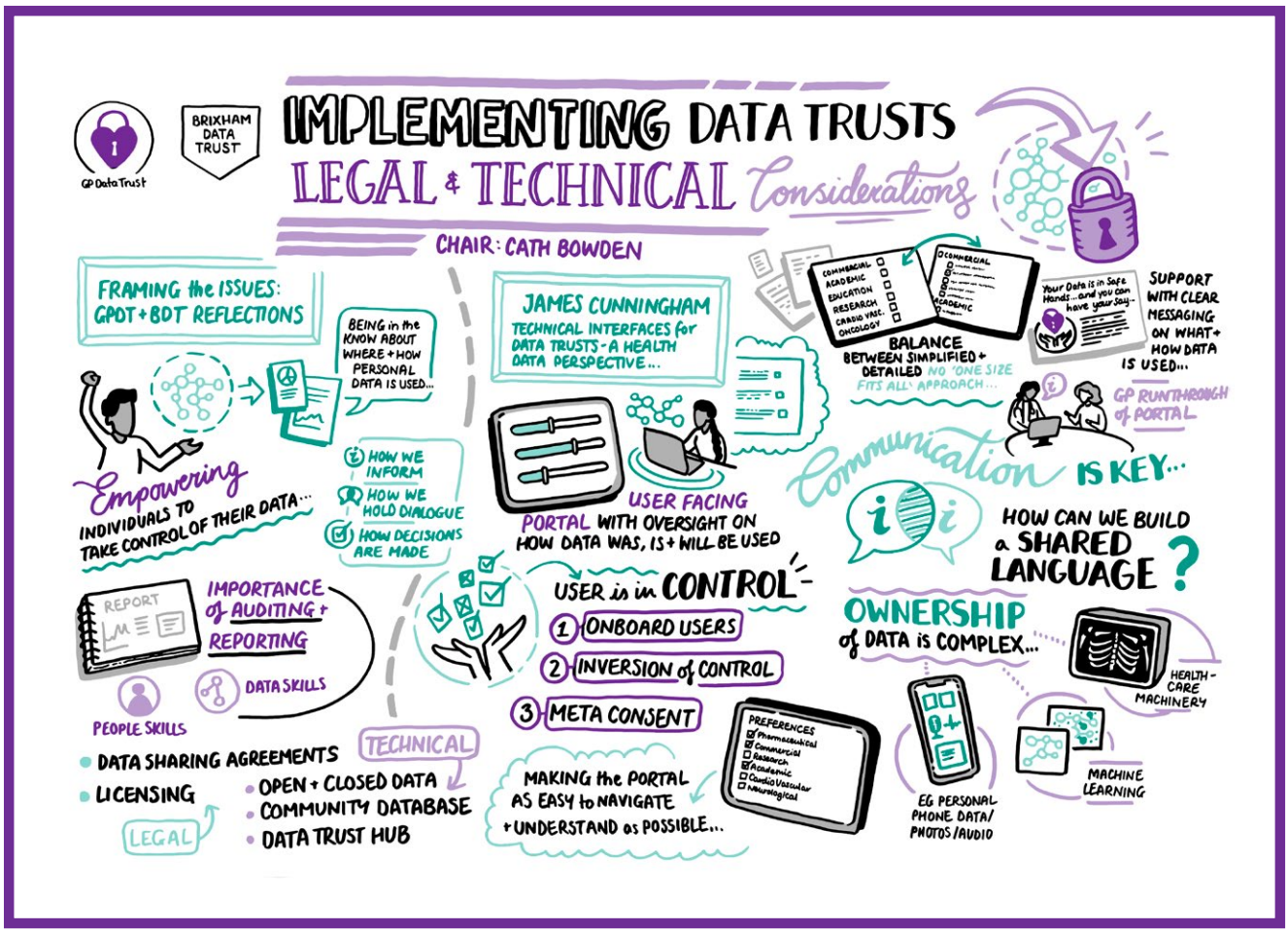
MANCHESTER 1824

The University of Manchester

Implementing Health Data Trusts

General Practice Data Trust Project Report



Sarah Devaney, Cath Bowden and James Cunningham

March 2025

Acknowledgements

We express our grateful thanks to all those who participated in our stakeholder workshop and focus groups; to Use My Data and Understanding Patient Data for facilitating engagement and participation; to Amber Anderson for the visual minutes throughout this report.

We are enormously grateful to our funders, the Data Empowerment Fund, for funding this work.

This report provides an account of work undertaken in 2024 to explore the technical and governance challenges to establishing a Data Trust for Health Research.

The aim of our work was to explore how public trust can be prioritised in systems to share health data for research and planning. This will enable wider confidence that, where people choose to share their data, it will be used for beneficial health purposes in line with their values and preferences, as well as being kept safe and secure.

Cover image: Figure 1: Legal and Technical Consideration of Establishing Data Trusts (Stakeholder meeting in May 2024)

Chapter 1: Background

In this section we set out:

- The promise of and challenges in the use of health data for research
- The nature of Health Data Trusts;
- Emerging governance approaches to health data sharing for research; and
- The importance of engendering public trust, transparency and involvement in health data sharing.

The Promise and Challenge of Using Health Data for Research

The immense promise of NHS data as a research resource is a key driver for UK Government efforts to make it available to health researchers.[1] As was noted in the Goldacre and Morley review[2]:

> *'NHS data represents an exceptional and globally important resource. For 73 years the NHS has collected detailed records and data, on tens of millions of patients, from a huge and ethnically diverse population. Because of this diversity, analytic outputs created from NHS data can help save lives around the world.'*[3]

> *'73 years of complete NHS patient records contain all the noise from millions of lifetimes. Perfect, subtle signals can be coaxed from this data, and those signals go far beyond mere academic curiosity. They represent deeply buried treasure, that can help prevent suffering and death, around the planet, on a biblical scale. It is our collective duty to make this work.'*[4]

Similarly, Lord Darzi's 2024 report on the NHS in England notes that health data offers untapped opportunities to enhance care and shift services towards the community.[5] Most recently, the final report of a review of the UK health data landscape carried out by Cathie Sudlow (the Sudlow Report) was published.[6] It describes health data as 'critical national infrastructure that can underpin the health of the nation'.[7]

While the promise of the use of health data for research is widely lauded, much remains to be done to ensure that:

- The data set is accurate and representative,
- There is a suitable technical and governance structure to facilitate its availability to researchers, and
- The public have confidence and trust in the system.

