be found on the *Living With Data* website.
- ii. Documenting specific data uses in public sector organisations:** We produced accounts and visualisations of data uses to discuss with participants. For welfare and media, we partnered with the Government Department for Work and Pensions (DWP) and the British Broadcasting Corporation (BBC), who selected data uses for us. For health, we produced an account of a data use based on information in the public domain, and another based on research that one of us had undertaken into National Health Service (NHS) data flows (Medina Perea et al 2019). Figure 1 shows small versions of the visualisations we discussed with focus group and interview participants. Larger visualisations and more information

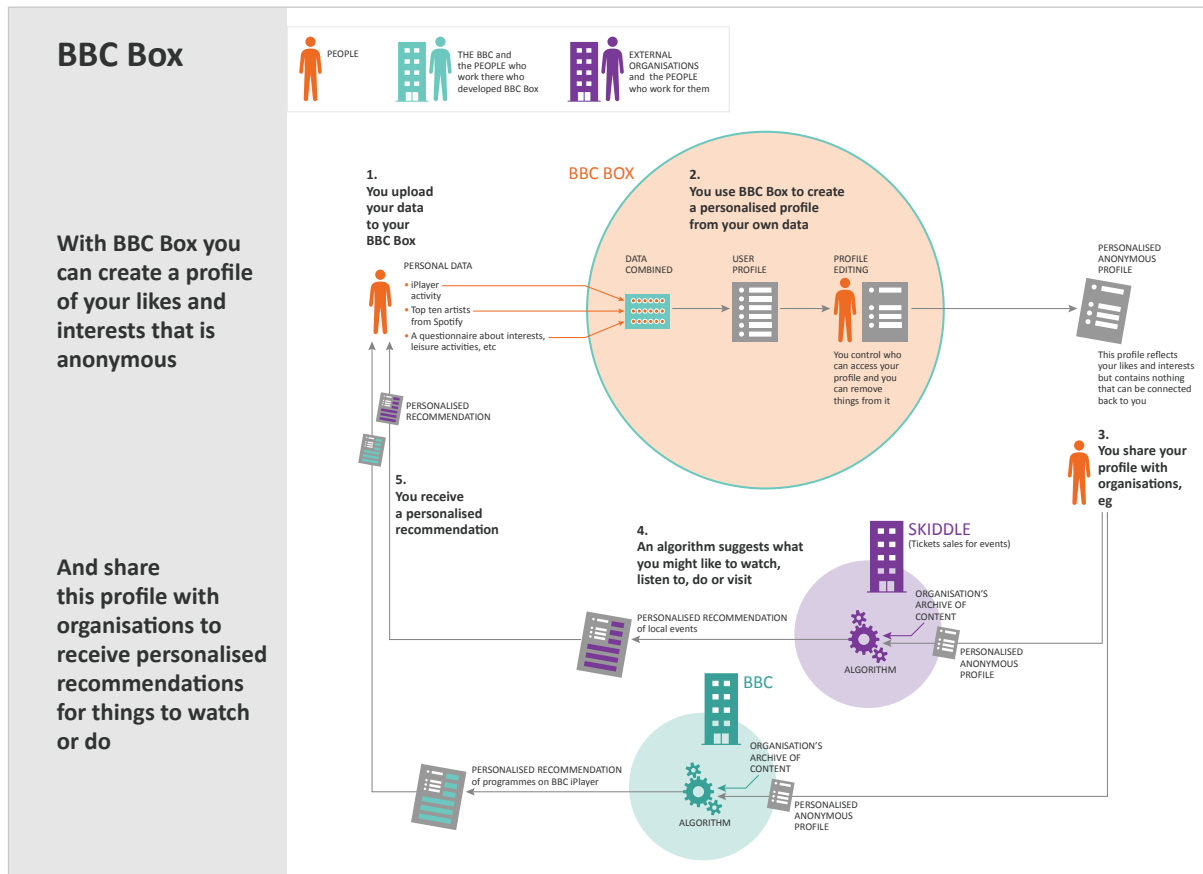
about data uses and our process can be found in the appendix and on the [Public Sector Data Uses](#) page of our website.

iii. Survey of public attitudes to data uses:

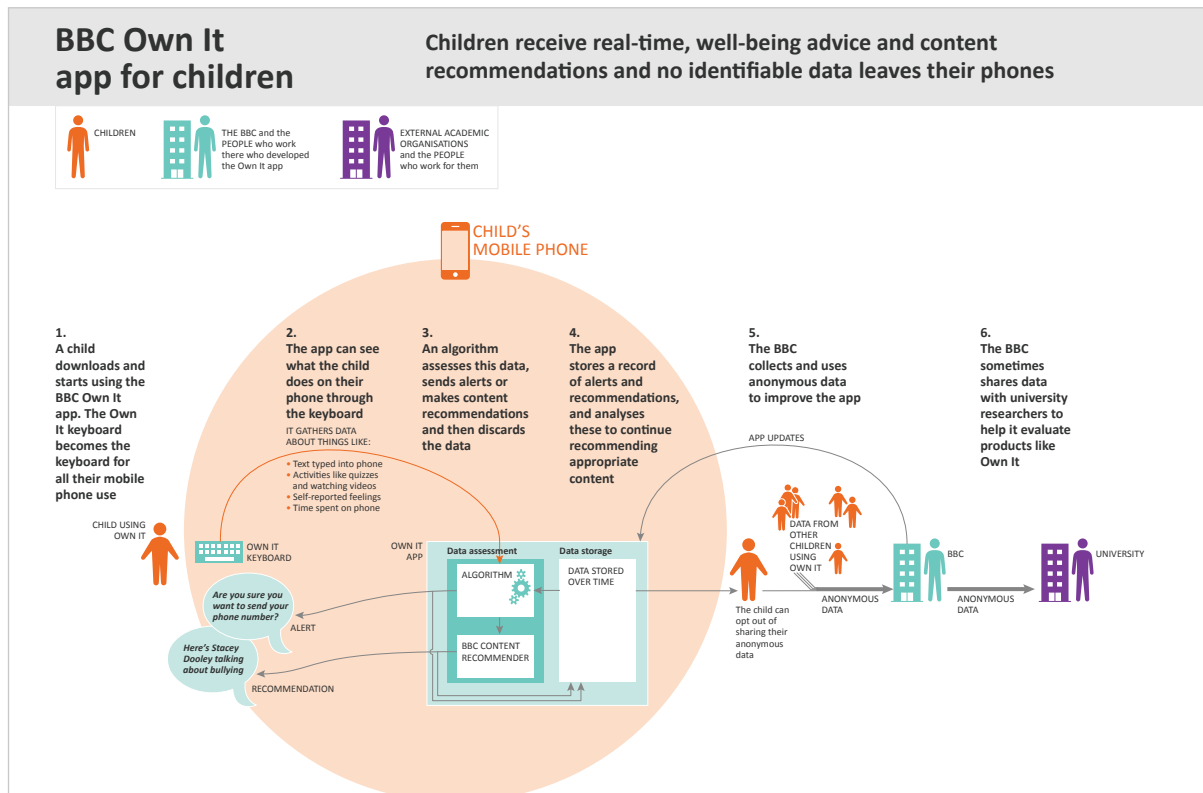
We carried out two waves of a survey of people's attitudes to data uses, the first at the end of 2020, the second at the end of 2021, each with 2000 adult respondents from the UK. The sample was nationally representative in relation to gender, age, income, disability and ethnicity, with 'boosts' of people born outside the UK, LGBTQ+ people, and people in receipt of the UK's main welfare benefit, Universal Credit, to enable analysis. In response to the vast majority of questions, we found no statistically significant difference between figures from the first and second waves of the survey. Therefore, in this report, we refer to the 2021 survey and we highlight any differences between the 2020 and 2021 surveys that we did find. More detail can be found in the appendix and in our [full survey report](#).

- iv. Focus groups and interviews:** We carried out interviews and focus groups with 112 adults in the UK from November 2020 to September 2021, in which we also explored general attitudes and awareness, and perceptions of specific public sector data uses, using the visualisations we produced as elicitation tools. In this report, we share some of the information about participants that we gathered. Sometimes the information is relevant to the things that they said and the views they expressed. Sometimes we share information to paint a picture of the person we are writing about, or to avoid 'othering' – that is, highlighting only minority characteristics. We don't share all of the information we gathered about each participant to preserve participants' anonymity, something which is especially important when writing about disadvantaged, vulnerable or minority groups (Fox et al 2021). We also use pseudonyms to refer to participants as a means of preserving anonymity. More detail can be found in the appendix and in our [full report of our qualitative research](#).

Figure 1: Images and short descriptions of the six data uses which were the focus of our research



BBC Box, a prototype which pulls together data about what users watch or listen to and gives them control over who has access to this data.

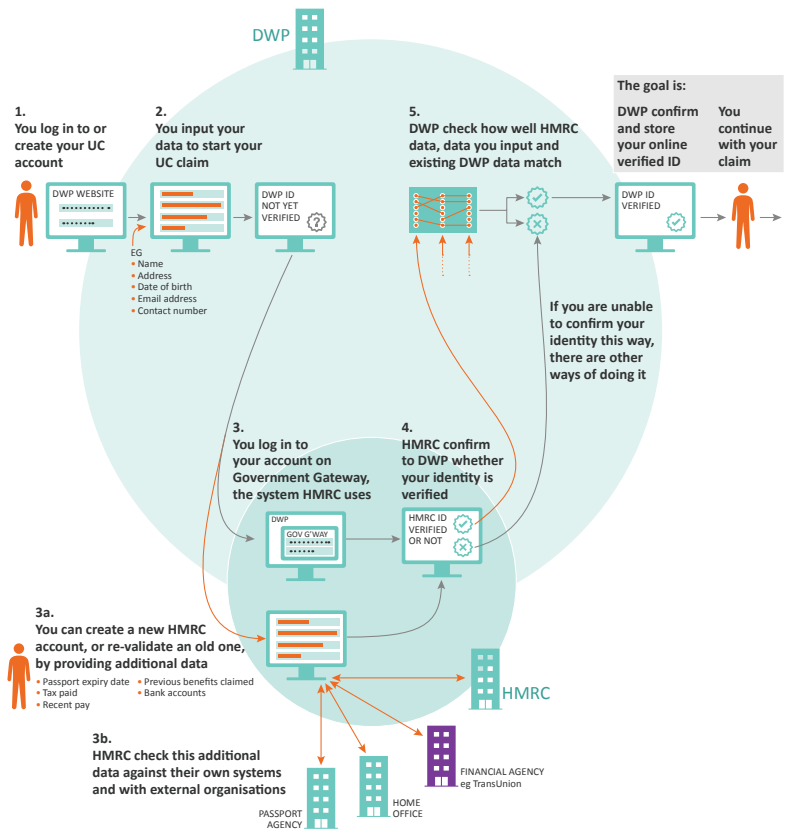


BBC Own It, a free app designed to support, help and advise children when they use their phones to chat and explore the online world, without adult supervision.

