



# Public perceptions of sharing and aggregating health data

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

# 1. INTRODUCTION

[Living With Data](#) is a research project funded by The Nuffield Foundation, which 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. 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).

On Living With Data, we produced accounts and visualisations of specific 'data uses' which we presented to and discussed with research participants in surveys (n = 2000 x 2), and in focus groups and interviews (n = 112).

We selected cases from public sector organisations because their data systems increasingly shape everyday life experiences, and yet they had received less attention than high profile commercial systems at the time of our research. We identified welfare, public service media and health as three domains on which to focus our research because they are core aspects of everyday life.

This document summarises what we found about people's perceptions of two health data uses which involve data aggregation and sharing: the NHS (National Health Service) Covid-19 Data Store and on an antibiotic prescribing research project. An overarching project report and reports on other sectors can be found in the Resources section of the Living With Data website, along with other publications from the project.

---

## 2. CASE STUDIES OF HEALTH DATA USES

We produced accounts of the NHS Covid-19 Data Store, a national data store to help organisations responsible for coordinating the Covid-19 response, which we chose because our research was carried out during the Covid-19 pandemic. We drew on information in the public domain, on government web pages and elsewhere. We also produced an account of an NHS antibiotic prescribing research project. We drew on prior research by one of us, Itzelle Medina-Perea, to produce this account. Itzelle's research drew on practitioners' accounts of the antibiotic prescribing project and additional sources provided by interviewees or that were publicly available.

Alongside other general questions, we presented a textual description of one of these, the NHS Covid-19 Data Store, to 1/3 of our survey respondents; the other 2/3 saw other public sector data uses. As the survey was completed by 2000 respondents on the two occasions we administered it, this means around 1300 of them answered questions about the NHS Covid-19 Data Store.

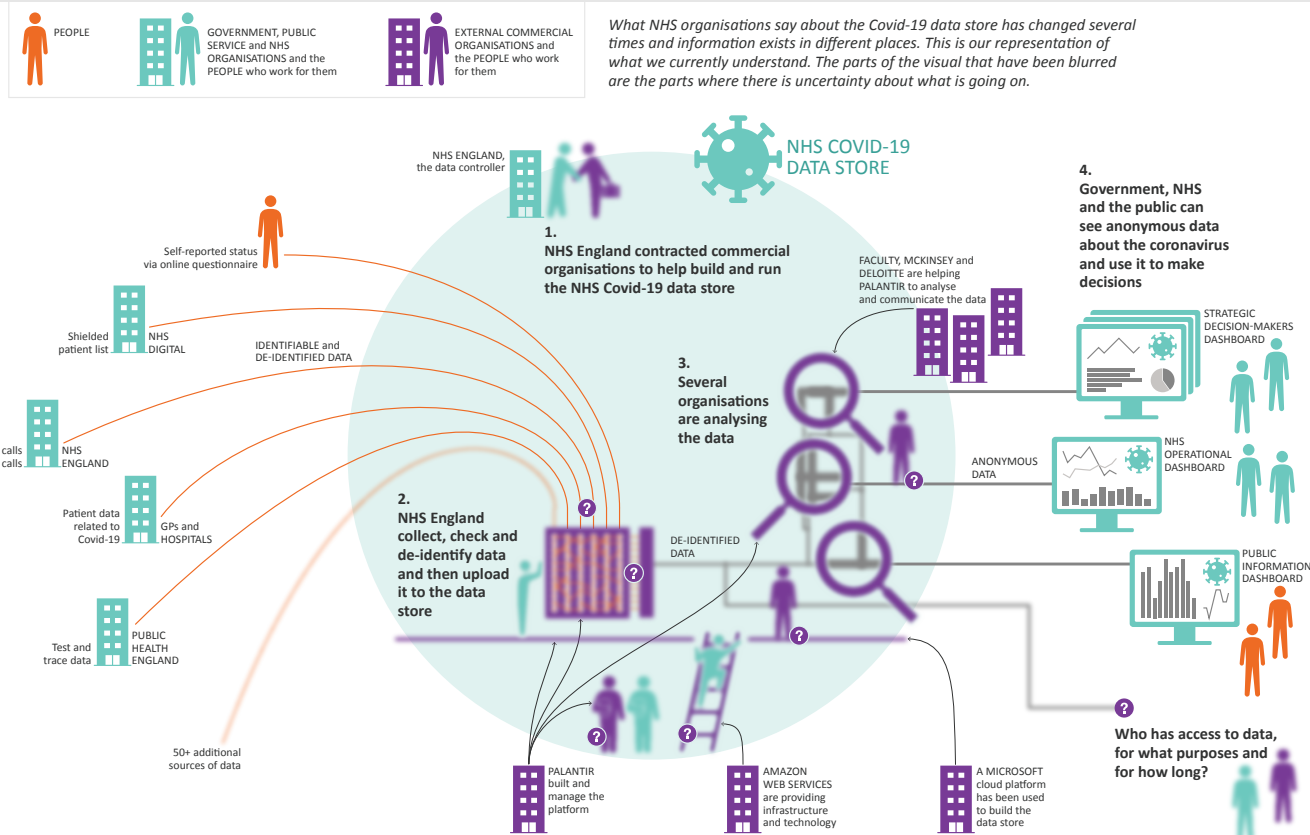
We ran the survey twice to explore change over time, and specifically to investigate whether the different phases of the Covid-19 pandemic affected attitudes to data uses. In response to the vast majority of questions that we asked, we found no statistically significant difference between figures from the first and second waves of the survey. Therefore, in this report, estimates of the proportions of people holding particular opinions refer to the 2020 survey. Where we did find differences between 2020 and 2021, we highlight these.