Unfortunately, government attempts to maintain public confidence and trust in its sharing of patient data have been undermined by its own attempts to implement strategies to gather such data. For example,

» In 2012 - 2013 the **care.data** scheme was proposed as a system with the power to collect and collate all medical information from NHS patients across primary and secondary care. However, practitioners and patients expressed deep concerns that there were inadequate safeguards or clarity around the potential uses of the data, as well as about the interaction of the scheme with existing privacy provisions. It was closed in 2016.[8]

» In May 2021, NHS Digital announced that they would collect patients' primary care or 'General Practice' (GP) data so that it could be used in a non-identifying form in

1 Department of Health and Social Care, Data saves lives: reshaping health and social care with data (updated version June 2022) https://www.gov.uk/government/publications/data-saves-lives-reshaping-health-and-social-care-with-data/data-saves-lives-reshaping-health-and-social-care-with-data

2 Goldacre, B & Morley, J. (2022) Better, Broader, Safer: Using health data for research and analysis. A review commissioned by the Secretary of State for Health and Social Care. Department of Health and Social Care, https://www.gov.uk/government/publications/better-broader-safer-using-health-data-for-research-and-analysis

3 Ibid, Executive Summary

4 Ibid, 17

5 Darzi A. Independent investigation of the NHS in England. Department of Health and Social Care, 2024. https://www.gov.uk/government/publications/independent-investigation-of-the-nhs-in-england

6 Sudlow C, Uniting the UK's Health Data: A Huge Opportunity for Society (November 2024). https://www.hdruk.ac.uk/helping-with-health-data/the-sudlow-review/. The report was commissioned by the Chief Medical Officer for England, the UK National Statistician and NHS England's National Director for Transformation.

7 Ibid, 11

8 See Sigrid Sterckx et al, 'You Hoped we Would Sleep Walk Into Accepting the Collection of our Data: Controversies Surrounding the UK care.data Scheme and their Wider Relevance for Biomedical Research' (2016) Med Health Care Philos, 19(2), 177-190.

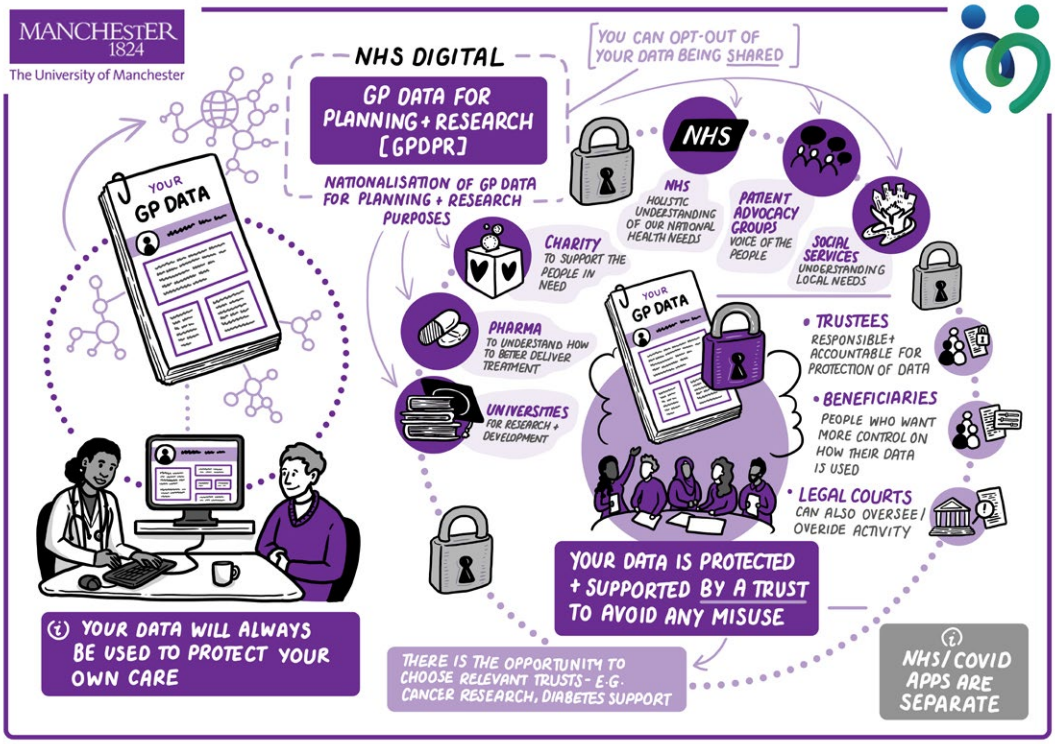


Figure 2: Visual Minutes - GPDPR and the Potential of Data Trusts

medical planning and research in their GP Data for Planning and Research (GPDPR) programme. Patients who did not want their data to be included in the programme had until 23rd June 2021 (subsequently postponed to 1 September 2021) to register an opt-out by printing, completing, and returning a form to their GP surgery. By June 2021, over a million people had registered an opt-out in this way, prompting NHS Digital to delay the programme to provide more time to speak with people about their concerns.

» In November 2023, the award of a 5-year, £330 million contract to run the UK's Federated Data Platform to Palantir (the US analytics company that charged the government £1 to process English health data during the pandemic and subsequently was awarded further health data processing contracts) caused additional disquiet about the government's disregard for patient and public preferences around health data sharing.

» In November 2023 there were reports that UK Biobank shared patient data with insurance companies despite previously indicating that they would not do so.[9]

It has been argued that the failure of care.data can be said to have been because its proposals fell foul of the 'social licence for research'. The social licence for research has been described as 'how the expectations of society regarding some activities may go beyond compliance with the requirements of formal regulation; those who do not fulfil the conditions for the social licence (even if formally compliant) may experience ongoing challenge and contestation.'[10] This argument can also be applied to the failure of GPDPR.

Since the publication of the *Data Saves Lives* strategy,[11] regulation and governance frameworks relating to patient data have gone, and continue to go through, significant amounts of change – NHS England, Health Education England, and NHS Digital have merged into a new NHS England; Integrated Care Systems have been established to develop shared plans across

9 Das, S. 'Private UK health data donated for medical research shared with insurance companies' The Guardian, 12 November 2023. https://www.theguardian.com/technology/2023/nov/12/private-uk-health-data-donated-medical-research-shared-insurance-companies

10 Pam Carter, Graeme Laurie, and Mary Dixon-Woods, 'The Social Licence for Research: Why Care.Data Ran Into Trouble' (2015) 41(5) Journal of Medical Ethics 404

11 Department for Health and Social Care, Data saves lives: reshaping health and social care with data, (June 2022) https://www.gov.uk/government/publications/data-saves-lives-reshaping-health-and-social-care-with-data/data-saves-lives-reshaping-health-and-social-care-with-data

