Public perceptions of data uses

Summary version of *Living With Data* Report on Qualitative Research

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

July 2022

WHO WE ARE AND WHAT WE DID

Living With Data is a research project funded by The Nuffield Foundation. It aims to understand people's perceptions of how data about them is collected, analysed, shared and used, and how these processes could be improved. We use the term 'data uses' as a short and accessible way of talking to people about these processes.

Concern about the potential harms of data uses means there is a lot of interest in ensuring that data 'works for people and society' (the aim of the Ada Lovelace Institute). But people's perceptions of data uses are rarely central to research, policy and practice, limiting efforts to achieve this aim.

To address this problem, we carried out interviews and focus groups with 112 adults in the UK from November 2020 to September 2021. We produced accounts and visualisations of specific public sector data uses in welfare, health and public service media to discuss with participants.

These can be found, along with a longer version of this summary report and other publications on our *Living With Data* website. This document summarises our main findings.

We recruited diverse participants, including significant numbers of people from disadvantaged or minority groups:

- **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.

FAIRNESS AND INEQUALITIES: WHAT 'GOOD' DATA USES LOOK LIKE

To understand public perceptions of better, improved or 'good' data uses, we asked our participants what **fair data uses** look like to them. In response to these questions and in our broader conversations with participants, we found that **the relationship between data uses and social inequalities** was an important factor in their reflections about good data uses.

'Good' or fair data uses:

Fair data uses are understood to combine a number of different things:

'Public good' purposes or **'social good' motivations** make data uses fair, whereas data uses that had other kinds of motivations or purposes were seen as unfair;

Data uses which discriminated against disadvantaged or minority groups, or which these groups might be excluded from accessing **were seen as unfair**;

Clear and honest information, that makes it possible to understand what happens to data and the purpose of a data use, is a characteristic of a fair data use;

Being able to choose a) to share data once you understand how it will be used and b) what happens to your data are characteristics of fair data uses.

> " In terms of the public sector, I think, if they are not led by profit and they're led by trying to provide certain services, I think that changes it. [...] I guess it's just about what the motivation is behind them doing certain things, and if that's fair, then the whole process should be fair.

Mbali, a black, British-African, heterosexual woman, with an annual household income £30,000-£39,000

Differing understandings of the term fairness made it hard for some participants to talk about what fair data uses look like. But most participants wanted data to be used in reasonable ways. Two participants asked what fairness meant and why it mattered, and then went on to indicate that they wanted data uses to be ethical and 'scrupulous' (Jim, a British-born participant aged over 65). Participants wanted data uses that avoided harms, even if they didn't see the relevance of using the term fairness in this context.

There's a relationship between thinking about fairness and thinking about inequality. Some of our participants felt that data uses were fair if they did not affect different groups unequally. If data uses negatively affected or discriminated against already disadvantaged or minority groups, they were seen as unfair. Participants wanted data uses not to discriminate.

Concern about inequalities and data uses

We found widespread concern about the potential for data systems to have negative consequences for people from disadvantaged and minority groups, for example because they are hard for some groups to use or engage with. Without explicitly using the term, a lot of participants appeared to be aware that data uses can reinforce inequalities, and that some data uses are more likely than others to deepen inequalities.

Belonging to a disadvantaged or minority group appeared to inform participants' perceptions of data uses and what they said about them, in relation to: education; economic status; age; dis/ability; gender identity and sexuality; English as an additional language; race and ethnicity; and the intersections of one or more of these characteristics. We are not suggesting that there is a direct correlation between belonging to a demographic group and attitudes to data uses. Our point is that demographic characteristics shape life experiences and in turn, those experiences shape perceptions of data uses.

> " 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 Turkish-born, heterosexual woman, with an annual household income of less than £10,000

A number of 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. Often, participants from one disadvantaged or minority group were concerned about the effects of data uses on another disadvantaged or minority group. In short, there was collective concern about the consequences of data uses for people from disadvantaged and minority groups.