We presented visualisations and verbal descriptions of both NHS data uses to participants in our focus groups and interviews, which ran from November 2020 to September 2021. Figure 1 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 Data Uses page of our website. Descriptions can be found later in this report.

Figure 1: visualisations of health data uses that we discussed in our research

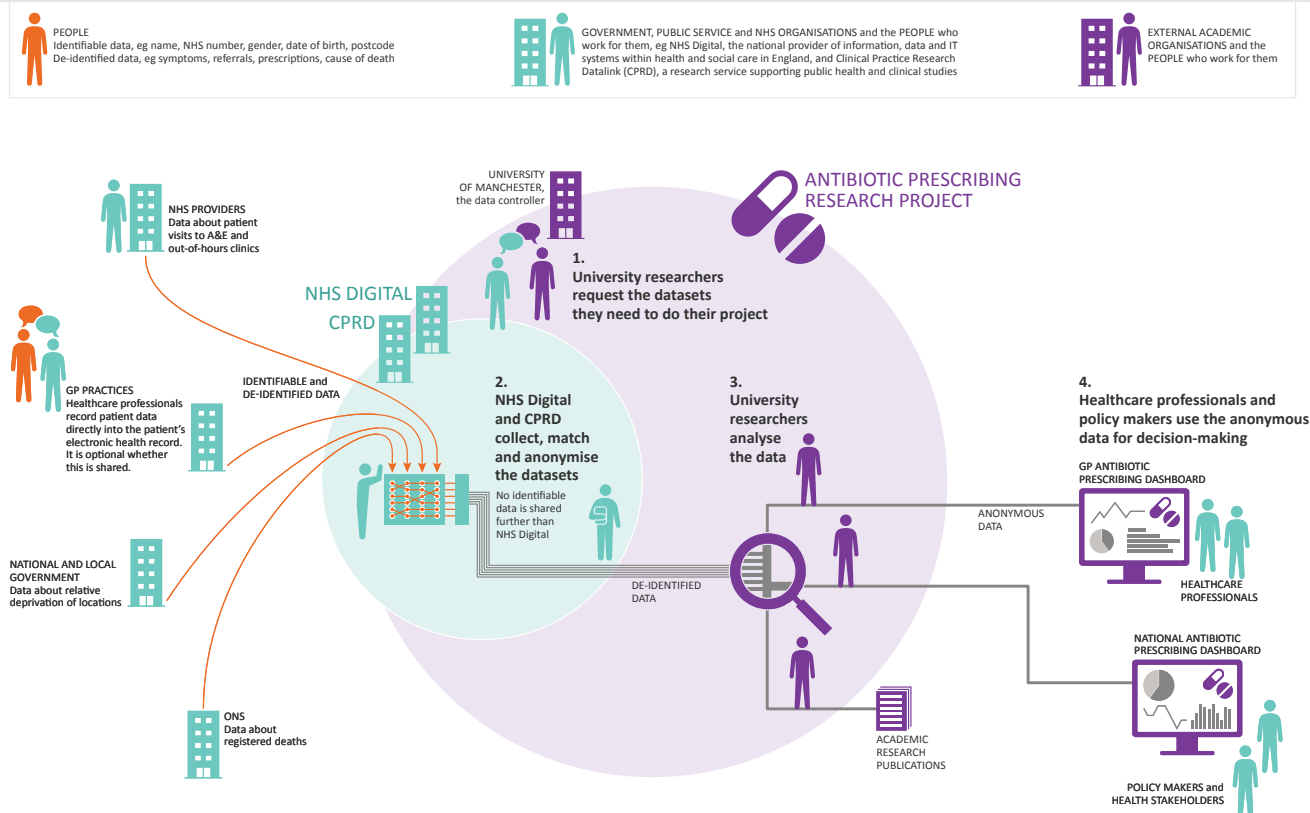
## NHS Covid-19 data store

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



## NHS antibiotic prescribing research project

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



For the interviews and focus groups, we grouped our example data uses into four themes: Data Matching; Data Ownership and Control; Data Sharing and Re-use; Algorithmic Processing. We discussed one theme in each focus group or interview, which means that each theme was discussed by approximately 1/4 of our participants. The themes that included health data uses were:

- **Data Matching:** this is where organisations match data from different datasets or databases, or compare data from one dataset with data from another. The Data Matching theme included the NHS antibiotic prescribing project and an example from welfare.
- **Data Ownership and Control:** this theme focused on who owns and controls data about us. It included two public service media data uses and the NHS Covid-19 Data Store.
- **Data Sharing and Re-use:** this theme focused on instances where data collected by one organisation is then shared with another organisation. It included the two NHS data uses, the NHS antibiotic prescribing project and the NHS Covid-19 Data Store, and a public service media data use.