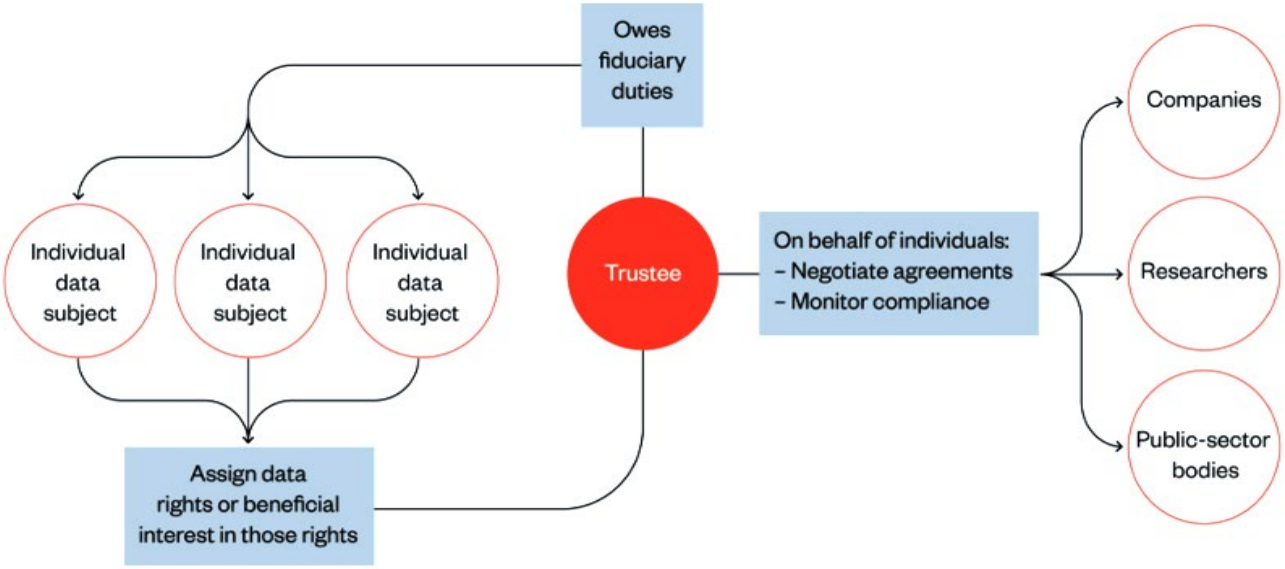


Figure 3: 'How Data Trusts Work' in Ada Lovelace Institute and UK AI Council, Exploring Legal Mechanisms for Data Stewardship, (March 2021)

health and care organisations; and efforts are underway to establish the Secure Data Environment as a secure data and research analysis platform.

Our projects sought to explore whether, within ongoing restructuring and development of data infrastructure in the NHS in England, Health Data Trusts could provide a form of governance of the sharing of health data for research which could help resolve the tension between the imperative to realise the benefits of health data for improving health and the challenges of doing so in a way which does not alienate the public.

Health Data Trusts

The General Practice Data Trust (GPDT) projects explored whether data trusts can offer a suitable solution to maximising researcher access to health data for beneficial research with appropriate respect for patients' preferences about the use of their data.

A Data Trust is a legal structure by which Trustees manage the sharing of individuals' (the Beneficiaries) data on agreed terms for the benefit of the Beneficiaries. The key features of a Trust are the fiduciary nature of the relationship between the Trustees and Beneficiaries, and the role of the court in providing additional safeguards by overseeing the Trust.

A Data Trust can be highly participatory, enabling individuals to exert control over the sharing of their data, but they can also involve delegating many of the decisions regarding data sharing to the Trustees. Therefore, in principle they can accommodate a range of people's preferences about the extent to which they would like to be involved in decision-making about this data sharing.

The fiduciary relationship is key as, rather than a contractual duty arising out of binding agreement between parties, it requires Trustees to act in good faith, to further the interests of the Beneficiaries, and to undertake its tasks with undivided loyalty. This is significant as it has the potential to address the particular vulnerabilities at stake with the use of data. The use of health data can create power imbalances between data users and data subjects. A fiduciary duty to represent and protect the interests of data subjects could be construed as including a duty to address those power imbalances, protecting individuals' ability to control their identity by controlling the data about them.[12]

12 For more discussion of these issues, see Delacroix, S. and Lawrence, N.D. 'Bottom-up data Trusts: disturbing the "one size fits all" approach to data governance. International Data Privacy Law', (2019). International Data Privacy Law 9(4) 236-252, https://academic.oup.com/idpl/article/9/4/236/5579842

Governance Approaches

In this project, we set out to establish what sorts of governance and technical approaches would need to be in place to engender public trust in health data sharing for research. Specifically, we wanted to know whether a Health Data Trust could provide the solution. Any Health Data Trust established would need to comply with existing and emerging health data governance frameworks.

Recent Recommendations on Health Data Sharing

The issue of suitable governance and technical frameworks is also a live question being considered by Government and other health data stakeholders. The central focus of the Goldacre and Morley review was to explore how to balance the preservation of patient privacy with the facilitation of access to NHS data by researchers and commissioners so that useful research and planning can take place. Their recommendations included to:

1. 'Build trust by taking concrete action on privacy and transparency: trust cannot be earned through communications and public engagement alone.

2. Ensure all NHS data policies actively acknowledge the shortcomings of "pseudonymisation" and "trust" as techniques to manage patient privacy: these outdated techniques cannot scale to support more users ...

3. Build a small number of secure analytics platforms - shared "Trusted Research Environments" - then make these the norm for all analysis of NHS patient records data by academics, NHS analysts, and innovators, wherever there is any privacy risk to patients, unless those patients have consented to their data flowing elsewhere....'[13]

The Sudlow report makes five key recommendations for the trustworthy use of health data:

1. 'Major national public bodies with responsibility for or interest in health data should agree a **coordinated joint strategy** to make England's health data **a critical national infrastructure**.

2. Leading government health and research bodies should establish a **national health data service** for England with **accountable senior leadership**.

3. The Department of Health and Social Care should oversee and commission a strategy for **ongoing coordinated engagement with patients, public, health professionals, policymakers and politicians**.

4. The health and social care departments in the four UK nations should set a UK-wide approach for data access processes and **proportionate data governance**.