DWP Confirm Your Identity

When you claim Universal Credit (UC) online DWP needs to check you are who you say you are before you can continue with your claim

All the processes shown here are secure and automated

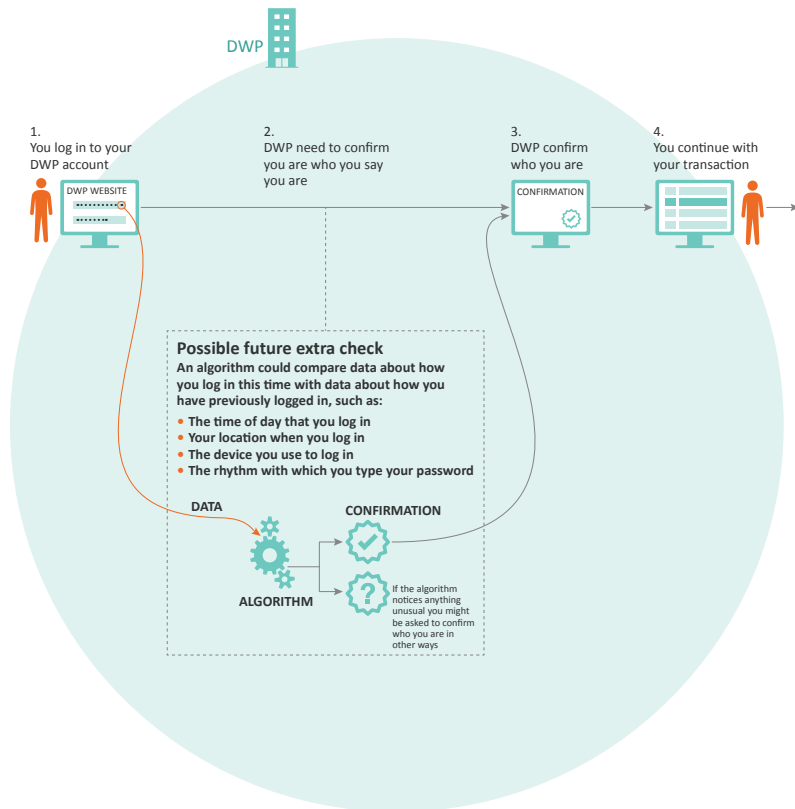


Confirm Your Identity, an identity verification process for Universal Credit payments which makes it possible to confirm identity online.

DWP Dynamic Trust Hub

It is increasingly possible to interact with DWP online. If you do this, the DWP needs to confirm you are who you say you are. As part of a larger project, it is considering a range of security checks that it could use to support this process.

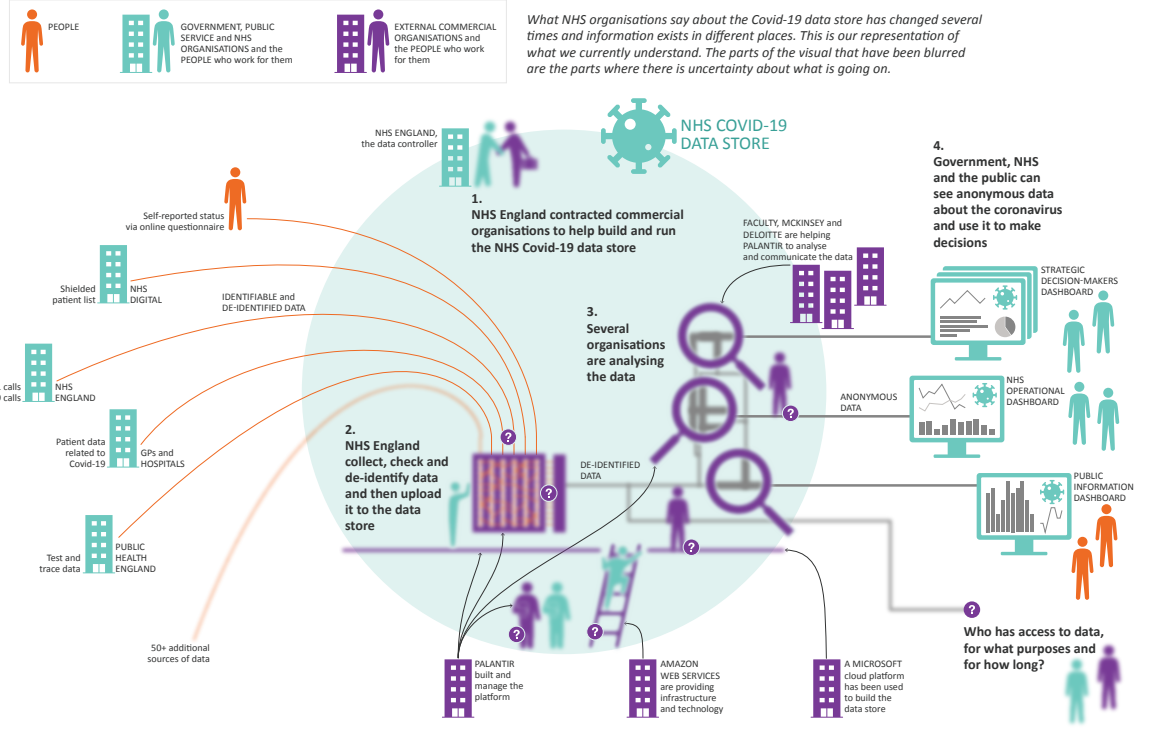
All the processes shown here are secure and automated



Dynamic Trust Hub, which explored a range of issues to enhance identity verification, including technology integration and possible security checks..

NHS Covid-19 data store

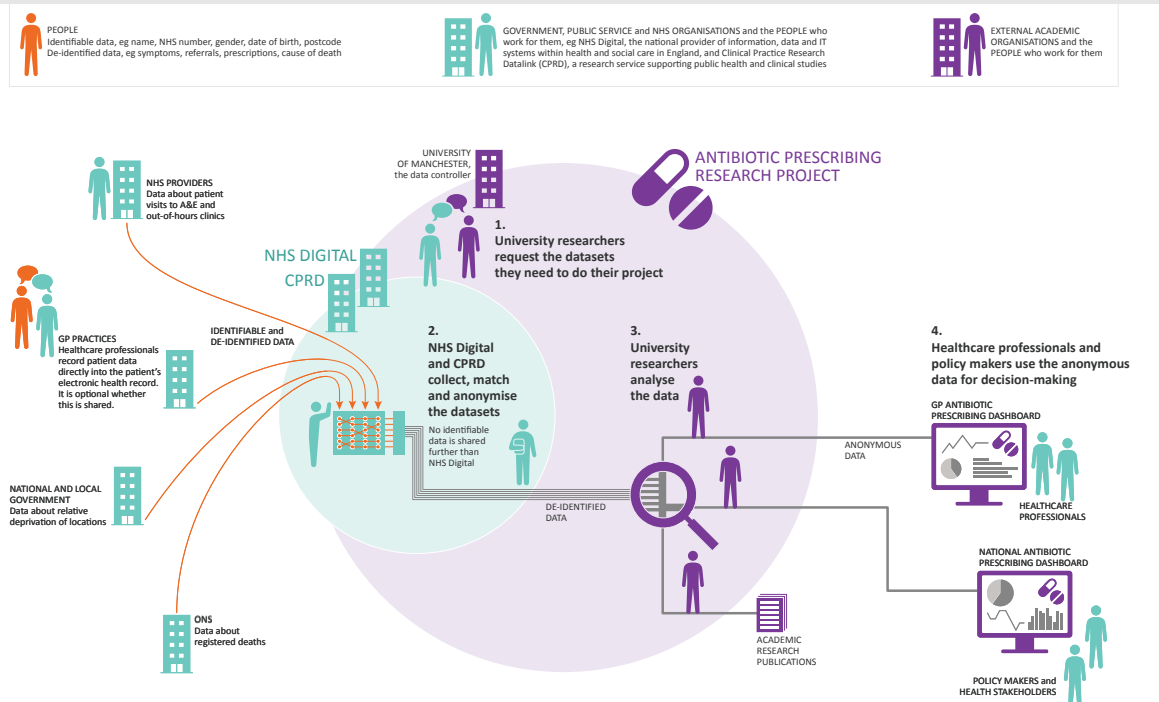
A national, secure data store to hold data in one place to help national organisations responsible for coordinating the Covid-19 response



NHS Covid-19 Data Store, a national Data Store to help organisations responsible for coordinating the Covid-19 response.

NHS antibiotic prescribing research project

Existing data about us is matched, linked and compared to help GPs use antibiotics better



NHS antibiotic prescribing research project, exploring ways to address the public health crisis of antibiotic resistance.

3. Inequalities and different perceptions of data uses

We all live with data, but our experiences of *Living With Data* are not all the same. They are shaped by who we are, our identities, and our wider social and political environment. On *Living With Data*, we found that belonging to a disadvantaged or minority group informed people's perceptions of data uses.

People from disadvantaged or minority groups expressed different views about data uses to their advantaged or majority counterparts. Education, economic status, age, dis/ability, gender identity and sexuality, English as an additional language, and race and ethnicity appeared to inform participants' perceptions of data uses some of the time. To be clear, we are not suggesting that there is a direct correlation between belonging to a demographic group and perceptions of data uses. Our point is that demographic characteristics shape life experiences and in turn, those experiences shape perceptions of data uses.

We identified small differences in perceptions of data uses between disadvantaged, minority and advantaged, majority groups. All differences reported are statistically significant at the 95% level. For example:

- In the survey, disabled people were more positive about the re-use and sharing of health data for research purposes, and more concerned by commercial companies providing data-driven public services, than people who did not have a disability.
- The survey also showed different degrees of trust in different sectors or organisations' data uses across ethnic groups. We found that white people trusted the police's data uses more than Black, Asian and other racialised people (a term we use following [Sobande and others](#) and acknowledging

criticisms of the term BAME). The survey also revealed age-related differences. People aged 65 and older trusted their GP to be open and transparent about what they do with people's data, to use personal data responsibly, and to keep data safe, much more than the youngest 18-24 age group.

- The survey also found that LGBTQ+ people were less likely to trust health organisations than heterosexual cisgender respondents. We found this in our interviews and focus groups too, where a number of LGBTQ+ participants expressed concern about uses of sexual health data, because of the ways that this kind of data can be mobilised for harmful ends.
- Interviews and focus groups showed that the intersection of demographic characteristics was important in people's reflections on the data uses that we showed them and that we describe in the appendix below. For example, Gulay, a Turkish woman who is quoted below, noted that she found it hard to understand the DWP's uses of data for identity verification because English is not her first language. At the same time, knowing that data about her was safe and secure was really important, because of her status as a refugee.

“ This is hard for me because it’s in English and long. I know I am not accept, not continue. I just accept. [...] I’m refugee. I come to this country because of problems in my country. I don’t want my location to be shared with this [Turkish] government. This is important for me. I know this country [UK] is very safe and – it is okay. [...] But I don’t want to share with other countries, like my country.