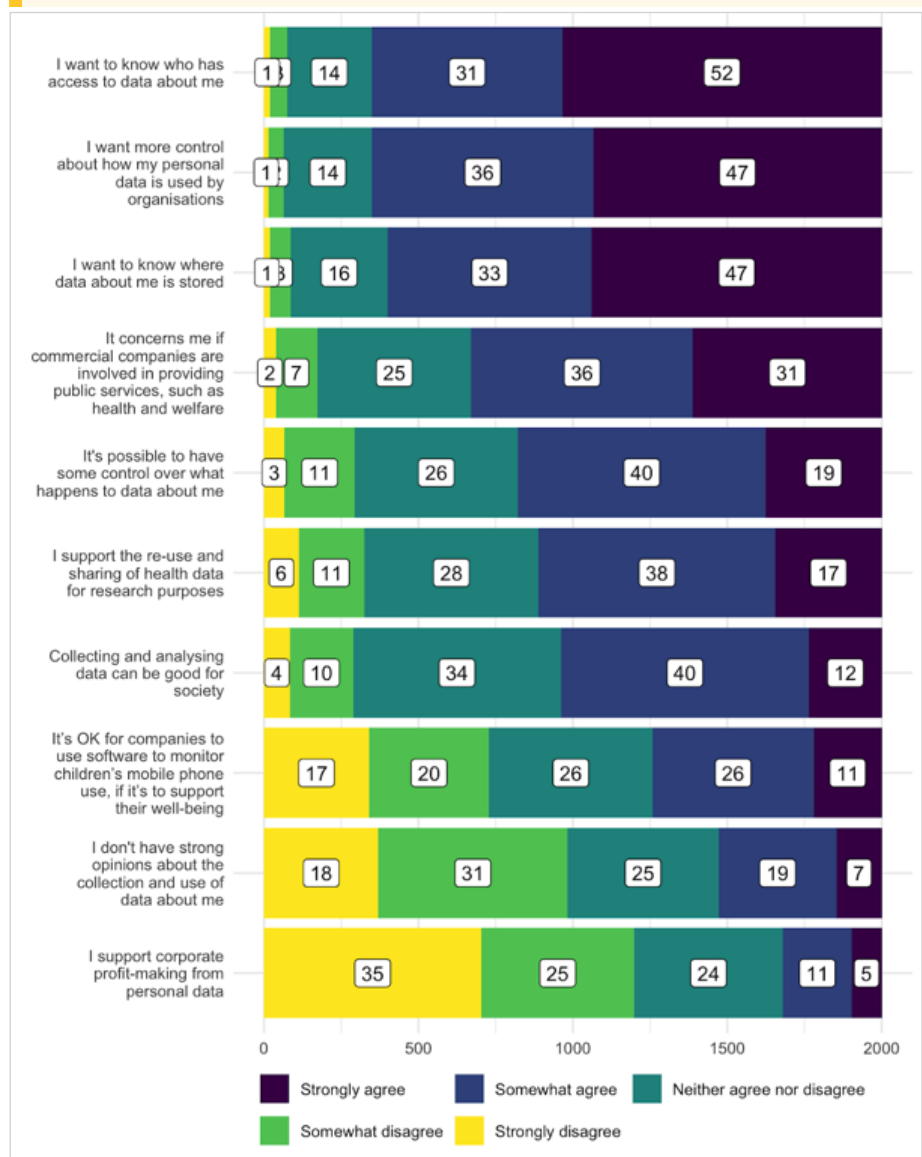
Visualisations of these themes can be found on the [Producing accounts of data uses](#) page of our website.

### 3. PUBLIC PERCEPTIONS OF CASE STUDY USES OF HEALTH DATA: WHAT WE FOUND

#### 3.1. Context matters: 'social good' rationale influences perceptions of and trust in health data uses

In a section of the survey which aimed to gauge respondents' attitudes to data uses, we presented them with ten statements and asked them to indicate how much they agreed or disagreed with each one. Statements and responses can be seen in Figure 3. Here we found 55% of respondents agreed or strongly agreed with the statement 'I support the re-use and sharing of health data for research purposes.' Five statements received more support (eg 83% agreed 'I want to know who has access to data about

Figure 3: Statements with which respondents were asked to state their agreement or disagreement.



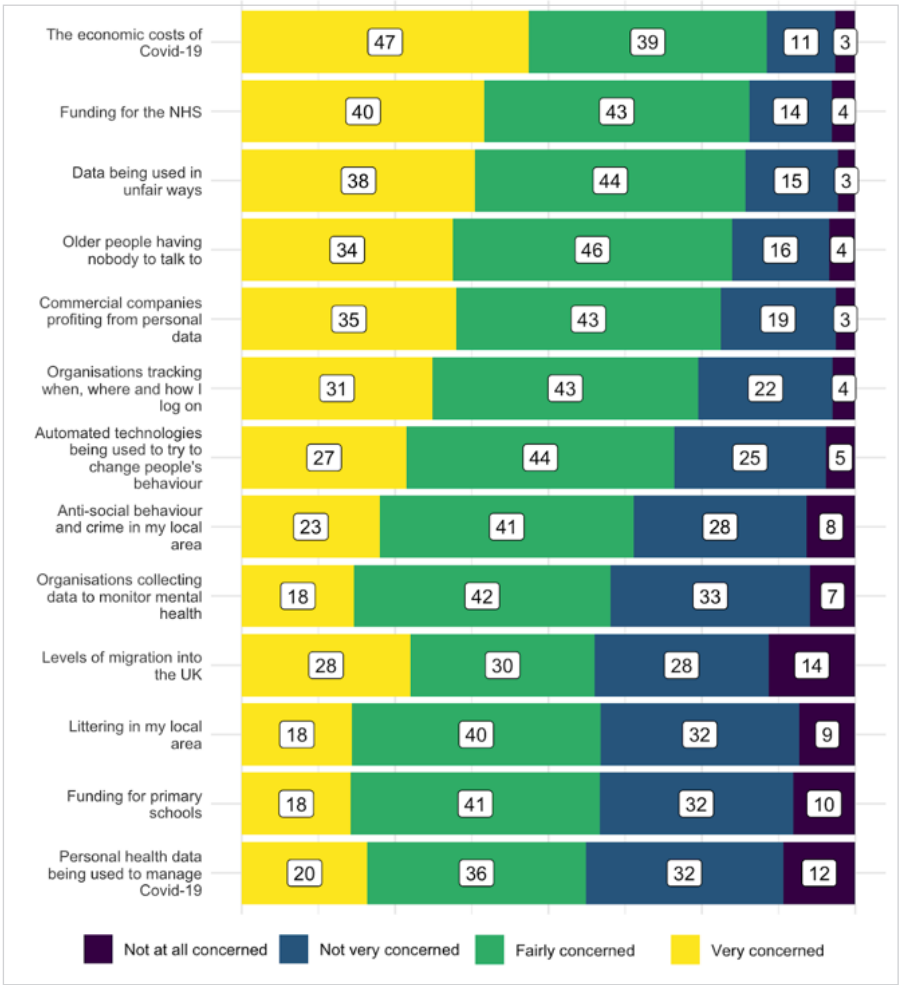
me') and four statements received less support (eg 26% agreed 'I don't have strong opinions about the collection and use of data about me').