5. National organisations in the four UK nations should develop a UK-wide system for **standards and accreditation of SDEs** holding data from the health and care system.'[14]

In relation to the governance of health data, the Sudlow Report states that:

> *"A UK-wide approach to streamline data access processes and foster* **proportionate and trustworthy data governance** *will enable more and better health data analysis and research. The aim should be for* **trusted** *researchers and analysts conducting* **responsible** *analyses* **in the public good** *to be able to rapidly access the de-identified data they need, while* **ensuring that data cannot be inappropriately accessed.**
>
> *The approach should be set by the health and social care departments of the UK's four nations and* **developed with patient and public involvement.** *It should confront legal and regulatory complexity by providing clear guidance on current approaches, proposing improvements that* **reduce unwarranted variation**, *and recommending where new or revised legislation is needed."*[15]

13 Op cit, 10

14 Sudlow C, Uniting the UK's Health Data: A Huge Opportunity for Society (November 2024). https://www.hdruk.ac.uk/helping-with-health-data/the-sudlow-review/ (emphasis added)

15 Ibid (emphasis added)

While our work supports this statement, the success of any health data sharing policy will depend on the detail and, crucially, the definition of the terms in bold above.

The Sudlow Report is likely to set the course for health data governance, at least in the medium term, making the work of the GPDT project all the more relevant in the contribution it can make to determining the meaning of proportionate and trustworthy data governance developed with and ensuring ongoing patient and public involvement.

The Structure of the Data Sharing Landscape

In this report, we do not provide a detailed account of the legal frameworks governing permissions to access health data under the UK GDPR and the Data Protection Act 2018. Rather, this report is based on the assumption that researchers seeking access to such data will be legally compliant.[16] It focuses instead on the following existing and emerging health data sharing technical and governance frameworks within which a Health Data Trust would need to function:-

National Data Opt-Out

The drive to develop a system to allow patients to opt out of sharing their healthcare data for research and planning came from the public's response to care.data and led in 2018 to NHS Digital's announcement of the 'national data opt-out programme' in England (NDOO). The NDOO allows patients to choose if they do not want their health data to be used for purposes beyond their individual care and treatment - for research and planning.

However, the NDOO does not apply in certain circumstances, including if the confidential patient information is being used to protect public health, or there is a legal requirement to disclose it, in some circumstances in which there is an overriding public interest to disclose it, and in some circumstances relating to research purposes.[17] Nor does it apply to anonymised health data.[18]

By November 2024, 3.61 million people in England had opted not to share their health data for purposes beyond their care. This equates to 5.4% of the population registered with a GP.[19] Opt-outs reduce the sample size and therefore accuracy of the dataset, especially as some groups are more likely to opt-out than others. For example, women are more likely to opt-out than men, and those in London and the North-West have the highest two opt-out rates regionally.[20]

The NHS Research Secure Data Environment (SDE) Network

The Goldacre and Morely review noted that the promise of the NHS dataset for improving understanding about health is hampered by the myriad ways and spaces in which it is captured and stored, not all of which are connected to each other, and recommended the setting up of 'Trusted Research Environments' to overcome these flaws. The *Data Saves Lives* strategy therefore included a commitment to a move to a system of 'data access as default' for the secondary uses of NHS health and social care data.

NHS England is pursuing a network of regional hubs which aims to make linked data accessible for research purposes. As part of this, funding was awarded to 11 teams to develop regional Secure Data Environments (SDEs) which together constitute the NHS Research SDE Network covering the whole of England.[21] NHS England stated that one of the five foundational layers of its digital delivery and transformation drive is 'Research Secure Data Environments to provide secure access to anonymised data for research and innovation, enabling patients and people to gain faster access for innovative treatments (medicines, med-tech, bio-tech, AI, vaccines and pharmacogenomics)'.[22] The NHS England Secure Data Environment now gives approved researchers conducting approved projects access to relevant de-identified data.[23]

16 For details on applicable legal frameworks see guidance provided by the Information Commissioner's Office at https://ico.org.uk/for-organisations/

17 NHS England, Understanding the National Data Opt-Out, (16 May 2023) https://digital.nhs.uk/services/national-data-opt-out/understanding-the-national-data-opt-out#:~:text=Setting%20or%20changing%20an%20opt,about%20their%20opt%2Dout%20choice.

18 Ibid.

19 NHS England, National Data Opt-Out (last edited 21 November 2024) https://digital.nhs.uk/services/national-data-opt-out#:~:text=opt%2D-out%20choices.-,Status%2C%20service%20level%20and%20current%20usage,to%20Friday%2C%20excluding%20Bank%20holidays.

20 NHS England, National Data Opt-Out Open Data Dashboard, https://digital.nhs.uk/dashboards/national-data-opt-out-open-data

21 The NHS Research Secure Data Environment Network https://digital.nhs.uk/data-and-information/research-powered-by-data/sde-network

22 NHS England, Transformation and innovation delivery update (1 February 2024) https://www.england.nhs.uk/long-read/transformation-and-innovation-delivery-update/

23 NHS England Secure Data Environment, https://digital.nhs.uk/services/secure-data-environment-service

SDEs are data storage and access platforms which protect privacy and security of health and social care data being used for research and analysis. They are 'designed to help provide safer, more secure and faster access to different types of NHS data as well as ensure consistent standards and interoperability.'[24] Data can be accessed and analysed by approved researchers without the data leaving the SDE. SDEs act in accordance with the *Secure Data Environment for NHS Health and Social Care Data Policy Guidelines*[25] which set out 12 requirements across the 'five safes' requirements for data protection:

- Safe settings - the environment prevents inappropriate access, or misuse
- Safe data - information is protected and is treated to protect confidentiality
- Safe people - individuals accessing the data are trained, and authorised, to use it appropriately
- Safe projects - research projects are approved by data owners for the public good
- Safe outputs - summarised data taken away is checked to make sure it protects privacy.

It is intended that, '[i]n improving access to NHS data, the SDE Network will enable world class research to prevent, diagnose, and treat our biggest healthcare challenges. This will improve patient care, support innovation and help us sustain the NHS in the future.'[26]

The SDE Network incorporates:

- The NHS England SDE, which supports health and social care research of approved organisations; and
- 11 interoperable regional SDEs across England, each bringing together integrated care boards (ICBs) with local universities and industry groups to build on research partnerships.

A Community of Practice (SDE C of P) has been established to connect SDEs and any partner organisations. The SDE C of P has set up a number of working groups to consider matters such as 'information governance and ethics; technology infrastructure; and communications and patient and public involvement and engagement.'[27]

Integrated Care Systems and Secure Data Environments

Integrated Care Systems (ICSs) were established in 2022 and are partnerships that bring together NHS organisations, local authorities and others to take collective responsibility for planning services, improving health and reducing inequalities across geographical areas.[28]

ICSs have four aims:

- Improving outcomes in population health and health care
- Tackling inequalities in outcomes, experience and access
- Enhancing productivity and value for money
- Helping the NHS to support broader social and economic development.[29]

Each ICS has one integrated care board (ICB). Each ICB has access to their own version of the NHS Federated Data Platform to connect their information in a safe and secure manner and to collaborate, and coordinate and plan care.[30]

24 NHS England, Accessing data for research and analysis, . https://transform.england.nhs.uk/key-tools-and-info/data-saves-lives/secure-data-environments/accessing-data-for-research-and-analysis/#:~:text=The%20NHS%20Research%20SDE%20Network%20is%20an%20interoperable%20England%2Dwide,ensure%20consistent%20standards%20and%20interoperability.