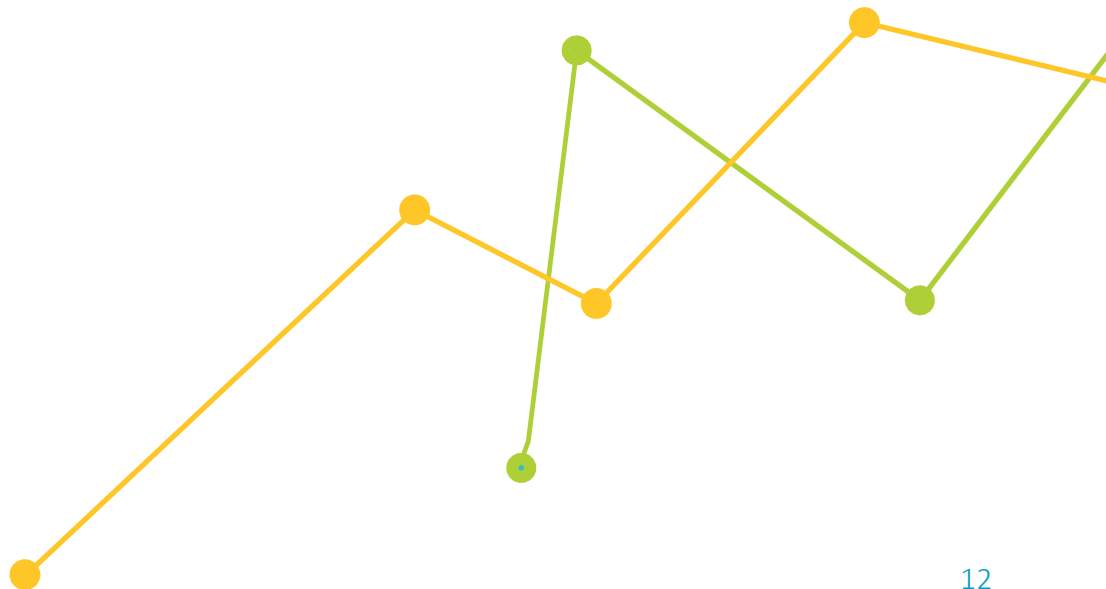
Gulay, a heterosexual, Turkish woman, with an annual household income of less than £10,000

”

IN SHORT:

Differences and inequalities matter when it comes to perceptions of data uses. Not all data uses are the same – we say more about this below – and people experience them from different social positions. Social inequalities play a role in shaping people’s experiences of data uses, and therefore their understandings and perceptions of them.

There is no one ‘public’ and no such thing as ‘the public’s attitudes to data uses’. The idea of a singular public obscures the differences and inequalities that characterise diverse publics and their perceptions of data uses. This means it is important to look beyond headline findings about public perceptions of data uses, and it is better to talk about ‘diverse publics’ than ‘the public’.



4. Data solidarities

Although we found that social inequalities play a role in shaping people's perceptions of data uses, as reported above, the differences that we found in the survey were not large. In the focus groups and interviews, we found widespread concern about the potential for the public sector data uses that we discussed to have negative consequences for people from disadvantaged and minority groups.

Participants spoke about the need for data uses to be inclusive 'for all communities', sometimes listing groups who they perceived to be excluded from certain processes. Without explicitly using the term, a lot of participants appeared to be aware that data uses can reinforce or deepen inequalities, and that some data uses are more likely than others to do so.

So social inequalities matter in two different ways. First, belonging to a disadvantaged or minority group informs people's perceptions of data uses. This confirms what has been found in relation to experiences of data uses (eg by Eubanks 2018 and Noble 2018). Second, and perhaps more surprisingly, social inequalities matter as a collective concern: participants who didn't belong to a disadvantaged or minority group were still concerned about how these groups might be more negatively impacted by data uses than others. Often, participants from one disadvantaged or minority group were concerned about the effects of data uses on another disadvantaged or minority group. In short, we found that concern about the consequences of data uses for people from these groups, while not universal, was very common.

As an example, in the quote opposite, Tahira, a charity worker, is talking about BBC Own It, a free app designed to support, help and advise children when they use their phones to chat and explore the online world without adult supervision. Tahira expressed concern about the effects that socio-economic inequalities

A lot of people I know wouldn't normally have access to that kind of resource. [Pakistani parents sitting at home in the UK] wouldn't know where to reach out to, because they've not been educated in this country, for example, or just don't know. [...] So, for me that's the fair one, if I was to look at it from that lens. [...]

One family member I know, like family friend kind of thing, she's on it. She will check the kids' phone, she will – you know, she is like really – without being too aggressive, she knows how to manage that. Whereas other parents are working three jobs, they haven't got the time or the, you know, know-how of what to check. Even if they were to ask that child, 'Oh, what are you doing?' They'd say, 'Oh, yeah, I'm just doing this'. And they're like, 'Alright'.

Tahira, a heterosexual, Pakistani woman, aged 45-54, with an annual household income of £50,000-£69,000

have on parents' ability to engage with their children about their phone and app usage. She was positive about how Own It could provide access to helpful resources to children whose parents otherwise might not be aware of or able to access them. At the same time, she was concerned that not all parents are able to support their children's mobile phone use, precisely because of these inequalities.

Across LWD surveys, focus groups and interviews, participants were especially concerned about the possibility of the DWP data uses that we discussed with them reinforcing inequalities, when compared to the health and media data uses which were also discussed. Accessibility was seen as particularly important with regard to welfare data uses,

because welfare services are essential to people who are likely to be disadvantaged by structural inequalities. As noted above, the welfare data processes that we discussed with participants, Confirm Your Identity and Dynamic Trust Hub, both related to online identity verification and how to enhance it, for example through additional security checks. Focus group participants felt that the additional security checks that were being considered within these two projects could be difficult for certain groups to engage with, such as people who do not have access to the relevant technology in their homes, as Huso, a Black British woman, quoted below, noted. This concern about data uses in the welfare setting demonstrates that context is an important factor in shaping diverse people's perceptions of data uses.

Who wants to be filling out security checks in a public library, where they can see what you're doing on the computer, because the computers are placed so that everyone can see what you're doing on the computer? You don't want everyone seeing all that information about you. But if that's the only place you have to actually use a computer, [...] that's really bad.

Huso, a black Black British, heterosexual teacher, aged 25-34

IN SHORT:

People don't want data uses to have negative consequences for people from disadvantaged and minority groups.

We found that our participants don't want data uses to have negative consequences for people from disadvantaged and minority groups. Their concern that data uses might reinforce inequalities could be seen as a form of data solidarity, with solidarity understood as 'standing in unity with others and showing support in the struggle for justice' (Nikunen 2019, p15).

Data policy-makers and practitioners must acknowledge that data uses reinforcing inequalities is a widespread concern. They also need to understand the potentially discriminatory impacts of different data-driven systems, in order to be able to overcome them.

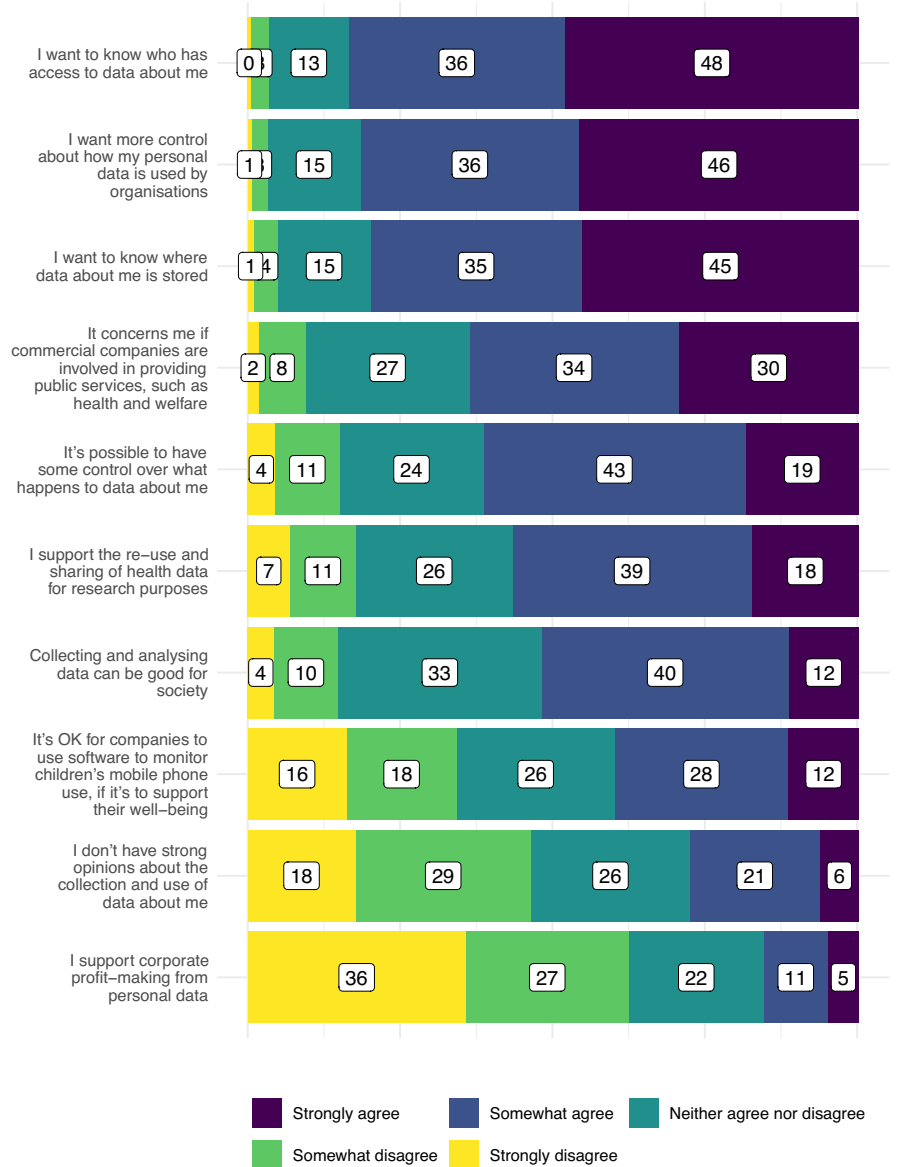
5. The role of commercial companies in public sector data uses

Across LWD surveys, focus groups and interviews, we found confusion and concern about the involvement of commercial companies in the provision of the specific public data uses that we discussed with participants and that we describe in the appendix below.

In our surveys, to gauge respondents' general attitudes to data uses, we presented them with ten statements and asked them to indicate how much they agreed or disagreed with each one. The statements we presented to respondents can be seen in Figure 2. As can be seen, respondents wanted to know who has access to data about them (84% of respondents agreed or strongly agreed with the relevant statement), they wanted more control over how their data is used by organisations (82%), and they wanted to know where data about them is stored (80%). Collectively, these statistics indicate broad concern amongst the public regarding who gets to access and use their personal data.