> " 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, with an annual household income of £50,000-£69,000

Participants were most concerned about the possibility of welfare data uses reinforcing inequalities, compared to the other public sector data uses we discussed with them. Data-driven discrimination was particularly concerning in this context because people who are disadvantaged by structural inequalities often depend on welfare services. A small number of participants raised concerns about inequality in discussions about public service media and health data uses. These tended to relate to access and accessibility or the circulation of data beyond its original context, rather than the specific data uses themselves.

HOW PEOPLE UNDERSTAND DATA USES

How data about us is collected, analysed, shared and used is difficult to understand. This is because data uses are complex, explanations can be long and hard to follow, often by design, and sometimes they are not available at all. Greater transparency or improved data literacy are sometimes proposed as solutions to this problem, which assumes that more and better information and skills will improve understanding. But how do people come to understand data uses? Understanding is an important prerequisite to developing opinions about data uses, yet it is rarely the focus of discussion in policy, practice and research circles. On *Living With Data,* we found that people work around the challenges of understanding data uses by **imagining** what happens to data, how processes work and their impacts on others. People also draw on their **everyday experiences** to understand data uses, and sometimes, they don't understand them.

The role of imagining in understandings of data uses

We found that imagining played an important role in participants' perceptions of data uses. 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, supposed and assumed things to fill in assumed gaps. The fact that participants often imagined that there was more to data uses than what they saw and were told suggests that the link between information and understanding is more complex than is sometimes assumed. Clear and transparent information about data uses does not always result directly in understanding. We think that historical data misuses, security breaches and lack of transparency may lead participants to assume that there must be information missing from accounts of data uses, and they imagined to fill in real or imagined gaps.

Participants imagined:

1. **What might happen to data in the future,** for example if governance arrangements or prevailing norms change. Data policymakers and stakeholders often advocate for transparency about data uses as they operate in the present, but people are concerned about possible future uses of gathered data, which suggests awareness that data uses change over time. For example, some participants imagined possible future data misuses by commercial companies involved in the provision of healthcare data systems. We suggest that this is because of the lack of clarity about their access to personal health data and the value attached to health data by individuals and by corporations.

2. **Differences between how data uses are said to work in theory and how they actually work in practice.** Participants sometimes looked beyond what they are shown or told about data uses, imagining a gap between how data uses are said to work 'on paper' and how they work in real life, sometimes drawing on their own everyday life experiences.

3. **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 than participants themselves would be.

> " 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 [laughs].

Matthew, a white, gay man born in Zambia, with an annual household income of less than £20,000-£29,999

> " You can't get the genie back in the bottle [...] Once the data's out there and shared and used, then you can't recapture that, you can't sort of regain, you know, the control over it.

Lewis, a white, heterosexual man born in the UK, with an annual household income of £50,000-£60,000

What helps people understand data uses, and what doesn't

Participants drew on their everyday life experiences to 'fill the gaps' between the accounts of data uses that we shared with them and what they imagined happens in practice or might happen to data in the future. Everyday experiences were an important resource that participants drew on in order to understand the specific public sector data uses that we discussed with them. For example, one couple questioned the collection of fingerprint data at their children's small kickboxing club. Another participant noticed that every time he went to an appointment for a Covid vaccine trial in which he was participating, he was told that the uses to which his data will be put have changed and asked to consent anew.

Talking about their everyday experiences instead of the data uses that we presented to them could be seen as participants going off on tangents. Instead, we suggest that linking our questions and examples to their everyday lives, participants talked about data uses in ways that connected them to their realities. Participants' observations about the role of data in their everyday lives sometimes served as opportunities for reflection on the politics of data uses or for small-scale acts of agency.

But sometimes people don't understand data uses. The involvement of third party or commercial organisations in public sector data processes was particularly confusing for participants. Lack of clarity about contractual arrangements leads to confusion, which in turn leads to concern.