25 Department of Health and Social Care, (December 2022) https://www.gov.uk/government/publications/secure-data-environment-policy-guidelines/secure-data-environment-for-nhs-health-and-social-care-data-policy-guidelines#what-a-secure-data-environment-is

26 https://transform.england.nhs.uk/key-tools-and-info/data-saves-lives/secure-data-environments/how-will-secure-data-environments-be-delivered/ (accessed 29 October 2024)

27 https://transform.england.nhs.uk/key-tools-and-info/data-saves-lives/secure-data-environments/how-will-secure-data-environments-be-delivered/

28 NHS England, What are Integrated Care Systems? https://www.england.nhs.uk/integratedcare/what-is-integrated-care/

29 Ibid.

30 NHS England, How does the NHS Federated Data Platform Work? https://www.england.nhs.uk/digitaltechnology/nhs-federated-data-platform/learn-about-the-fdp/how-does-the-nhs-federated-data-platform-work/

Public Trust, Transparency and Involvement

In principle, health data can legally be shared for purposes that are considered a public task and where the data is necessary for research purposes. Patient consent for data sharing is not a *legal* requirement in this situation, in fact the GDPR sets out that consent would not be the appropriate legal basis where an imbalance of power exists between the data subject and the data controller, for example, where the data controller is a public authority and the data subject depends on the services it provides.[31] Therefore, consent would not be the appropriate legal basis for the processing of confidential patient data by the NHS.[32] However, there are other reasons why consent should be sought at least in some way, including helping to ensure that researchers are complying with the social licence for research.

It is likely that if consent is sought in an appropriate way, it will be given as, in addition to the policy drive to enable access to health data, there is a lot of patient support for this too.[33] However, a significant gap in policy documents which consider access to health data for research, is *how* patients can be given more choice and control over how their data is used for research purposes beyond a broad opt-out.[34] This is despite participants in the public workshops conducted as part of the Sudlow review emphasising 'the importance of patients having ownership and control over their health data…'[35] Our earlier work found that in addition to patients needing to know that the system will keep their data secure, it will also deal with it in a way that reflects their values, something that has been called 'trust in commitment'.[36]

As seen above, failing to properly embed respect for patients' choices can lead to breaches of the social licence for research, and a disengagement by the public with data sharing initiatives.

In response to the acknowledged mistakes made in implementing GPDPR, the Government revised *Data Saves Lives*, stating that in the updated version:

> *'We are putting public trust and confidence front and centre of the safe use and access to health and social care data. The data we talk about is not an abstract thing: there is an individual, a person, a name behind each piece of data. That demands the highest level of confidence. It is their data that we hold in trust and, in return, promise to use safely to provide high-quality care, help improve our NHS and adult social care, develop new treatments, and, as a result, save lives.'*[37]

This aspect of the strategy led to the setting up of the Data Strategy Advisory Panel,[38] an independent stakeholder advisory group, and a programme of public engagement.[39] There is also an ambition in due course to establish a Data Pact to set out clearly and simply how the NHS and social care uses health and care data and what the public has the right to expect. It is intended that this would give the public confidence that the health and care system is a trustworthy custodian of data.[40]

It is this policy environment that our work takes place in and hopes to shape, by offering insights into practical solutions for how data governance can offer individuals more choice and control over their health data.

31 Paragraph (43) GDPR – Recital 43: https://gdpr-info.eu/recitals/no-43/

32 See [HRA GDPR Guidance] https://www.hra.nhs.uk/planning-and-improving-research/policies-standards-legislation/data-protection-and-information-governance/gdpr-guidance/

33 See for example, Understanding Patient Data, How do People Feel about the Use of Data? https://understandingpatientdata.org.uk/how-do-people-feel-about-use-data; and Cheshire and Merseyside Secure Data Environment Patient and Public Engagement, The Use of Administrative Health Data for Research (May 2024)

34 See for example the Sudlow report and Macon-Cooney B et al, A New National Purpose: Harnessing Data for Health The Tony Blair Institute for Global Change (21 May 2024) https://institute.global/insights/politics-and-governance/a-new-national-purpose-harnessing-data-for-health

35 Sudlow Report, 192.

36 Wolfensberger, M. and A. Wrigley (2021) 'Trust in medicine : its nature, justification, significance and decline.'

37 Op cit

38 NHS England, Data Strategy Advisory Panel, https://transform.england.nhs.uk/key-tools-and-info/data-saves-lives/data-strategy-advisory-panel/#:~:text=The%20Data%20Strategy%20Advisory%20Panel,and%20NHS%20England%20Transformation%20Directorate.

39 NHS England, Large-scale public engagement - Steering group terms of reference, https://transform.england.nhs.uk/key-tools-and-info/data-saves-lives/national-public-engagement-on-the-use-of-health-data/large-scale-public-engagement-steering-group-terms-of-reference/

40 NHS Transformation Directorate, 'Response to 'Developing a Data Pact', a Report on the Public Engagement Undertaken by the Patients Association' (27 September 2023) https://transform.england.nhs.uk/key-tools-and-info/data-saves-lives/response-to-developing-a-data-pact-a-report-on-the-public-engagement-undertaken-by-the-patients-association/

Data Stewardship

Given the high level of public support for health data research that is being frustrated by a lack of public trust in the current approach, innovative approaches to data governance are gaining attention as a potential solution.

This challenge prompted the Ada Lovelace Institute to produce a report in 2021, exploring the legal mechanisms for data stewardship more broadly. It stated that:

> *'The challenges of the twenty-first century demand new data governance models for collectives, governments and organisations that allow data to be shared for individual and public benefit in a responsible way, while managing the harms that may emerge.'*[41]

There are three main models for how this could be done: Data trusts, data cooperatives, and corporate and contractual mechanisms. Our work focuses on the Data Trust option, which the Ada Lovelace report defines as follows:

> *'…a vehicle for individuals to state their aspirations for data use and mandate a trustee to pursue these aspirations. Data trusts can be built with a highly participatory structure in mind, requiring systematic input from the individuals that set up the data trust. It's also possible to build data trusts with the intention to delegate to the data trustee the responsibility to determine what type of data processing is to the beneficiaries' interest.*
>
> *The distinctive elements of this model are the role of the trustee, who bears a fiduciary duty in exercising data rights (or the beneficial interest in those rights) on behalf of the beneficiaries, and the role of the overseeing court in providing additional safeguards.*
>
> *Therefore, data trusts might work better in contexts where individuals and groups wish to define the terms of data use by creating a new institution (a trust) to steward data on their behalf, by representing them in negotiations about data use.'*[42]

One way of instilling trust in applicable governance systems is for a formal person or body to be appointed to represent patients' data sharing preference. Research has shown public support for some sort of central authority overseeing sharing.[43] This activity has been termed "data stewardship". We define health data stewardship as,

> *'A trusted participatory system, overseen by an intermediary accountable to the community of beneficiaries, which is responsible for the preservation and management of health data and/or the rights relating to them, and for the promotion and protection of the values and aspirations of data subjects'.*[44]

This chapter has provided an outline of the current landscape within which a Health Data Trust would need to work. We now move on to summarise our previous findings in relation to patient views on their health data.