While these findings show a general concern about what happens to data about them, other responses to this question reveal that there is a specific concern about commercial companies accessing or using personal data for the purposes of

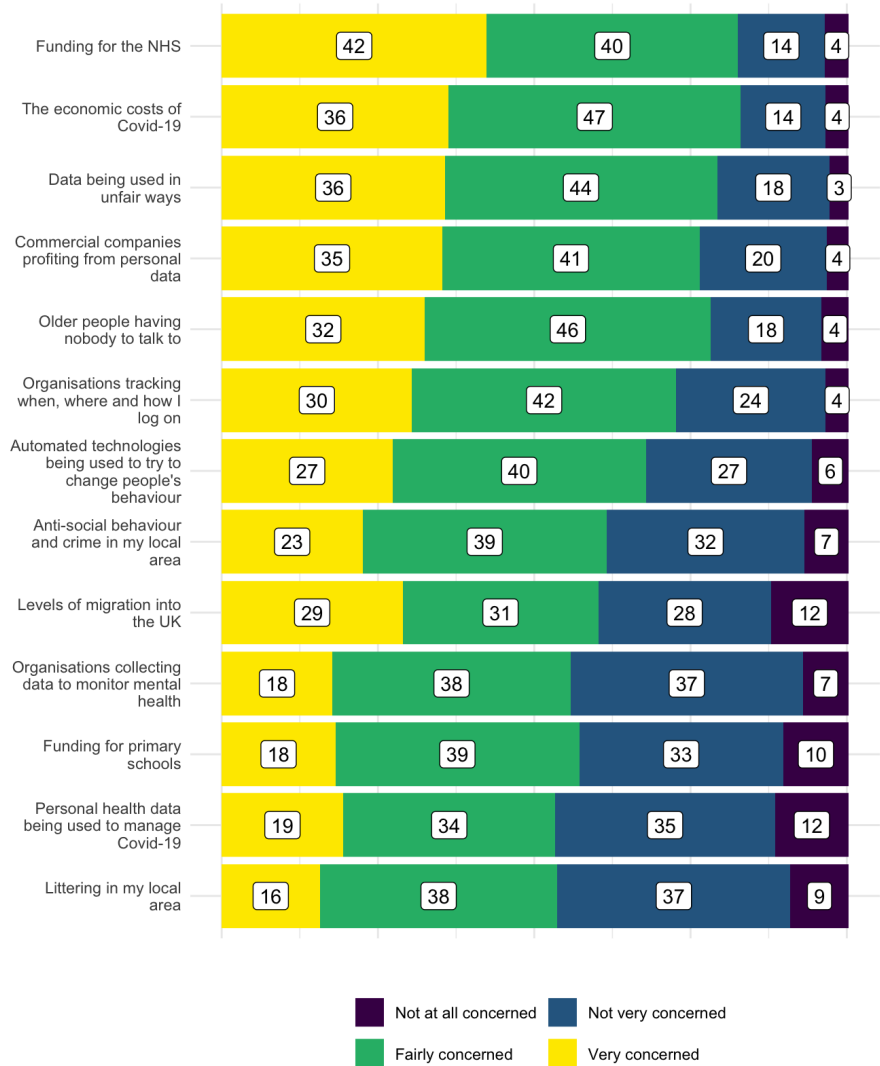
Figure 2: Answers as percentages to: Please indicate how strongly you agree or disagree with the following statements.



profit-making. For example, Figure 2 also shows that 63% of respondents disagreed or strongly disagreed with the statement ‘I support corporate profit-making from personal data.’ 22% of respondents neither agreed nor disagreed with the statement, and only 16% agreed or strongly agreed, with only 5% strongly agreeing.

Elsewhere in the survey, we asked respondents how concerned they were about issues relating to data uses, compared to other everyday concerns. From a list of 13 concerns, in which funding for the NHS and the economic costs of Covid-19 were the biggest concerns, commercial companies profiting from personal data was the 4th biggest concern, as shown in Figure 3. Organisations tracking when, where and how people log on to online systems was the 6th biggest concern, further evidence of concern about commercial companies’ data uses, given that they may undertake such tracking. In these findings, it is striking that concern about commercial involvement in and profiting from uses of data ranks closely to major issues like the Covid-19 pandemic and the ability of public health services to provide the health services that the nation requires.

Figure 3: Answers as percentages to: In general in your daily life, how concerned are you about each of the following?



A final finding from the survey that is relevant to the involvement of commercial companies in public sector data uses comes from the free text fields, in which we invited respondents to comment on their responses to our questions. Here, $\frac{3}{4}$ of respondents who mentioned the NHS Covid-19 Data Store, one of our health data use cases, expressed concern about the involvement of commercial companies. It is also interesting to note that, in 2020, 78% of survey respondents were fairly or very comfortable about their NHS patient data being added to the NHS Covid-19 Data Store. In 2021, the figure was significantly lower at 70%. This was one a small number of cases where there were significant differences between responses to surveys undertaken in 2020 and 2021. This decrease in comfort could result from vaccination roll-out and a feeling that the pandemic was more under control, or it could result from increased awareness of and concern about the involvement of private companies in the Data Store, something which received a small amount of media coverage.

Furthermore, this apparent broad support for gathering data to manage the Covid-19 pandemic is called into question by the fact that, in free text fields, most expressions of concern were about this data use case, and most of these were about the involvement of commercial companies like Amazon Web Services, Microsoft and Palantir in the Data Store. Respondents expressed their concerns by imagining future, worrying scenarios, in which commercial organisations profited from, leaked, misused or sold personal data.

In focus groups and interviews, the involvement of third party or commercial organisations in the public sector data processes that we presented to participants was the main source of misunderstanding or confusion. For example, when told Microsoft, Google, Amazon Web Services and other commercial organisations were providing infrastructure and processing tools for the NHS Covid-19 Data Store, Diane, a white British, heterosexual woman quoted below, asked why they needed to know personal medical information, which suggests that she thought these companies would get direct access to the data in the Data Store.

Why did they feel the need to contact Amazon, Microsoft? What's the benefit? This is supposed to be NHS, so it's supposed to be, in my mind, medical information. Why does Amazon need to know that? Why?

Diane, a white British, heterosexual woman, 55-64, with an annual household income of £40,000-£49,000

We also found concern and confusion about the involvement of third parties in the DWP's online identity verification processes that we discussed with participants. Ruby, a heterosexual, British-Chinese woman, aged 18-24, who has an annual household income of £40,000-£49,000, appeared to believe that data would be transferred from one organisation to another as part of online identity verification, which is not the case. Rather, data held by one organisation is checked, anonymously, against data held in another organisation, as part of the identity verification process.

IN SHORT:

There is confusion and concern about commercial involvement in public sector data uses, which sometimes derives from lack of clarity.

There is broad concern about what happens to personal data, and specific concern about commercial companies profiting from accessing and using such data. Some of the time, concern results from confusion, caused by lack of clarity about the nature of commercial company involvement.

Broader concerns about data uses may also derive from lack of clarity: it's likely that what causes people to want to know more about what happens to their data is an absence of clear information about such matters. More clarity about commercial involvement in public sector data uses may diminish confusion, but it won't necessarily diminish concern because, as we highlight in Section 7 below, it is not the case that the more people understand data uses, the more positive their attitudes towards them are.

6. Unpacking concern

People are not only concerned about public sector data uses. 52% of our survey respondents agreed or strongly agreed that collecting and analysing data can be good for society. This shows that people can recognise the benefits of data uses, whilst also having some concerns about them. Throughout our research, we found that people are supportive of data uses if they perceive them to have ‘public good’ intentions.

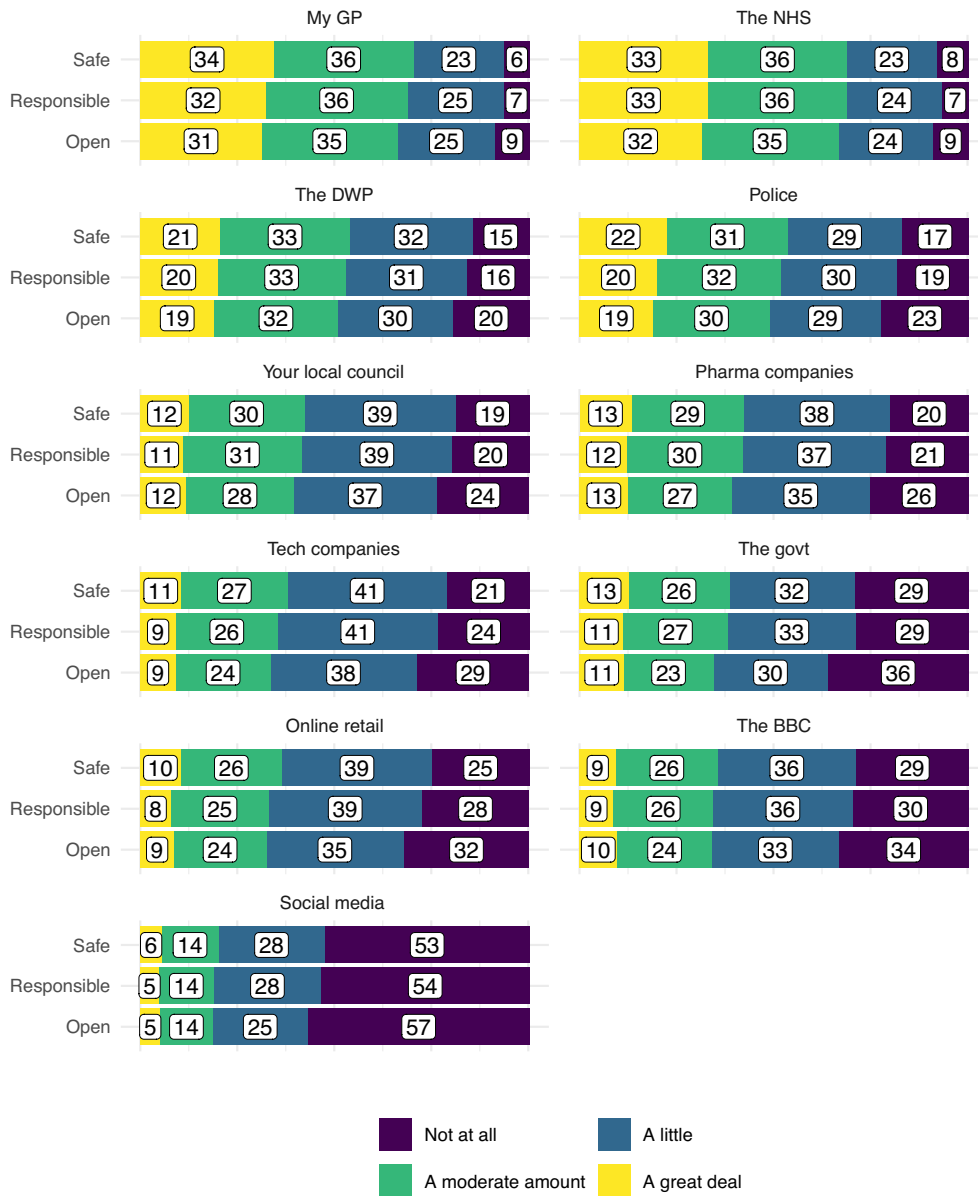
A majority of participants were supportive of the NHS Covid-19 Data Store because its primary purpose was deemed to be in the public interest. But such support is not unconditional. People support data uses for social or public good purposes if these do not appear to be compromised by three factors. The first factor is negative impacts on people from disadvantaged and minority groups, discussed in Section 4 above. The second factor is the involvement of commercial companies, discussed in Section 5 above, which was a major concern for some of our research participants.