Later in the survey, to understand how concerns about data uses compare with other concerns, we asked respondents how concerned they were about two lists of issues, one broad issue (eg the economy, pandemic diseases and immigration) and the second more narrow. Figure 4 shows responses to the second, narrower list. As can be seen, in the first wave of the survey, personal health data being used to manage Covid-19 was the least concerning issue, with 56% of respondents fairly concerned or very concerned about it, compared to 86% of respondents who were concerned about the economic costs of Covid-19 and 83% of respondents who were concerned about funding for the NHS.

In the second wave, personal health data being used to manage Covid-19 had become the second least concerning issue. This was because another issue from the list, 'littering in my local area', had become less of a concern, not because Covid-19 health data use was any more of a concern.

In the survey, we also asked respondents about whether they trusted different sectors and institutions, including the NHS, to: a) keep their data safe, b) gather and analyse data about them in responsible ways, and c) be open and transparent about what they do with data.

Figure 4: In general in your daily life, how concerned are you about each of the following?



We did this in order to gauge whether trust varies across institutions or across data uses. In other words, we explored whether trust in organisations and sectors in general influences attitudes to those same organisations' or sectors' data uses.

Overall responses can be seen in Figure 5. As can be seen, levels of trust were consistent across the three data uses that we asked about (keeping data safe, gathering and analysing data in responsible ways, and being open and transparent about what is done with data) across all sectors and institutions. This consistency in degrees of trust across the three data processes suggests that respondents' trust in sectors or institutions influences their trust in the same sectors' or institutions' data processes. In other words, sectoral or organisational context is an important factor when it comes to attitudes to data uses.

Consistent with the findings of other surveys (eg ODI 2018), levels of trust were highest for the healthcare sector or organisations. 'My GP' was ranked highest out of 11 options, and the NHS was ranked second highest. 72% of respondents said they trusted the NHS a great deal or a moderate amount to keep their data safe, 69% trusted the NHS to gather and analyse data in responsible ways and 67% trusted the NHS to be open and transparent about what they do with people's data.

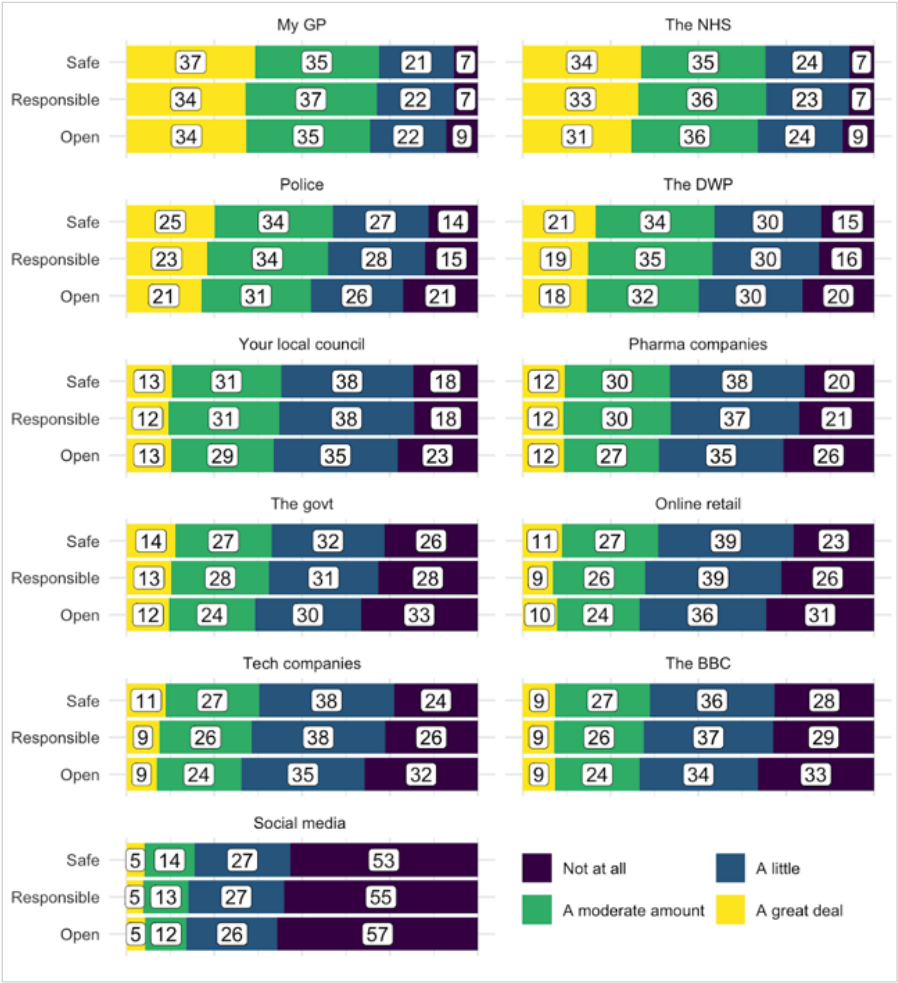
In the section of the survey about specific public sector data uses, respondents read a description of the ways in which data in the NHS Covid-19 Data Store is collected, shared, and used by different organisations.