41 Ada Lovelace Institute, 'Legal Mechanisms for Data Stewardship' 2021, available at https://www.adalovelaceinstitute.org/report/legal-mechanisms-data-stewardship/#fn-6

42 Ada Lovelace Institute, 'Legal Mechanisms for Data Stewardship' 2021, available at https://www.adalovelaceinstitute.org/report/legal-mechanisms-data-stewardship/#fn-6

43 Neumann V, Davidge G, Harding M, Cunningham J, Davies N, Devaney S, et al. (2023) Examining public views on decentralised health data sharing. PLoS ONE 18(3): e0282257. https:// doi.org/10.1371/journal.pone.0282257

44 Bartlett B et al, 'Health data stewardship: achieving trust through accountability in health data sharing for research' Law, Innovation and Technology (2024) 16(2), 517–557

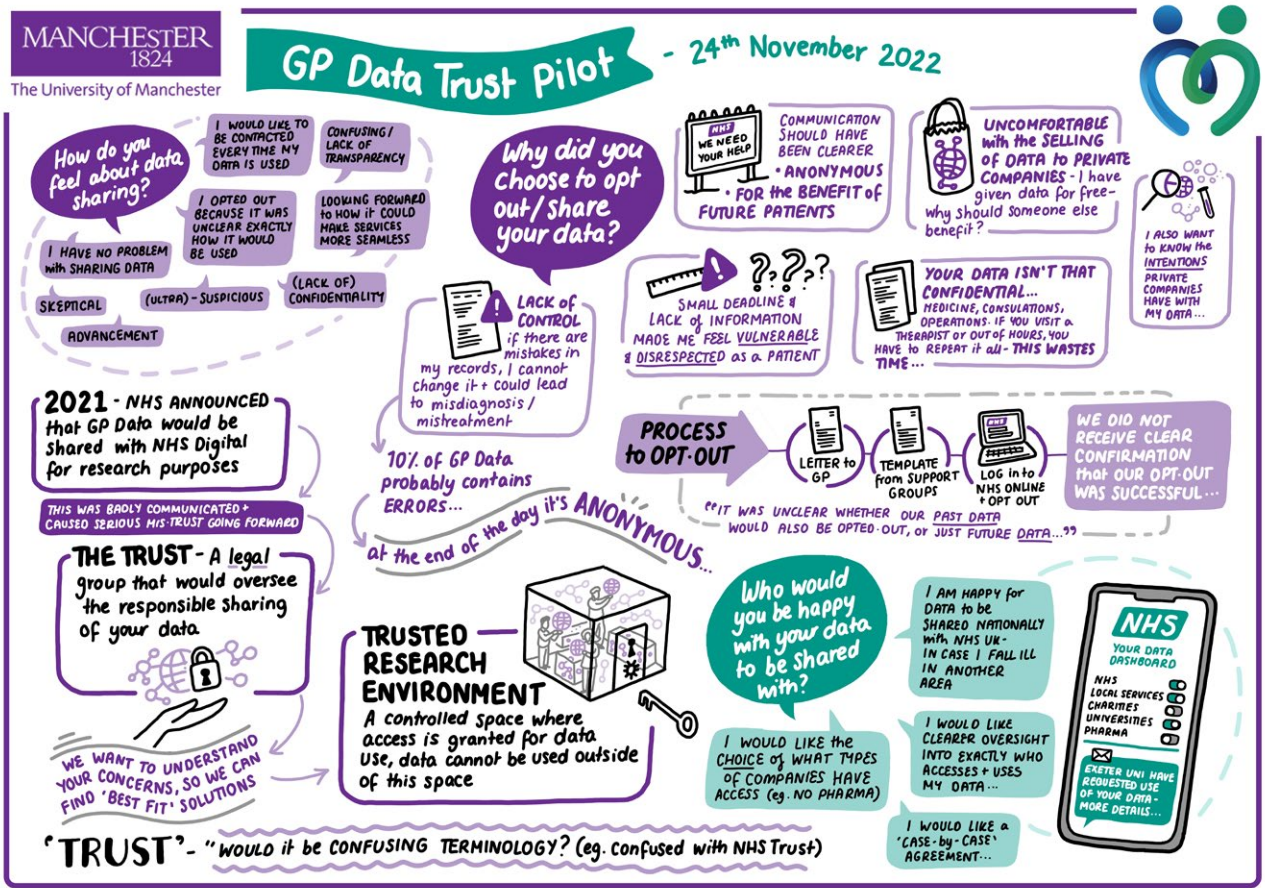


Figure 3: Animated minutes of focus group on 24 November 2022

Chapter 2: GP Data Trust Pilot Project

In response to the large scale opt-out from NHS Digital's GPDPR programme, in the first phase of our GP Data Trust pilot study, we carried out surveys, focus groups and interviews with patients, GPs, and other stakeholders including campaign groups and others exploring solutions to this problem. Our aim was to investigate why people opted out of sharing their GP data in this way. We also wanted to explore the nature of the concerns people had, and how law and regulation might respond in order to alleviate those concerns and enable medical research to be as inclusive as possible.

This image represents one of the Focus Group discussions:

Some of the key points in the focus group discussions were:

» Communication around the GPDPR should have been clearer with more information about what was being done and why.

» People had been made to feel vulnerable and disrespected by the way the GPDPR programme was handled.

» Concern was expressed over data being shared with commercial organisations. The intentions of these organisations were not understood, nor was it clear who would benefit from the research.

» There was also concern over the accuracy of health records. Participants had identified mistakes in their medical records and were concerned that research using them would lead to an inaccurate picture.

» Overall, participants wanted to know how their data would be used and they wanted it to be their decision.

These issues were also explored in interviews with people who indicated that they had opted out of sharing their GP data with the GPDPR programme, and with other stakeholders exploring solutions to this problem.