The third factor which informs people’s degree of support for data uses is the context in which they take place. In our survey, we asked respondents how much they trusted 11 different sectors or organisations to keep data about them safe, to gather and analyse data in responsible ways, and to be open and transparent about what they do with personal data. We present responses to these questions in Figure 4 below. We found that trust in data uses varies, depending on the organisation or sector undertaking them, not on the data use. In other words, trust in data uses in a given sector or organisation were consistent across the three things we asked about: openness, responsible data use, and transparency. In terms of sectors and organisations, health sector practitioners and organisations, such as GPs and the NHS, were most trusted. Social media, media and technology companies were the least trusted.

As noted above, we also explored how concern about data uses compares with other concerns. From the list of concerns presented to survey respondents and shown in Figure 3, personal data being used to manage Covid-19 was the least concerning issue on the list. This provides further evidence that health-related data uses tend to concern people less than data uses in other sectors. Thus, although some data uses concerned respondents a great deal, others concerned them much less. In short, context matters with regard to people’s degree of concern about data uses.

Demographic differences also matter in relation to concerns about data uses. It’s not the case that all people are equally concerned about all data uses in all contexts. Some data uses in some contexts are more concerning to some groups than others. As discussed in Section 3 above, belonging to a disadvantaged or minority group informed people’s perceptions of data uses. Disabled people were more positive about health data re-use than people who did not have a disability, white people trusted the police’s data uses more than Black, Asian and other racialised people, older people trusted their GP more than the youngest 18-24 age group, but LGBTQ+ people trust health organisations less than heterosexual cisgendered respondents.

Figure 4: Answers as percentages to: How much do you trust <organisation> to: keep data about you safe?; gather and analyse data about you in responsible ways?; be open and transparent about what they do with data about you?



IN SHORT:

Concerns about data uses are context-specific, not universal, and informed by inequalities.

Some people are concerned about some data uses, but not all people, and not all data uses. There is some support for data uses, if they are deemed to be for the public or social good. Context, inequalities, who's involved, and demographic difference play a role in the degree of concern that exists about specific data uses. Therefore, we need to be specific and precise when we talk about public concern about data uses.

7. Understanding and imagining data uses

Understanding data uses is an important prerequisite to developing opinions about them, yet it is rarely the focus of research and discussion, in policy, practice and research circles (research by one of us, Oman 2021, is an exception). This is surprising, given that how data about us is collected, analysed, shared and used is difficult to understand. Difficulties can arise for a number of reasons.

It could be that the data uses themselves are complex, explanations are long and hard to follow, perhaps even by design, or they may not be available at all. As we saw in Section 5, confusion can also arise with regard to the role played by commercial companies or other third party organisations in public sector data uses.

It is therefore vital to consider how people come to understand data uses, given the importance of and difficulties in understanding them. This is especially important given that clear and transparent information about data uses (Pasquale 2015) and improved data literacy (eg the ODI's 2022 mapping of UK government activity on data literacy) are commonly assumed to be the best solutions to misunderstandings of data uses.

In our research, we found that clear information about data uses does not result directly in people understanding them. As discussed in the appendix below, we produced visualisations of the public sector data uses that were the focus of our research, through rigorous, iterative processes, involving a professional visualisation designer and experts in accessible communication. Nonetheless, focus group and interview participants often assumed that there was more to these data uses than what they saw or were told in our discussions with them. As a result, they imagined a number of things about the public sector data uses that we discussed with them.

You know, if the government knows that I'm gay, for example, and a particularly right wing administration comes in, they know where I live, they know where I am. They can make it a legal requirement for companies to share that data with the government. They could very quickly arrest me or whatever. That has crossed my mind. I mean, it seems like a sort of dystopian fantasy but at the same time, it's possible. I mean, it's happened in the past in history and so it's the thing of you don't trust Facebook, but you do trust the government, but why? I mean, it's not as though politicians are morally superior to business people. They're often the same thing.

Matthew, a white, gay, Zambian man with an annual household income of less than £30,000

For example, they imagined:

- What might happen to data in the future, if governance arrangements or prevailing norms change, as seen in the quote from Matthew above.
- Differences between how data uses are said to work in theory and how they actually work in practice.
- The experiences and perceptions of people from disadvantaged or minority groups to which they didn't belong, who may be more negatively impacted by data uses, as seen in the examples in Section 3.

By imagining, we mean building or creating a mental image of something that is not present at the moment of expressing an opinion. We are not suggesting that the things that participants imagined were imaginary – that is, existing only in the imagination. Rather, participants imagined and supposed things to fill in assumed gaps, possibly because historical data misuses, security breaches and lack of transparency led them to assume that there must be information missing from our accounts. As a mechanism to work around the challenges of understanding data uses, participants imagined what happens to data, how processes work and their impacts on others.

The fact that participants often imagined that there was more to data uses than what they saw or were told demonstrates that the link between information and understanding is complex.

Our survey also produced important findings about how people's understanding of data uses relates to their attitudes towards them. Underlying arguments for more transparency about data uses or better data literacy is an assumption that the 'problem' with people's perception of data uses is one of information and understanding. For example, the Royal Statistical Society (2014) identified a 'data trust deficit' which, it was argued, is characterised by limited understanding accompanied by high levels of suspicion. Our research challenges the implicit assumption here that the more people understand data uses, the more positive their attitudes will be. In contrast, we found that the more aware of data uses that people were, the more critical and cautious about them they tended to be.

In the survey, we collected data about different aspects of people's understanding of and attitudes towards data uses. Having done this, we used latent class analysis to classify respondents into a number of different latent classes, using the polCA package in R.

Table 1: Understanding groups, attitudes groups, descriptions and percentages of respondents in each group

Group name	Group description	% of respondents
Understanding groups		
Aware	generally correctly identify true and false statements	25
Believers	generally respond that the statements they are presented with are true	38
Disbelievers	generally respond that statements are false, even when they are in fact true	14
Don't knows	generally state that they don't know the answer to awareness questions	23
Attitude groups		
Critical	strongly disagree with some statements that are positive about data practices, and strongly agree with others that are negative about data practices	36
Cautious	tend to agree or disagree in the same directions as the Critical group, but not strongly	34
Neutral	overwhelmingly respond with 'Neither agree nor disagree'	13
Agree	overwhelmingly either agree or strongly agree with all statements	17

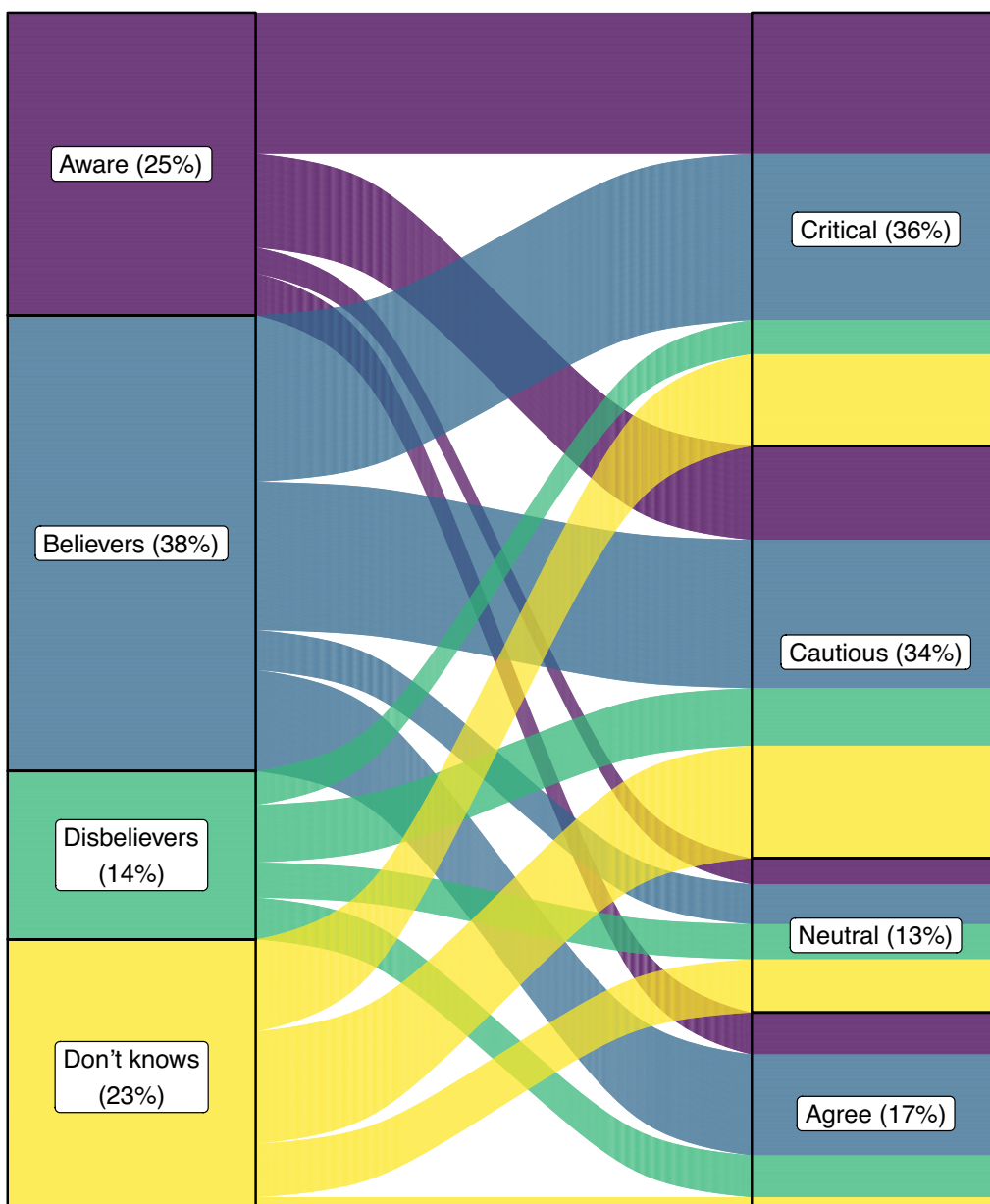
Latent class analysis is a subset of structural equation modelling, used to estimate latent groups from a number of categorical variables. Models were estimated for between 2 and 10 latent classes in each case, with each number estimated ten times, and with 100,000 iterations each time. We identified four ‘understanding’ groups and four ‘attitude’ groups and we classified respondents into one group in each cluster as shown in Table 1.

Figure 5 shows the relationship between understanding and attitude, where the width of bars denotes the fraction of survey respondents who were members of both latent classes

connected by the bars: the thicker the bar, the stronger the relationship. 46% of respondents categorised as Aware were in the Critical group, significantly more than the 36% of the overall sample who were in the Critical group.

Aware respondents were also significantly less likely to be in the Neutral group, at 9% compared with 13% of the overall sample, and slightly less likely to be in the Agree group, at 14% compared with 17% of the overall sample. Thus we found that people who had greater awareness of data uses tended to be more critical and cautious about them.