Writing an accurate description of the NHS Covid-19 Data Store was challenging, because details were originally not fully available, they changed several times, and some contradictions existed in the available information. Experts on patient and health data on the Living With Data advisory board helped us to ensure the description was as accurate as possible in this context. The final version included in the 2020 survey is reproduced in Figure 6.

Respondents were asked how comfortable they felt about their NHS patient data being added to the Data Store. Around three out of four respondents were fairly or very comfortable about their NHS patient data being added to the NHS Covid-19 Data Store (78%). Comments in free text fields suggest that this is because of support for its purpose. In the focus groups and interviews, where we discussed the NHS antibiotic prescribing research project (a description of which can be found in Figure 7) as well as the NHS Covid-19 Data Store, we also found that many participants were comfortable with the health data uses that we discussed with them because they saw them as for the public or social good.

For example Tanya, a white, British lesbian woman, aged 55-64, with a long-term condition and an annual household income of £30,000-£39,000, described fair

Figure 5: 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?



data uses as ‘for the greater good [...] for health and benefits, for people, for us to learn, develop and grow in a way that’s going to make society a fairer place and people have the best opportunities in their life’.

Tanya felt that the NHS antibiotic prescribing research project was fair because of her ‘sense of where it’s motivated from’. Likewise, Brook, a white, heterosexual man, aged 35-44, who is non-verbal and whose views were spoken by his facilitator, said ‘[NHS Antibiotic Research] it’s clear, I understand why they’re doing what they’re doing’. And Astrid, non-binary, lesbian/gay, unemployed, aged 35-44, from an ‘other white’ background, said:

“ we’re talking about a research project developed in response to the public health crisis around the overuse of antibiotics, [...] the NHS is a very, very valued and cherished institution in this country, and that is probably beyond individual perceptions across the board, more of an idea of, you know, the NHS trying to do the right thing by the citizens. *Astrid*



Figure 6: description of the NHS Covid-19 Data Store data use from the survey

## NHS Covid-19 Data Store

**NHS (National Health Service) organisations, including NHSX (responsible for digital changes to national healthcare), have been commissioned by the government to develop a national, secure data store to hold data in one place to help national organisations responsible for coordinating the Covid-19 response.**

Patient data related to Covid-19, from GPs and hospitals, as well as call data from 999 and the 111 coronavirus helpline, lab test data from Public Health England, and data from the Office for National Statistics (ONS) is checked by NHS England and then uploaded to the secure NHS data store. The data in the NHS data store does not identify any individual in its current state. However, it may be possible for somebody with the right skills to re-identify some of this data.

Commercial organisations, including Microsoft, Google, Amazon Web Services, Palantir Technologies and Faculty, are collaborating in the development of the NHS COVID-19 data store. They are providing cloud platform and infrastructure technologies for the functioning of the data store, data

collection tools and data processing software, data modelling and support with data analytics capacity and capability. They have access to patient data to enable them to fulfil their roles.

NHSX says that when the pandemic subsides and the outbreak is contained, they will close the Covid-19 data store. They say that the data processing will stop and all data will be either destroyed or returned to the NHS once the public health emergency situation has ended.

What NHS organisations say about the Covid-19 data store has changed several times and information exists in different places. Patient data groups are concerned that not enough detail has been provided about contracts with partners to fully understand who has access to data, for what purposes and for how long. They are concerned that commercial companies may be able to use data from the data store to develop their technologies and potentially increase their profits.

Figure 7: description of an NHS antibiotic prescribing research project used in focus groups and interviews

## NHS antibiotic prescribing research project

**The NHS research project about antibiotic prescribing. Prescribing too many antibiotics is a problem, because it means that people are becoming resistant to them. Because of this, antibiotic resistance is a public health crisis. The research aims to understand antibiotic prescribing in order to reduce it and so address the antibiotic resistance crisis.**

The NHS Antibiotic Prescribing Research Project gathers data from a range of sources (eg visits to GPs and hospitals, data about deaths, or about the deprivation of local areas).

University researchers request access to the datasets they need to do their project.

This data is linked together by NHS Digital and CPRD (Clinical Practice Research Datalink, a service supporting public health and clinical studies). In this process, the only organisation that receives identifiable patient data is NHS Digital, the statutory body in England with legal authorisation to receive this type of data.

The linked data is shared with University of Manchester researchers who analyse the data to understand the factors

that affect antibiotic prescribing and to try to improve how this is done. Individuals can't be identified by the people with whom the data is shared.

The researchers produce dashboards for GPs and health policy makers to improve / reduce antibiotic prescription.

### From the Data Matching theme:

Potential benefits of the NHS antibiotic prescribing research project: address public health crisis of antibiotic resistance

Concerns about the NHS antibiotic prescribing research project: it requires the use of patient data.

### From the Data Sharing and Re-Use theme:

Potential benefits of the NHS antibiotic prescribing research project: address public health crisis of antibiotic resistance

Concerns about the NHS antibiotic prescribing research project: 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?)