Phase 1 Findings

Trust and control

All participants in the interviews in our phase 1 study expressed a desire to contribute their data to medical planning and research. However, they had reluctantly opted-out of doing so due to a lack of trust that their data would be managed in a way that was acceptable to them. The GPDPR programme complied with the relevant legal requirements, but this was not sufficient to indicate trustworthiness for the participants in our study; the requirements of the social licence went beyond the legal requirements.

The main reason for this lack of trust was the perception that the data was not treated as belonging to patients or as being 'my data' in the way that the programme was presented. Very little information had been available about the GPDPR programme; participants were not told who the data would be shared with, what the benefits would be, and who would be responsible for keeping the data secure. The short timeframe of one-month patients had to register an opt-out before data collection would begin, led to feelings of being pressured and rushed. This, along with the opt-out rather than opt-in nature of the decision, caused participants to feel that they were not given the opportunity to make an informed, considered choice. This was important in developing the sense that health data was not being seen or treated as 'my data' and led to many people's decision to opt-out of sharing their health data.

Anonymised data

Another important finding from phase 1 of our study, was that even if primary care data could be fully anonymised, many of our participants felt that this would not weaken their personal connection to the data.

> *'It comes down to, regardless of the anonymisation, my data is my data and I feel that even if it's anonymised, I still have a right to say how it's used and where it goes'.*

This is an important finding as it indicates that the current efforts to foster trust in health data sharing through ensuring anonymity, are unlikely to be sufficient for those who share our participants' views.

Ethical considerations

One of the reasons that participants in phase 1 of our study wanted to control the use of their data was that they were concerned about the purposes for which their data might be used. For some, this concern extended to fully anonymised data that related to them; they wanted to ensure that any data that related to them was not used for purposes that would conflict with their religious or ethical beliefs.

This can be explained by understanding that for these participants, 'my data' refers to the data being connected to their sense of identity and individuality; it is part of what makes them who they are. Once health data is understood as belonging to individuals in this constitutive sense, the importance of being able to control the purposes one's data is used for becomes clear; if an individual's health data is used for purposes she finds objectionable, this might cause her to feel as if *she* is being used for purposes she finds objectionable.

In light of these key findings, the next step was to explore how the sharing of health data for research purposes could be governed so that it reflects peoples' fundamental connection to their data by giving them the control over it they had told us they wanted.

Chapter 3: General Practice Data Trust Project Phase 2

In this chapter we set out the aims, activities and findings of Phase 2 of the GPDT work in 2024 (GPDT2).

Aims and Objectives of Phase 2

By the end of our pilot stage project, it was apparent that the next key areas of enquiry lay in technical and governance areas.

Our core aims were:

- From a technical perspective, to develop and trial a patient-facing dashboard to facilitate the expression of choices and preferences in relation to sharing health data for research purposes.
- From the governance perspective, to explore the detail of the type of legal structure appropriate for this task, and the elements such a structure would need to contain in order to maintain the trust of people sharing their data for research purposes.

Overview of General Practice Data Trust Phase 2 Activities

GPDT2 project undertook a series of activities designed to determine how patient preferences and our technical research could be translated into a functional prototype dashboard.

- The process began with the development of use cases informed by the findings from the first GPDT project. These use cases outlined the primary interactions, workflows, and decision-making processes which would be required within the portal, providing a basis for the design and implementation of key features. The goal was to ensure that the portal could address the needs of patients, researchers, and governance bodies in managing consent and health data sharing.
- A prototype dashboard was designed.
- A stakeholder workshop was held in partnership with Prospect Brixham to explore practical considerations around establishing data trusts, using GPDT2 and Prospect Brixham as case studies. This workshop examined broader questions relating to data trust governance, privacy, and operationalization, highlighting similarities and differences between the two case studies. Feedback on the portal prototype, presented during the workshop, was used to inform revisions to its functionality and to gather feedback on the direction and complexity of choices made around the representation and presentation of consent options. These revisions included adjustments to improve usability and alignment with stakeholder expectations.
- Following the workshop, two focus groups were conducted to explore user interactions with the portal in greater depth. A pilot focus group presented the proof-of-concept and centred on how participants engaged with the hierarchical representation of consent preferences. Based on the outcomes of this session, three user scenarios were developed to structure discussions in the second focus group. These scenarios illustrated the complexities of consent decisions and framed discussions around the usability, ethical considerations, and technical challenges of the portal. Feedback from these focus groups informed further refinements to the prototype and highlighted areas for future development.
- We set out our findings in a blog post with Understanding Patient Data.[45]

We provide further detail on these activities below.

Activity 1: Use Case Analysis for Portal Development

Overview of Use Cases

The initial design of the GPDT2 Portal was guid-

45 Bowden C, Devaney S, Cunningham JA, Health data education is the key to unlocking the benefits of health research, https://understandingpatientdata.org.uk/news/health-data-education-key-unlocking-benefits-health-research

ed by a series of use cases that reflected the intended interactions and user journeys within the system. These use cases aimed to capture both typical and complex scenarios to ensure the portal could accommodate a range of requirements and preferences from different stakeholders, including patients, trustees, and researchers.

The primary use cases focused on the following activities:

- User registration;
- Consent management;
- Patient interactions with data preferences;
- Governance activities; and
- Project request workflows.

Use Case 1: User Registration and Setup

This use case covered the initial on-boarding process for new patient users of the dashboard. The goal was to streamline the identification, consent verification, and account setup steps, ensuring users could quickly understand and engage with the portal.

Key features developed were:

- Identification verification using NHS details.
- Account creation with strong security measures (e.g., two-factor authentication).
- Initial consent preferences setup, allowing users to specify their data sharing preferences at registration.
- Introduction to the trust's governance structure and policies.

Design considerations informing this development were:

» Ensuring a secure registration process while minimizing user friction.

» Providing clear explanations of governance and consent options to inform user choices.

Use Case 2: Data Preferences Management

The portal was designed to enable users to easily configure and update their data sharing preferences. This use case focused on giving users control over how their data could be used, including options for specific research categories and individual projects.

Key features developed were:

- A dedicated section for managing data preferences, accessible via the user dashboard.
- Options to consent to different categories of research (e.g., publicly funded, pharmaceutical, privately funded).
- Ability to opt-in or opt-out of specific projects requesting access to data managed by the trust.
- Real-time updates and conflict-highlighting when user selections override or conflict with governance decisions.

Design considerations informing this development were:

» Providing a clear interface for users to express consent preferences and easily review or change their choices.

» Incorporating visual indicators and dynamic feedback to highlight consent conflicts or overrides.

Use Case 3: Patient Interaction and Activity Review

This use case emphasized the need for transparency and accountability by allowing users to review the activities performed on their data. It covered functionalities that let patients audit data usage, understand governance decisions, and receive notifications about new projects.