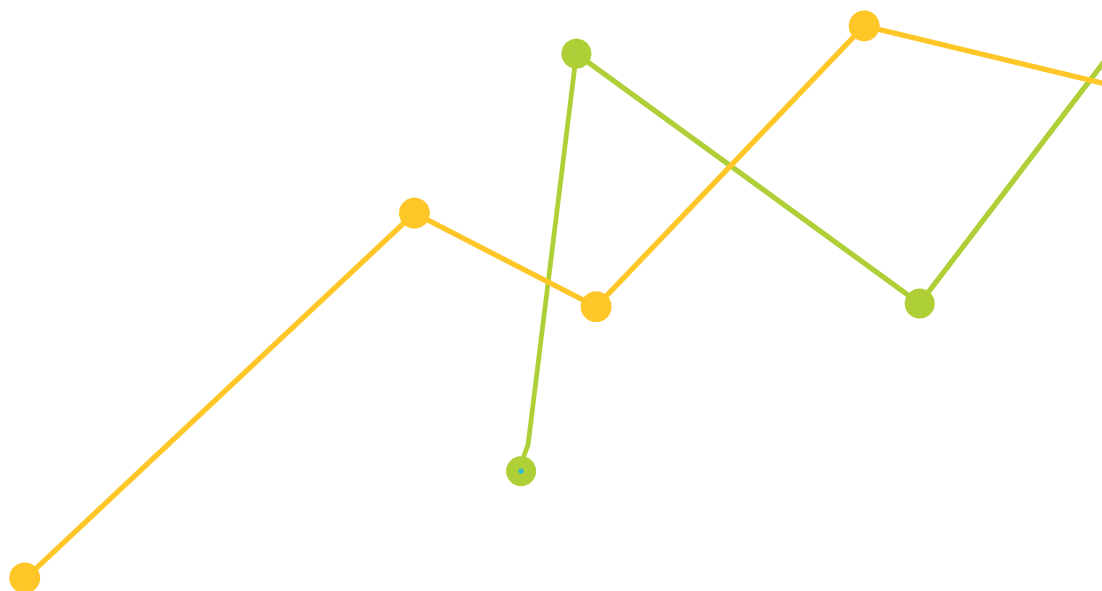
Figure 5: the relationship between understanding groups (on the left) and attitude groups (on the right).



It is important to note that we do not conclude from this finding that less information about data uses should be made available! As noted in Section 5, people want information about what happens to their personal data, for example who has access to data about them and where such data is stored. What is at stake here is not how much people know or are aware, but rather the characteristics of the object of knowledge or awareness. What is needed is not more or less information about data uses, but better data uses.

IN SHORT:

How people understand data uses is complex, is shaped by imaginings, and shapes opinions. The findings we describe above suggest that how people understand data uses is complex, it is shaped by imaginings and, in turn, it shapes opinions. Our findings call into question the proposition that what is needed is more transparency, or better literacy, about data uses. Instead, better data uses are needed. Our findings also point to the importance of researching the processes by which people come to understand data uses.



8. What our findings tell us about communicating about data uses

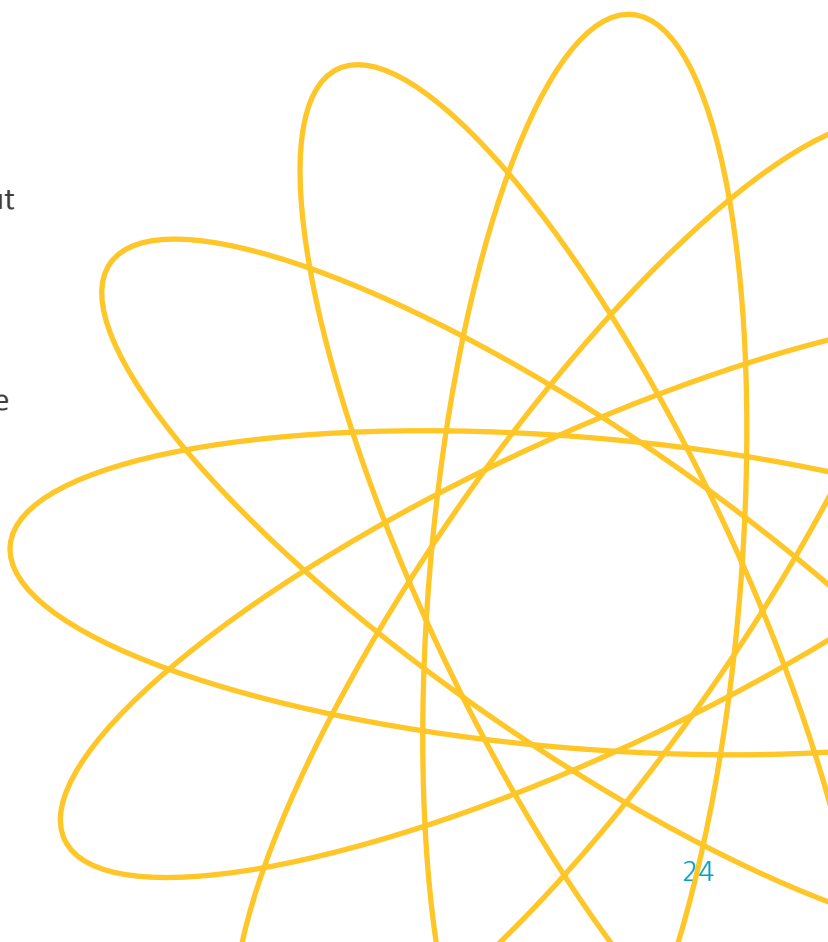
Our efforts to produce balanced and accurate accounts of public sector data uses on *Living With Data* and our findings from discussions of these accounts have led us to a number of conclusions about how to communicate about data uses.

Communication was central to some focus group and interview participants' ideas about what makes a data use acceptable. Some participants said that clear and honest information, which makes it possible to understand what happens to data and the purpose of a data use, is a characteristic of a fair data use. This is a pre-requisite to another characteristic of a good, fair data use for some – being able to choose what happens to personal data. In other words, for a data use to be fair:

- you need to understand it;
- then, you can choose whether to share your data;
- then, you can choose what happens to it.

Putting these findings together with other findings reported above – about concerns about the ways in which data uses might reinforce inequalities, and confusion and concern about the involvement of commercial companies in data processes – suggests that communication about data uses needs to actually communicate useful information to people, not just tick a transparency box. Explanations need to focus on what matters to people. This includes addressing what might go wrong, and responding to the imagined gap between data uses as described on paper and in practice.

It includes addressing possible future as well as actual current uses of data, responding to imagined future changes. And it includes addressing concerns about whether and how some groups might be more negatively affected by data uses than others, responding to people's worries about the experiences of more disadvantaged others. In other words, how data uses are communicated is as important as ensuring they are communicated. Communication should be motivated by a genuine desire to enable understanding, and it should focus on information that is of interest to diverse publics.



9. Conclusions and recommendations

A. Different people have different perceptions of data uses

Different people have different perceptions of different data uses. Stakeholders across policy, practice and research should avoid generalising about public opinions of data uses. Because of these differences, we need to look beyond headline findings about public opinions of data uses.

For data policy-makers and practitioners, this could mean consulting or commissioning evidence reviews, which synthesise the findings from the growing pool of empirical research on this topic. For researchers, it could mean undertaking such evidence reviews. It is also important for all stakeholders to regularly consult diverse publics, because data uses and perceptions of them change, and because structural inequalities influence what different people think of different data uses.

B. Data inequalities can lead to data solidarities

The differences described above relate to members of the public themselves, the data uses in question, their contexts and consequences. They are shaped by structural inequalities: it is not the case that everyone has equal access to or is equally affected by uses of data in data-driven systems.

As a result of unequal experiences of structures and systems, data-driven or otherwise, some groups are more concerned about some data uses than others. For example, Black, Asian and other racialised people are more concerned about what the police do with people's data, and LGBTQ+ people are more concerned about data uses in the health context. These results confirm the findings of other commentators on data and inequality (eg Eubanks 2018 and Noble 2018).

What is more surprising is that these differences were often small, that people from different groups were aware of structural inequalities and concerned about data uses that reproduce or exacerbate them. Our findings about the roles inequalities play in perceptions of data uses reinforce the recommendations in (A) above and (C) and (D) below, and they confirm the importance of putting people at the centre of data policy-making and data practices. This should be done in the following ways:

- in how stakeholders think about public perceptions of data uses (discussed in (A) above);
- in how they communicate about data uses (discussed in (C));
- and in data uses themselves, which may need to change or be desisted in response to this finding, as discussed in (D).

C. Communication should enable genuine understanding & respond to what matters to people

The aim of communicating about data uses and data-driven systems should be to enable genuine understanding in the people whose data feeds such systems. Communicating about data uses should not be a box-ticking exercise. In other words, it's not enough just to communicate. How data uses are communicated is important. Communication about data uses needs to respond to what matters to people and address their concerns.

We found that this includes: concerns about what might go wrong in a data-driven system, what might change in the future, and how disadvantaged communities might be more negatively or adversely affected than other groups. Various stages of data system design and implementation lend themselves to forms of communication that foster understanding, as we learnt from visualising data uses from prototypes to fully operational systems.

Our focus group and interview participants felt that our visualisations of public sector data uses were clear and helpful. These could be used as models for communicating visually about data uses, for those people for whom visuals are accessible. For people with visual impairments, alternative, text-based explanations are also needed. These recommendations require some understanding of the audiences to whom such communication is targeted, and thus they build on the recommendations in (A) above. Our recommendations about communication are especially relevant to policy and practice stakeholders who implement data uses. They are also relevant

to researchers exploring people's opinions of data uses, because it is important to ensure that people understand data uses prior to eliciting their opinions about them.

D. Change or desist data uses that are not human-centric, eg that discriminate or from which private companies profit

However, good communication alone is not enough. What is needed most is better data uses, not more or clearer information about them alone. We found widespread concern about data uses, consistent with the findings of other surveys into public perceptions of data processes and systems. This consistent finding communicates a strong message to data policy-makers and practitioners about public dissatisfaction with existing data uses. Furthermore, the people who know most about data uses are the most concerned about them.

This again also suggests that data uses, or aspects of them, are concerning. It is likely that if data uses continue unchanged, the public will continue to be concerned, regardless of how effectively communication enables understanding. In short, data uses need to change, in order for there to be greater public support for or confidence in them. Good communication is a component of a good data use, but good communication alone won't change a concerning data use into a good one.

What do good data uses look like? The answer to this question is not straightforward, because of the differences in perspectives, contexts and consequences that we highlight in this report. 'Good' needs to be defined on a case-by-case

basis, in consultation with diverse publics, whose views may change over time. With that said, it's clear from our research that sharing data that has been gathered for pro-social or the public good with commercial companies who will make a profit from it is widely disliked. Our survey respondents trusted social media and tech companies with their data the least, so involving them in the provision of public sector data infrastructures or services is likely to cause concern. Therefore, public sector data practitioners like the DWP, BBC and NHS, whose data uses were the focus of our research, should consider alternative ways of delivering data-driven services. This will not be easy, as global technology companies monopolise the provision of particular technologies and technical infrastructures, but it is not impossible. Such changes to the data ecosystem may produce the changes in public opinion that policy-makers and practitioners are keen to see, such as greater support for or confidence in data uses.