### 3.2. 'You can't get the genie back in the bottle': Concerns about data sharing

Alongside the support for health data uses that we found, we also found some concern about the involvement of commercial companies in the provision of related services or infrastructures. In the section of the survey which presented statements in order to gauge attitudes (shown in Figure 3), we found 67% of respondents agreed or strongly agreed with the statement 'It concerns me if commercial companies are involved in providing public services such as health and welfare'. Disabled people were more concerned about commercial companies providing public services than people who are not disabled (70% compared to 65% respectively). From the same list of statements, only 16% of respondents agreed or strongly agreed with the statement 'I support corporate profit making from personal data' and 60% disagreed.

This finding was confirmed in the section of the survey about concerns, where, in the second list of narrow issues, commercial companies profiting from personal data was ranked the 5th biggest concern of 13. The economic costs of Covid-19, funding for the NHS, data being used in unfair ways and older people having no-one to talk to ranking above this issue. We also found differences between groups (statistically significant at the 95% level). People in receipt of Universal Credit (UC) were more concerned about commercial companies profiting from personal data than people not in receipt of UC (27% compared to 18% of). Black (42%) and Asian (38%) people were more concerned than White people (31%), and LGBTQ+ people (39%) were more concerned than heterosexual cisgender people (32%). Black, Asian and other racialised people and LGBTQ+ people were more concerned across the board than other groups. Overall, 72% of White British people were fairly or very concerned about the issues listed, compared with 79% of Black people and 80% of Asian people. The equivalent figures are 73% for heterosexual cisgendered people and 76% for LGBTQ+ people.

Furthermore, in the survey, despite high levels of comfort with the NHS Covid-19 Data Store, comments in free text fields about the Covid-19 Data Store revealed concern about data sharing. More than half of these expressions of concern were about the involvement of commercial companies.

Concerns were often expressed as imagined future scenarios, most of which were negative, and many of which involved commercial organisations profiting from, leaking, misusing or selling data in the future. Examples included:

- 'They might use this data to force people to take the vaccine depending on their health condition'
- 'I am slightly concerned that my personal data will fall into 3rd party companies against my will'
- 'It's just maybe they will share it to somewhere its not safe'
- 'Just worried their might be a mix up that might come up in future to bite one'
- 'There will be breaches because systems are poor and staff overworked and will make mistakes'
- 'I think they will use that data for other purposes that we won't agree with when we sign up'

In focus groups and interviews, some participants also expressed concern about who is involved in the provision of data-driven services. A number of participants expressed concerns about future uses of data in the NHS Covid-19 Data Store, stating that data could be used for reasons not foreseen, intended or described at the time that data uses are implemented or in planning.

One was Craig, a white British, heterosexual man, 35-44, with an annual household income of £30,000-39,000 and no long-term conditions. Craig was concerned that as a result of future data sharing, data could end up 'in the wrong hands' from the perspective of what might be in the interests of social good. The data could then inform decisions that impact negatively on broad or narrow groups of people. Craig used insurance as an example:



If some data goes to get used to discriminate for insurance against one particular health condition, or pre-existing health condition – and I think that's very feasible even in England. If you're trying to get insurance and you fill in your questionnaire and [...] they've done a mass analysis on this data to find out, well, if you've had COVID and you're of whatever demographic, we're not going to give you insurance. I don't have any confidence in the government not to go and allow unscrupulous companies access to that data. *Craig*



Lewis, a white, heterosexual man, aged 55-64, with a household income of £50,000-£59,000 and no long-term conditions, was also concerned about potential future sharing of data from the NHS Covid-19 Data Store, stating that once shared, ‘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’.

Governments and political conditions can change, a number of participants noted, and this might lead to changes in who has access to data, how data is interpreted, and how it is used. This was a concern of Teddy, a white British, heterosexual man, aged 65+, retired with a household income of £50,000-£69,000 and a health condition. When talking about who has access to data in the NHS Covid-19 Data Store, Teddy noted that the UK government had invested a lot of money in managing the pandemic, and it may seek to recoup its spending by changing the law so that the data in the store became more widely available. Todd, a white, British gay man born in the UK, aged 25-34, with an annual household income of less than £10,000 and no long-term conditions, and his partner Matthew, a white, British gay man born in Zambia, aged 35-44, a web developer with an annual household income of £20,000-29,000 and no long-term conditions, also noted that political change could mean that personal data – about sexual orientation or gender identity, for example – could be used against people in dystopian or dictatorial futures. Diane, a white, British, heterosexual woman, aged 55-64, with an annual household income of £40,000-£49,000, said this about the NHS Covid-19 Data Store:



It’s almost too big, isn’t it? If they’d said, right, we’re going to collect this information for that reason, then that’s fairly simple. But what’s happened, it’s like a spider’s web. And the trouble is, then you lose, you know, on these little spider’s legs, you lose a) what it’s all about and b) it opens up to, first of all it going wrong and not working properly, and people then getting the information for means that they shouldn’t have. You know when you look at it in more detail [...] it’s, you know, it does make you sort of change your mind. *Diane*

These quotes speak to broader concerns about commercial company involvement in the NHS Covid-19 Data Store.

### 3.3. Trust and concern vary across groups

Belonging to a disadvantaged or minority group appeared to inform how participants perceived data uses. To be clear, 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. In turn, those experiences shape perceptions of data uses.

In the survey, we found that older people were more trusting of their GP. 74% trusted their GP to be open and transparent about what they do with data, 81% to gather and analyse data in responsible ways, and 80% to keep patient data safe, compared with 65%, 62%, and 62% respectively for the youngest group. LGBTQ+ people were less likely to trust their GP (30%) than heterosexual cisgender respondents (35%). Disabled people were more likely to trust their GP and the NHS data uses in general (38% compared to 33%, and 36% compared to 30% respectively), than respondents who did not report a disability.