REFLECTIONS ON TALKING TO PEOPLE ABOUT DATA USES

We made a choice to share accounts and visualisations of specific data uses with participants, in order to make details about potentially unfamiliar data processes accessible to them. We did this because we believe that showing and telling people about data uses enables them first to understand them and then to develop and express opinions about them. We told participants about one proposed benefit and one potential harm of each data use, because people can't assess potential benefits and harms if they don't know what they are. The particular accounts and visualisations and the specific data uses that we used in our research shaped our findings.

Our decision to work in this way has helped us think about what 'good' transparency about data uses looks like. Participants said that the visualisations we showed them helped them to understand the data uses we discussed with them, and they felt that organisations communicating about their own data uses could adopt a similar approach. We believe that good transparency communicates meaningful information about data uses to diverse publics, rather than being a PR exercise or designed to tick a transparency box. We think that good transparency should aim to contribute positively to enhancing people's agency and reducing power asymmetries in relation to data uses. We have developed a set of principles for good transparency, or what we call critical transparency practice, which we write about in a journal article.

It's important to remember that perceptions of data uses are not static. What happens to data can change, decisions about how and where to share data change, and public attitudes change in response to these other, contextual changes. Feelings play an important role in understandings and perceptions of data uses, in imaginings and in reflections on fairness and on social inequalities. Feelings also change, as people learn more about data uses in a focus group conversation, or as proposed benefits and potential harms of data uses are revealed to them.
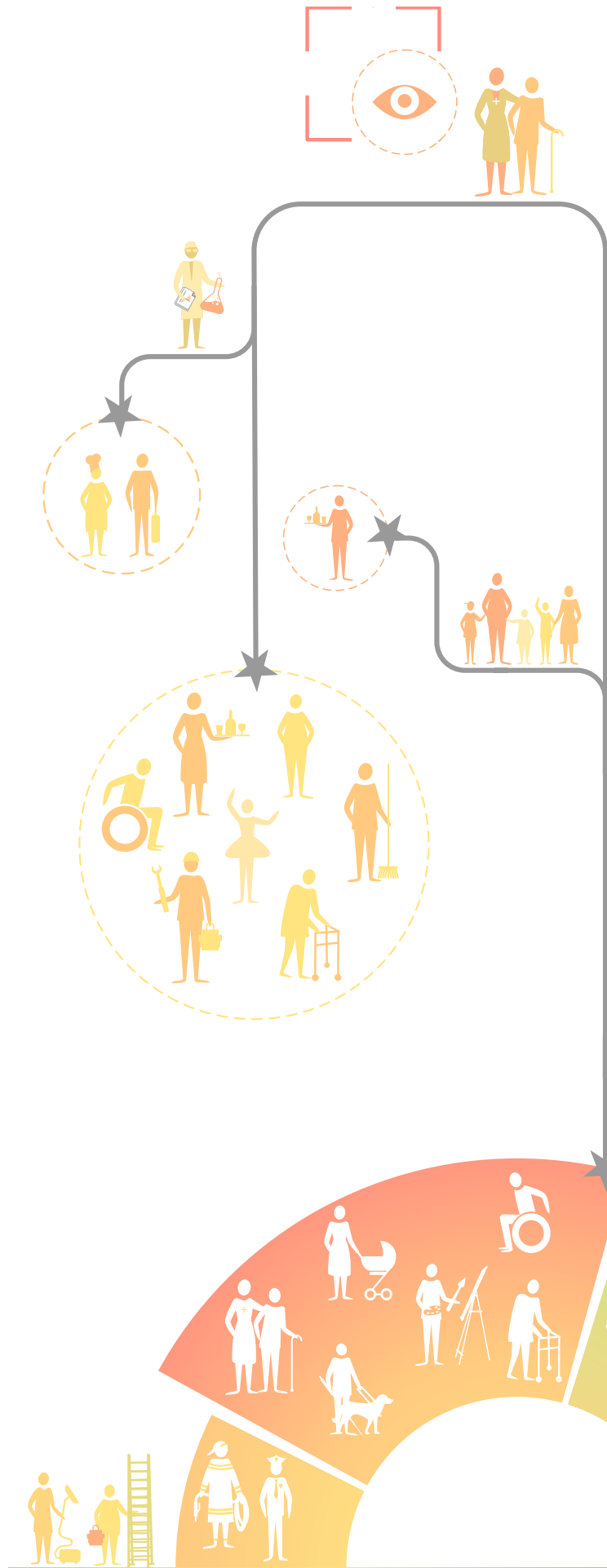
CONCLUSIONS & RECOMMENDATIONS:

It's important to remember that we are researching public perceptions of data uses in order to improve data uses, not to improve perceptions. Understanding public perceptions is necessary so we can then turn our attention to data uses and data-driven systems and think about how they should change. The recommendations below identify what data policymakers, practitioners and researchers can do to improve specific data uses and the data-driven ecosystem more broadly.

1. Address inequalities

There is widespread concern about the potential for data uses to have negative consequences for people from disadvantaged and minority groups and to reinforce inequalities, and people don't want that. This is more concerning in some contexts than in others, eg welfare. Data policymakers and practitioners need to acknowledge differences, in data uses and in public views about them, and avoid simplistic understandings of public attitudes.

Recommendations:

1.1. Avoid generalising about public opinions about data uses. There is no, one 'public,' and the idea of a singular public obscures the differences and inequalities that characterise 'diverse publics', which we suggest is a better term.

1.2. Understand the potentially discriminatory impacts of different data-driven systems, in order to a) communicate clearly about them and b) seek to overcome them.

1.3. When communicating about data uses, provide honest information about potential harms, especially for disadvantaged or minority groups, because this concerns people.

1.4. 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.

2. Ensure data uses have fair characteristics

Data uses in the public or social interest, that are supported by clear and honest information that is easy to understand, were seen to be fairer than data uses without these characteristics. Fair data uses enable choice, about whether to share data and about what happens to it, and they don't negatively impact some groups more than others.

Recommendations:

2.1. Consider not engaging in data uses that concern people, for example sharing data intended for pro-social or the public good with commercial companies who will make a profit from it

2.2. Build choice into data processes, for example about what data is collected and what happens to collected data.

3. Be aware of and respond to imaginings

Because people imagine, suppose and assume things about data uses, more information will not necessarily result in more understanding of them, but addressing what matters to people in explanations may help. 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 communicate

information that is of interest to diverse publics. Transparency efforts need to make it possible for people to understand what happens to data. They should speak to people's concerns about what might go wrong, what might change in the future and how disadvantaged communities might be affected. And more research into what motivates people to imagine, suppose and assume is needed.

Recommendations:

3.1. Communicate clearly about data uses in order to actually communicate, not just to tick the transparency box. Explanations need to focus on what matters to people, for example: addressing what might go wrong (responding to the imagined gap between data uses as described on paper and in practice); possible future as well as actual current uses of data (responding to imagined future changes); and whether and how some groups might be more negatively affected by data uses than others (responding to participants' imagining the experiences of more disadvantaged others).

3.2. Commit to supporting a change of culture around data uses, in which data is not misused, security is not breached, transparency is sustained and purposeful, because what has happened in the past can influence how people perceive data uses in the present.

3.3. Use visuals to communicate data uses and ensure explanations can be easily translated to other languages. These simple steps could significantly improve people's understanding of data uses.

3.4. Support or commission further research to improve understanding of imaginings, where they come from and how to address them.

4. Review commercial company involvement in public sector data systems

Commercial company involvement in public sector data systems concerned participants in our focus groups and interviews as well as our survey. It also confused them, sometimes because of a lack of clear information about the nature of their involvement.

Recommendation:

4.1. Support or commission further research into the specific aspects of commercial company involvement in public sector data systems that are concerning and confusing, how to communicate complex public-private partnerships, and whether visualisations, of the kind we used and participants appreciated, can help to communicate complexity and overcome confusion.