In short, data uses themselves need to change, so they eliminate harms and are in the public or social interest. Sometimes, in order to do these things, specific data uses need to stop, such as those which are ostensibly pro-social but from which commercial companies profit, or those that systematically discriminate against already disadvantaged groups. In other words, much more is needed than simply acknowledging differences and getting the communication right. Of course, if data uses overcome inequalities and are stripped of aspects which concern people, then public perceptions of them will probably improve.

E. Future research is needed on i) the role of commercial companies in public sector data uses & ii) the role of imagining in public understandings of data uses

Our findings on *Living With Data* also point to areas for further research. The involvement of third party, commercial companies caused both confusion and concern. Is it possible for them to be involved in the provision of public sector data-driven services in ways which reduce or eliminate concern?

Are some forms of involvement less concerning than others (eg the provision of hardware, storage capacity, processing, analytics)? Or is their involvement in and profiting from ostensibly pro-social data uses consistently seen as problematic? Are some (types of) companies more concerning than others? We also found that people often imagine that there is more to these data uses than what they see or are told, despite efforts to communicate about data uses appropriately and accessibly. Are there forms of communication that minimise imaginings, is that desirable, or are imaginings productive, and if so, in what ways? Addressing these questions requires researching how people understand data uses, a gap in research into public perceptions of data uses to date. These recommendations are aimed at social researchers.

F. Who our recommendations are for

Our findings and recommendations are relevant to the areas of policy and practice which we reference in Section 2 above and in which we situate our research: the Data Act, the Digital Strategy, and strategy and guidance on specific types of data, such as health.

They are also relevant to the three national public sector organisations on whose data uses we focused in our research: NHS, BBC, and DWP. Because data uses and public perceptions

of them are diverse, context-specific and have varied consequences, the precise ways in which our recommendations are implemented should be tailored to attend to these differences. As we note above, what better data uses look like needs to be defined on a case-by-case basis, in consultation with diverse publics. We summarise our recommendations and who they target in Table 2 below.

Table 2: summary of our recommendations and who they are relevant to

LWD Recommendations	Who are the recommendations targeted at?		
	Policy-makers	Data practitioners & professionals	Researchers
CHANGE HOW WE THINK & TALK ABOUT PUBLIC PERCEPTIONS OF DATA USES AND ACKNOWLEDGE INEQUALITIES AND DIFFERENCES			
Avoid using phrases like ‘the public thinks X about data uses’	✓	✓	✓
Be specific: whose views, about which data uses, in which contexts?	✓	✓	✓
Recognise the value of evidence reviews rather than single or standalone studies about public perceptions of specific data uses	✓	✓	✓
Regularly consult diverse publics on their views of data uses	✓	✓	✓
Understand how specific data uses might disadvantage or discriminate against certain groups, in order to change them	✓	✓	✓
CHANGE COMMUNICATION			
In communications about data uses, address the things that concern people: what might go wrong, what might change in the future, how people from minority or disadvantaged groups might be affected	✓	✓	
Consider using visuals to communicate data uses, in addition to clear text	✓	✓	
CHANGE OR HALT CERTAIN DATA USES			
Change or consider not engaging in data uses that concern people. For example, don’t share data initially gathered for pro-social or public good motives with commercial companies who will make a profit from it.		✓	
Change or consider not engaging in data uses that disadvantage or discriminate against certain groups		✓	

10. References

- Bates J., Lin Y-W., Goodale P. (2016) 'Data journeys: Capturing the socio-material constitution of data objects and flows'. *Big Data & Society*. [10.1177/2053951716654502](https://doi.org/10.1177/2053951716654502).
- Centre for Data Ethics and Innovation (CDEI) (2021). Britain Thinks: Complete Transparency, Complete Simplicity. <https://www.gov.uk/government/publications/cdei-publishes-commissioned-research-on-algorithmic-transparency-in-the-public-sector>
- Dencik, L. and Cable, J. (2017). Digital citizenship & surveillance, *International Journal of Communication*. <https://ijoc.org/index.php/ijoc/article/view/5521>
- Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police and punish the poor*. St Martins Press.
- Fox, J. et al (2021) 'Open Science, Closed Doors? Countering Marginalization through an Agenda for Ethical, Inclusive Research in Communication', *Journal of Communication*, 71 (5), pp 764–784, <https://doi.org/10.1093/joc/iqab029>.
- Geburu, T. et al (2021). Datasheets for Datasets. *Communications of the ACM*, December 2021. <https://cacm.acm.org/magazines/2021/12/256932-datasheets-for-datasets/fulltext>
- 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>
- Kalluri, P. (2020). Don't ask if artificial intelligence is good or fair, ask how it shifts power. *Nature*, 7 July 2020. <https://www.nature.com/articles/d41586-020-02003-2>
- Katwala, S. (2021) *Race and opportunity in Britain: finding common ground*, British Future, https://www.britishfuture.org/wp-content/uploads/2021/03/Race-and-opportunity-in-Britain.Final_.30.3.21.pdf.
- Kennedy, H. et al (2020) *Living With Data Literature Review*. <https://livingwithdata.org/resources/living-with-data-literature-review/>
- Malik, S., Ryder, M., Marsden, S., Lawson, R. and Gee, M. (2021) *BAME: A report on the use of the term and responses to it*, BBC, <https://bcuassets.blob.core.windows.net/docs/csu2021325-lhc-report--bbchighres231121-132836254614117870.pdf>.
- Medina Perea IA, Bates J & Cox A (2019) [Using data journeys to inform research design: Socio-cultural dynamics of patient data flows in the UK healthcare sector](https://doi.org/10.1145/3287560.3287596). *iConference 2019 Proceedings*.
- Mitchell M. et al. (2019). Model cards for model reporting. Proceedings of the conference on fairness, accountability, and transparency (FAT* '19), Association for Computing Machinery, New York, NY, USA. <https://doi.org/10.1145/3287560.3287596>
- Nikunen, K. (2019). *Media Solidarities: emotions, power and justice in the digital age*. Sage University Press.
- Noble, S. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York University Press.
- Oman, S. (2021). *Understanding Well-being Data: Improving Social and Cultural Policy, Practice and Research*. Palgrave Macmillan. Available at: <https://link.springer.com/book/10.1007/978-3-030-72937-0>.

Open Data Institute (ODI) (2022). Why data literacy matters. <https://theodi.org/article/data-literacy-and-the-uk-government-report>

Pasquale, F (2015). The Black Box Society: the Secret Algorithms That Control Money and Information. Harvard University Press.

Pew Research Center (2019) 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/>

Royal Statistical Society (2014). Royal Statistical Society research on trust in data and attitudes toward data use / data sharing, 2nd October 2014. <https://www.statslife.org.uk/images/pdf/rss-data-trust-data-sharing-attitudes-research-note.pdf>

Yates et al (nd). Me and my big data: understanding citizens data literacies research report. <https://www.liverpool.ac.uk/media/livacuk/humanitiesampsocialsciences/meandmybiddata/Understanding,Citizens,Data,Literacies,Research,,Report,Final.pdf>

11. Appendix: further information about methods & evidence review findings

i. Evidence Review

We conducted a review of original empirical research published in grey and academic literature, that focuses on public perceptions of, attitudes towards and feelings about data uses and related phenomena such as AI and facial recognition technologies. Our review was published in May 2020, and focuses on literature which was published between 2015 and 2019.

We carried out a systematic search of online academic research databases and a manual search, that began with literature with which we were already familiar, and then snowballed out. We excluded a) literature about children’s understandings and perceptions of data practices because this is a specialist area beyond our remit, and b) literature focused on the health domain because high quality syntheses of literature focusing on this domain already exist. The grey literature we reviewed focused on the UK, whereas academic literature was international. We summarise the main findings of the evidence review and the questions that they led us to ask in our subsequent research in Table 3 below.

Find out more about what we did, how we did it, and what we found in our [evidence review summary](#) or our [full evidence review report](#).

Table 3: Main findings of evidence review and how they informed our research

Evidence review finding	How it informed our research
People have some understanding of data uses Emotions play a role in perceptions of data uses	What does it mean to understand a data use? How is understanding measured or interpreted? Is understanding the best term, or is awareness or knowledge better? What role do emotions play in understanding or awareness of data uses?
People are concerned about data uses, but they are not only concerned Context matters, people trust some sectors more than others with their data	What factors influence whether people are concerned about data uses or not? This means considering: sector or organisational context; consequences, for selves and for others; how concerns about data uses compare with other concerns; whether concerns change over time.
Differences & inequalities appear to inform perceptions of data uses Opinions differ amongst experts regarding how data uses need to be improved	How do demographic differences and social inequalities inform perceptions of data uses? This means thinking not only about whether different people perceive data uses differently, but what diverse members of the public think about the relationship between data uses and inequalities. What do diverse members of the public think about how data uses need to be improved?

ii. Documenting specific data uses in public sector organisations:

We produced accounts of specific data uses which we presented to focus group and interview participants for discussion. We selected cases from the public sector because public sector data uses shape everyday life experiences, and yet had received less attention than high profile commercial data uses at the time of our research.

As stated in Section 2, we focused on welfare, media and health data uses. To produce accounts of data uses in the first two domains, we partnered with the government Department for Work and Pensions (DWP) and the British Broadcasting Corporation (BBC). The BBC and DWP data uses were therefore selected by our contacts within these partner organisations. For the latter domain, health, we produced an account of a data use based on information in the public domain, and another account based on research that one of us, Itzelle,

had undertaken into National Health Service (NHS) data flows (Medina Perea et al 2019).

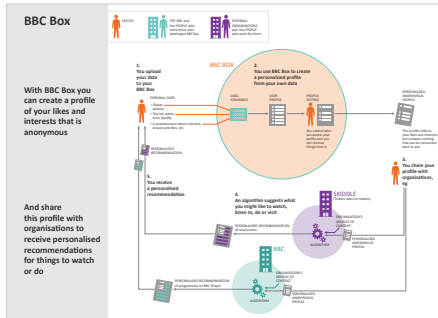
To produce accounts of data uses, we deployed some of the techniques used by one of us, Jo, (in Bates et al 2016) to map data flows and frictions, as part of an approach they call Data Journeys. This process involved developing detailed knowledge of each data use, through interviews and textual analysis of partner organisation and publicly available documentation. In the data uses which were the focus of our research, data did not always flow, or go on a journey – indeed, some data uses aimed to limit the movement of personal data in order to give people more control over data about them.

Figure 6 below shows small versions of the visualisations we shared with focus group and interview participants. Full-size visualisations of data uses can also be found on the [Public sector data uses](#) page of our website. Figure 6 also includes a short description of each data use, and proposed benefits and potential harms for each one, which we described to participants in order to ensure that they had some understanding before embarking on discussions of them.

Figure 6: Short descriptions, thumbnail images, benefits and harms of the six data uses which were the focus of our research

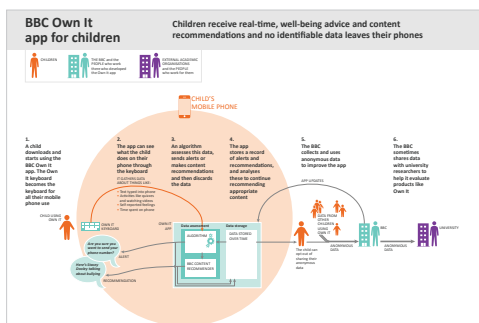
BBC Box, a prototype which pulls together data about what users watch or listen to and gives them control over who has access to this data.