There were also differences in comfort with the NHS Covid-19 Data Store by ethnicity, gender, sexuality, and qualifications. Black, Asian and other racialised people were more likely to feel comfortable with their NHS patient data being added to it than white respondents (73% were very or fairly comfortable, compared with 66% of White people). Men were more comfortable with it than women (73% compared with 64% for women). LGBTQ+ respondents expressed more discomfort with the NHS Covid-19 Data Store than heterosexual cisgender respondents.

Most of these survey findings were confirmed in our focus group and interview research. For example, some of our older participants appeared to be more comfortable with NHS data uses than data uses in other sectors. Alan, a white, British heterosexual man, aged 65+, retired, and his partner Betty, a white, British heterosexual woman, aged 65+, also retired, both of whom have no long-term conditions, were positive about data uses in the NHS Antibiotic prescribing research project. For example, Alan said:



If we’d been asked to cooperate in our hospital environment with a thing like this, we would have jumped at it, wouldn’t we? *Alan*

Betty agreed, and other older participants also concurred. For example, Bruce is a white, British heterosexual man born in the UK, aged 65+, retired with an annual household income of £20,000-£29,000 and has long-term conditions. He said of the NHS Covid-19 Data Store ‘I would have thought that the more medical information they have, the better they can target treatment’.

Some participants with long-term conditions also felt the same. In one focus group on the Data Matching theme, participants agreed that they trusted the NHS and did not mind its uses of their data. Celeste, a black British-Caribbean, heterosexual woman, aged 35-44, unemployed, with an annual household income of £9,999 or less, explained:

“

If something does happen, they should be able to type your name into the system and be like, ‘Oh, this is such and such. She has recently come out of hospital for such and such’, when, what she was prescribed, it should all be updated and in a place, I don’t know where, but stored somewhere. *Celeste*

In contrast, NHS and health data uses seemed to concern LGBTQ+ participants more than heterosexual cisgendered participants. A discussion of the NHS antibiotic prescribing research project took place between Todd and Matthew, mentioned above, reflected on the importance of the kind of health data in question.

They said that they would feel differently if the research project used data about sexual health, compared to data about viruses like colds or the flu, because of the ways in which this kind of health data can be mobilised for harmful ends in a context of structural inequalities. Drawing on his awareness of how data uses might feed into existing discriminatory structures, Todd said:

“

The sexual angle is a really, really good point, particularly when it comes to LGBTQ and that community, because we know that some of the information about our sexual health in the past has been used to massively discriminate, to massively sort of chastise those communities. It’s interesting from that perspective that it could potentially take one person with a vendetta of some sort – or just a mistake, just a bit of human error, that could then cause ramifications for that particular group or population that maybe more – that therefore the risk might be heightened. *Todd*

Ellis is non-binary, bisexual, white British, aged 25-34, with an annual household income of £30,000-£39,000 and no long-term conditions. They were also cautious about health data uses. Towards the beginning of their interview, Ellis was asked whether they had ever changed their behaviour because of their knowledge or experience of data collection practices. Ellis said they had altered some of their social media activity to reduce producing data about their identity as a trans person, and they were selective about who they shared their health data with:

“

There’s some things, especially NHS things, where I don’t disclose honestly, like, medication, because so they’ll ask you like are you taking any regular medication [...] but if I perceive that the fact I’m taking hormones isn’t relevant to them I won’t let them know. So I’ll just lie about it, even though it’s probably confidential and you expect the staff to have training, I’ll just be like, right, it’s easier for me if I don’t mention that and just try and not flag up being trans at all.

Like at the opticians, or at the dentist, or like things like that. I get a bit paranoid that people will treat you differently and not necessarily in a bad way, just that they will realise and it will just feel, there will be a shift and they’ll treat you different. *Ellis*

### 3.4. The challenge of understanding health data uses and the role that imagining plays

Participants from across a number of different demographic groups drew attention to the fact that understanding data uses is challenging. For example, Richard and Lucinda, who both have long-term conditions, were asked whether they thought understanding what happened to their data was easy. Lucinda is a white British, bisexual woman, aged 35-44, who was born in the UK. She has an annual household income of £10,000-19,000 and is not in paid work. In answer to the question, Lucinda said 'oh god no, I don't know what happens'. Richard, a white British heterosexual male who was born in the UK, aged 45-54, with an annual household income of £10,000-£19,000, is dyslexic. He doesn't think he has ever 'read through a single terms and conditions' because doing so feels like 'an ordeal' to him. He said:



All reading is just, not all reading, but just things where it's not designed to be clear, like in any way confusing [laughs], or legal jargon, or any kind of potentially ambiguous – I just [sighs], it does me in *Richard*

One participant, Brook, mentioned above, was generally positive about data uses, but was concerned about 'those who require the most assistance' being 'adversely affected'. Brook's facilitator highlighted the important role that his support staff play in Brook's life, enabling him to live independently. With regard to data uses:



Brook can opt out because we're here to have that conversation and to assist him in that. [...] [Other people might not have such support] could end up being exploited in that. Because the onus is on them, so the person collecting the data will always say, 'But you didn't tell me no'.  
*Brook*

We found that imagining played an important role in participants' efforts to understand 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 in explanations. 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 sometimes imagined that there would be a disparity between how data processes are said to work in theory, and how they actually work in practice. They added to the accounts of data uses that we presented to them, imagining additional elements that helped them build a mental picture of a data process. For example, in a focus group discussion about the Data Matching theme Matthew, mentioned above, identified a couple of possible disconnects between how the NHS Antibiotic research project works in theory and how it might work in practice, both of which were based on personal experience. In the first case, he questioned the academic rigour which was said to underpin the NHS Antibiotic research project, based on personal experience of a disconnect between how data uses are said to work 'on paper' and how they work 'in reality'. This was a real-life experience relating to sexual health data, in which the fallibility of the human actors involved was the source of the disconnect. He had visited a sexual health clinic and had been asked for the details of previous sexual partners. He said:



I gave them details and then they said we won't be able to tell you whether or not they've contacted us, obviously, and I was like well yes, obviously. Then a week later, got a phone call, hey so and so hasn't called us. Did you give him your details? I was shocked because they were giving me personal details relating to somebody and, you know, they don't know whether this person, I mean, it could have been an abusive relationship, where that information could have had consequences for my safety, or their safety, whatever *Matthew*

He went on to link this experience to the NHS Antibiotic research project:



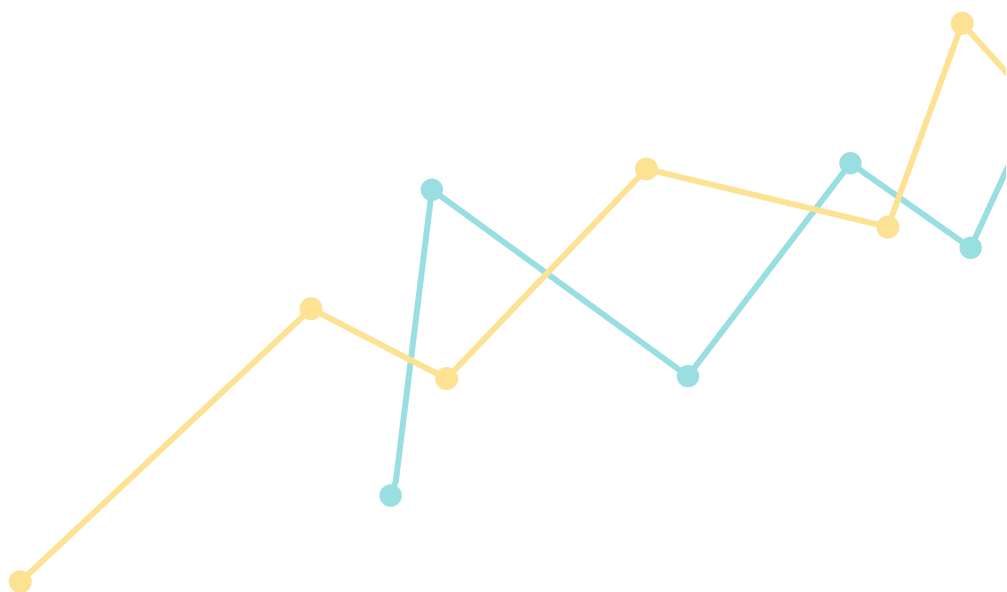
So, there is that. I mean, on paper this is great. On paper this is helping antibiotic research, you know, whatever. In reality, somebody who just has a target to meet for contact tracing doesn't care or doesn't know, considers me like, oh well he seemed nice. We can tell him, he won't mind. You know, I was nice. I'm not going to go spreading that information but they don't know that. *Matthew*

This real-life experience of a data confidentiality breach led Matthew to imagine that such incidents could happen again, and, in the context of sexual health, they could have negative consequences. He believed that LGBTQ+ sexual health data had already been used to discriminate in the past. 'Obviously STDs are seen very differently from the flu or a cold, that data can be used maliciously as well, and it's the safety around that data,' he said.

This experience led Matthew to question the reliability of other human actors in health data systems. Later, the fact that a friend works in a related department made him reflect on whether all stakeholders would be equally committed to ensure that claims about data privacy and confidentiality would be upheld. He said:



I can imagine doctors and nurses are very passionately concerned about data privacy, whereas a receptionist might just find it funny, you know, or might not have that same kind of commitment. Then the other thing is it's like somebody you know. I mean, one of my friends works in phlebotomy, you know, testing samples. There is a possibility that [...] something's labelled with my date of birth and she knows my date of birth. She would be able to go like, that's Matthew's date of birth, and make that connection. So it also relies on individuals and we all know individuals are not necessarily trustworthy or, you know, they're fallible. *Matthew*



## 4. CONCLUSIONS

Although there is widespread support for pro-social uses of health data, some groups support these more than others. NHS and health data uses concerned LGBTQ+ participants more than heterosexual cisgendered participants. Furthermore, there are concerns about the third parties with whom data might be shared. In the case of the NHS Covid-19 data store, respondents support its pro-social, public health aims and at the same time, they are concerned about the involvement of commercial companies and the lack of clarity about the ways in which these companies may have access to or profit from personal data in the future.

This finding matters for data-driven health services. As commercial companies are often involved in providing the technical infrastructure or processing capacity for public sector data-driven systems, concern about their involvement in or profit from public sector data systems should be taken into account by health services setting up such systems. High trust in GP and NHS data uses is in danger of being diminished by the involvement of less trusted parties, such as technology companies, in health data initiatives like the NHS Covid-19 Data Store (see Figure 5).

Most imaginings in survey free text fields were about possible future data misuses by commercial companies. We suggest that this is because of the involvement of a range of big tech companies in health data uses, such as the NHS Covid-19 Data Store, and the lack of clarity about their access to personal health data in the store.

### To address these issues, health service providers like the NHS could:

- Think carefully about third party or commercial company involvement in the provision of data-driven services.
- Provide clear information about third party or commercial company involvement in the provision of data-driven services.
- 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.
- Support or commission further research into the specific aspects of commercial company involvement in public sector data systems that are concerning, how to communicate complex public-private partnerships, and whether visualisations, of the kind we used and participants appreciated, can help to communicate complexity.