- Proposed benefit: people control their own data & who has access to it; people receive personalised recommendations
- Potential harm: whether data is secure if all in one place; whether individuals have the time & expertise to manage their own data; recommendations systems can recommend a narrow range of things, and people using them are not exposed to a range of issues of perspectives. Some people say this is not good for democracy



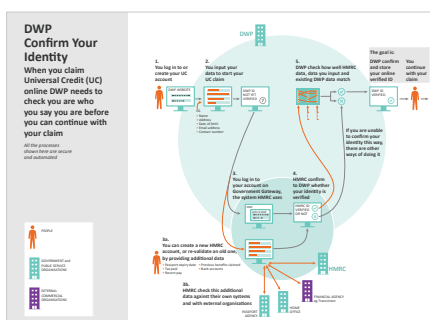
BBC Own It, a free app designed to support, help and advise children when they use their phones to chat and explore the online world, without adult supervision.

- Proposed benefit: children receive personalised wellbeing advice & support without identifying data leaving their device
- Potential harm: interventions in children’s wellbeing take place without parent/carer knowledge; potential for anonymous data to be shared with researchers, which may not be clear in data agreement; automated, makes assessments of & recommendations about children’s wellbeing, without the involvement of a parent/carer



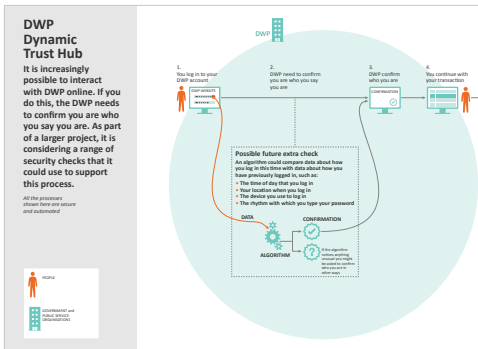
DWP Confirm Your Identity, an identity verification process for Universal Credit payments which makes it possible to confirm identity online.

- Proposed benefit: don’t need to confirm identity in person/with paper docs
- Potential harm: requires people to have HMRC account, passport, bank account, financial record, so excludes people with complex lives; negative consequences for people whose identities can’t be verified



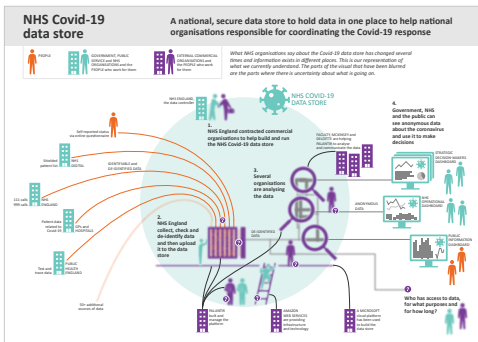
DWP Dynamic Trust Hub, which explored a range of issues to enhance identity verification, including technology integration and possible security checks.

- Proposed benefit: don't need to confirm identity in person/with paper docs & extra layer of security to the identity verification process
- Potential harm: requires people to have HMRC account, passport, bank account, financial record, so excludes people with complex lives; negative consequences for people whose identities can't be verified; inaccuracies in automated checks (eg people may not own their own devices) which may be taken as a proxy for unverifiable identity



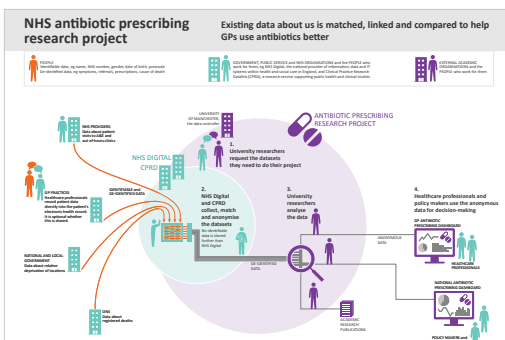
NHS Covid-19 Data Store, a national data store to help organisations responsible for coordinating the Covid-19 response.

- Proposed benefit: data gathered together in one secure place to inform the national response to the pandemic
- Potential harm: lack of clarity about who has access to data & the contractual arrangements with commercial organisations; commercial companies may be able to use data from the data store to develop their technologies and potentially increase their profits



NHS antibiotic prescribing research project, exploring ways to address the public health crisis of antibiotic resistance.

- Proposed benefit: addresses public health crisis of antibiotic resistance
- Potential harm: requires the use of patient data; whose data is shared (could it be children's?); who the data is shared with (could it be pharmaceutical companies, who could eventually profit from it?)



iii. Survey of public attitudes to data uses:

We carried out two waves of a survey of people's attitudes to data uses, the first at the end of 2020, the second at the end of 2021, each with 2000 adult respondents from the UK. Data was collected via the web survey platform Qualtrics, which recruited participants as well as hosting the survey.

The sample was recruited to be nationally representative of adults in the UK, in relation to gender, age, income and ethnicity. Numbers of respondents with disabilities were also nationally representative. There was additional recruitment (or 'boosts') of people born outside the UK, LGBTQ+ people, and people in receipt of the UK's main welfare benefit, Universal Credit, to ensure these groups were large enough for analysis and that the views of people in these groups were represented in our study. We did not oversample people with low educational qualifications, due to the complexity of the educational qualification infrastructure in the UK. As a result, our respondents have slightly higher qualifications than is nationally representative. Respondents whose answers suggested they were not paying attention – for example, by entering nonsense in free text fields – were filtered from the sample both during and after the data collection process, leaving an overall sample size of 2,000 in each wave.

We built on our analysis of survey questions that formed part of our evidence review in order to design our own survey. Sometimes this meant asking questions that had been asked in other surveys, for example about internet usage or awareness of data uses. Sometimes it meant revising such questions so they were more specific to our aims or the changing data landscape, for example in relation to awareness of and general attitudes to data

uses, trust in sector or organisations' data uses and concerns about data uses as they compare with other concerns. We devised question about attitudes to the specific public sector data uses that were the focus of our research.

More detail about what we did, how we did it, and what we found can be seen in our [survey summary](#) or our [full survey report](#).

iv. Focus groups and interviews:

112 people participated in our focus groups and interviews.

Because we were interested in the relationship between social inequalities and attitudes to data uses, we recruited demographically diverse participants. Demographic information was gathered via a 'Tell Us About You' survey, which 100 of 112 participants completed prior to the interviews and focus groups.

We gathered information about age, ethnicity, gender, sexuality, country of birth, nationality, employment status, household income, education, whether English was a first language, long-term conditions. We also asked participants about the newspapers that they read and their degree of interest in politics. Participants had the opinion to choose 'prefer not to say' to any of the questions. They could also choose to provide some information in free text fields, about their job title, the ages of children in their households who were under 18, their country of birth and the newspapers they read. We gathered this information to help us interpret the research data that we gathered, because previous research had suggested that these factors may play a role in shaping attitudes.

Our research took place during the Covid pandemic. Lockdown conditions made it difficult to recruit participants from certain disadvantaged communities who may have limited access to the tools and skills needed to engage in virtual meetings. Nonetheless, we managed to recruit a sample of which:

- 44% of participants were Black, Asian or other racialised people;
- 41% had household incomes of less than £19k a year;
- 37% had a long-term condition;
- 20% were LGBTQ+;
- 14% were 65 or older;
- 46% had two or more of these characteristics;
- 15% had none of the above characteristics.

We used definitions from the 2010 Equality Act to identify whether participants had disabilities, asking if they had any physical or mental health conditions or illnesses lasting or that were expected to last 12 months or more. Also following the Equality Act, we use the phrase 'had a long-term condition' to refer to participants with disabilities, although we recognise that conflating health and disability is sometimes challenged by disability activist and advocacy groups.

For the question about country of birth, the options were: UK; other European country; other non-European country; prefer not to say. Some participants chose to tell us their country of birth in free text fields. In response to the question about ethnicity, we presented participants with the 19 categories used in the Census in England, which we then aggregated into: White British; White Other; Black; Asian; Mixed ethnicities/other; prefer not to say. We recognise that there are more ethnicities than those we use in this report. We chose to use these six categories because to use more categories than this would not be productive in our analysis of our

112 participants we aimed to recruit for the focus groups and interviews. In this report, following [Sobande and others](#), we use the term 'Black, Asian and other racialised people' to refer to participants in the Black; Asian; Mixed ethnicities/other categories. We do this because we acknowledge criticisms that the term BAME (= Black, Asian and Minority Ethnic) is reductive ([#BAMEOver](#); see also Katwala 2021; Malik et al 2021). We also acknowledge that there is no one obvious term to use to describe participants who experience racism.

We describe as LGBTQ+ participants who answered 'no' to the question 'is your gender the same as the gender you were assigned at birth' and who described themselves as bisexual, asexual, lesbian/gay, queer or other. We recognise that there will be differences across these groups, but as above, given our relatively small number of participants, we felt that this grouping was most appropriate for our analysis.

Recruitment strategies included:

Social media pages/accounts: We posted messages inviting people to participate in our research on various social media groups and pages. We attached images to these posts which summarised key information about the project, and we accompanied these images with further details in the caption. We posted on:

- Neighbourhood-focused Facebook groups
- Community-focused Facebook groups (eg groups for Latin Americans, LGBTQ+ groups)

We also sought to recruit via the *Living With Data* Twitter account.

Charities and community organisations: We contacted relevant charities and organisations which worked with or had connections to groups of people to whom we wished to talk. Once contact with a gatekeeper at an organisation was established, the gatekeeper would either organise a focus group for us or

circulate details about our research within their community, via email, social media or word of mouth. People interested in talking to us would then contact our research team directly.

Hobby/interest groups: we contacted a large variety of groups that gathered together based on their interests. This included religious, activist and sports groups, and musical or creative classes. As with charities and community organisations, once contact with gatekeepers was established, they either organised a focus group for us or circulated details about our research.

Other: We leafleted targeted homes, such as council-owned tower blocks, once lockdown restrictions were eased, in order to attract participants in areas populated by communities with whom we wished to talk. We used local council data to identify suitable areas. We also used snowballing to recruit participants. In other words, sometimes participants would pass on project information to people within their own networks, who they thought would also be interested in taking part in our research.

We draw on some of these findings in this document. More detail about what we did, how we did it, and what we found can be seen in our [focus group and interview reports](#).

12. Acknowledgements

The *Living With Data* team would like to thank its funder, The Nuffield Foundation, for its support. The team would also like to thank the *Living With Data* advisory group for their input into all aspects of our research, especially Gaia Marcus from The Cabinet Office, Aidan Peppin from the Ada Lovelace Institute, Adrian Price from the Information Commissioner's Office, and Sam Smith from medConfidential for their frequent and very helpful insights and support.



Living With Data is funded by The Nuffield Foundation. 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-founder of the Nuffield Council on Bioethics and the Ada Lovelace Institute. The Foundation has funded *Living With Data*, but the views expressed are those of the authors and not necessary the Foundation.