Key Features:

- Activity logs displaying a detailed history of actions taken on user data (e.g., access by researchers).
- Notifications about new projects requesting access to user data.
- A governance section providing updates on recent decisions, board member actions, and policy changes.

Design considerations informing this development were:

- » Ensuring users have visibility into how their data is being used and by whom.
- » Offering a clear communication channel for users to provide feedback or ask questions about data usage.

Use Case 4: Governance and Decision-Making by Trustees

This use case outlined the functionalities needed by trustees to oversee the governance of the trust. It focused on enabling trustees to review project proposals, communicate with patients, and make collective decisions on data access.

Key Features:

- A tailored dashboard for trustees, featuring tools for proposal review, decision-making, and communication.
- Voting mechanisms for project approvals, with automatic calculations of results based on voting rules.
- Documentation tools for creating reports and records of governance actions.

Design considerations informing this development were:

- » Ensuring trustees have efficient access to decision-support tools.
- » Providing transparency and traceability of governance activities for accountability.

Use Case 5: Project Request and Patient Review Workflow

This use case described the process for managing project requests from researchers, including the assessment of these requests, patient notifications, and consent collection.

Key Features:

- Detailed project request forms submitted by researchers, including information on research goals, data requirements, and ethical considerations.
- Notifications to patients about new project requests, with options to review project details and provide consent.
- Feedback mechanisms for patients to ask questions or provide comments on project requests.

Design considerations informing this development were:

- » Balancing the need for comprehensive project information with a user-friendly interface for patient review.
- » Ensuring that patient consent decisions are accurately captured and respected.

Summary

The use cases defined in the initial stages of the project formed the basis for the technical design and implementation of the GPDT2 Portal. They provided a comprehensive framework for addressing the needs of different user groups, focusing on secure registration, flexible consent management, patient engagement, governance support, and transparent data access workflows.

Activity 2: Design of the GPDT2 Portal Prototype

The GPDT2 portal was designed as a central component of a future Health Data Trust to facilitate transparent and efficient management of consent to health data sharing. Its purpose is to empower users to make decisions about sharing their data which are informed by factors such as who the data will be shared with and for what research purpose. It also provides tools for governors to manage and oversee trust operations. By demonstrating features addressing the complex ethical, technical, and legal challenges associated with data governance, the portal aimed to serve as a proof-of-concept demonstration of a technical implementation of the Health Data Trust's broader objectives.

The design of the portal was intended to balance simplicity and functionality for users with the complexities inherent in managing granular consent options. It aimed to accommodate users with varying levels of expertise by presenting clear, categorized consent preferences while ensuring that advanced features were accessible for those requiring finer levels of control. Additionally, the portal provided trustees and administrators with tools to manage studies, review data requests, and maintain transparency and accountability in governance processes.

The portal was structured around three main functional areas:

1. The Registration Process, which guides new members through joining the trust and setting their initial preferences;
2. The Admin Portal, which facilitates governance activities and the management of data requests; and
3. The User Dashboard, which provides members with an interface to manage their consent preferences, track recent activity, and access governance information.

Registration Process and Onboarding

The registration process for the GPDT2 portal was designed to introduce potential members to the Health Data Trust and guide them through the necessary steps to join. It incorporates structured stages to ensure users are informed about the trust's purpose, governance structure, and the implications of their consent decisions.

The registration begins with an **introduction to the trust**, including background information on its goals, ethical framework, and governance. This section outlines the role of members within the trust and provides details about how their data is managed and the processes for participating in the trust's governance activities.

The next stage covers **consent decisions and their impact**, detailing how user preferences affect the use of their data in research projects. This stage includes information on high-level consent options, such as following governance decisions, and more granular choices, including opting in or out of specific projects. A **Q&A section** is included to address frequently asked questions and is intended to clarify uncertainties about the trust and participation.

The process then moves to **legal agreements**, which outline the trust's data-sharing policies and the rights and responsibilities of members. Another section explains the **identifiable data retained by the trust**, describing its role in maintaining membership records and verifying consent. The registration completes the membership process with the **submission of user preferences and account creation**.

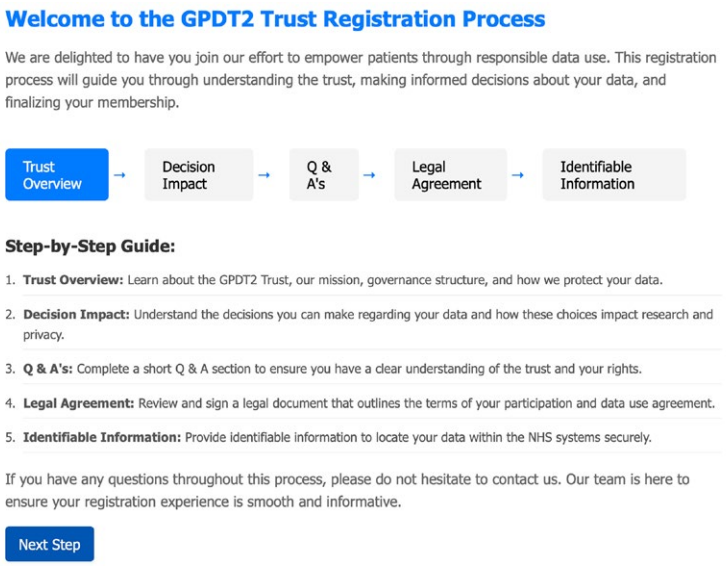


Figure 4: The Welcome page of the GPDT2 Prototype Portal

Governance Portal

The governance portal is designed to provide trustees with tools to manage and oversee the trust's operations. It focuses on key aspects of governance, including the approval of studies, the management of data access requests, and the dissemination of governance information to members.

The **study approval section** allows trustees to review and manage applications for data access from researchers. Trustees can examine project proposals, consider their alignment with trust goals and ethical standards, and either approve or reject them. The decisions are recorded for transparency and are reflected in the user dashboard.

The **governance information section** provides details on the trust's structure and operations. This includes information about the board members, recent governance decisions, and policies in effect. This section ensures that members have visibility on the actions and decisions of the trust's leadership.

The **data access request management section** enables trustees to monitor ongoing requests, assess feedback from members regarding specific projects, and track how data is used within approved research. Importantly it also allows those overseeing the Trust to actively engage with the preferences expressed by members, for example by obtaining feedback from members on specific projects, allowing them to raise concerns and ask for further details etc. This section therefore goes beyond typical approaches which offer a one-off binary choice between opt-in and opt-out by facilitating interaction between trustees and members, allowing for governance that incorporates user input and maintains trust transparency.

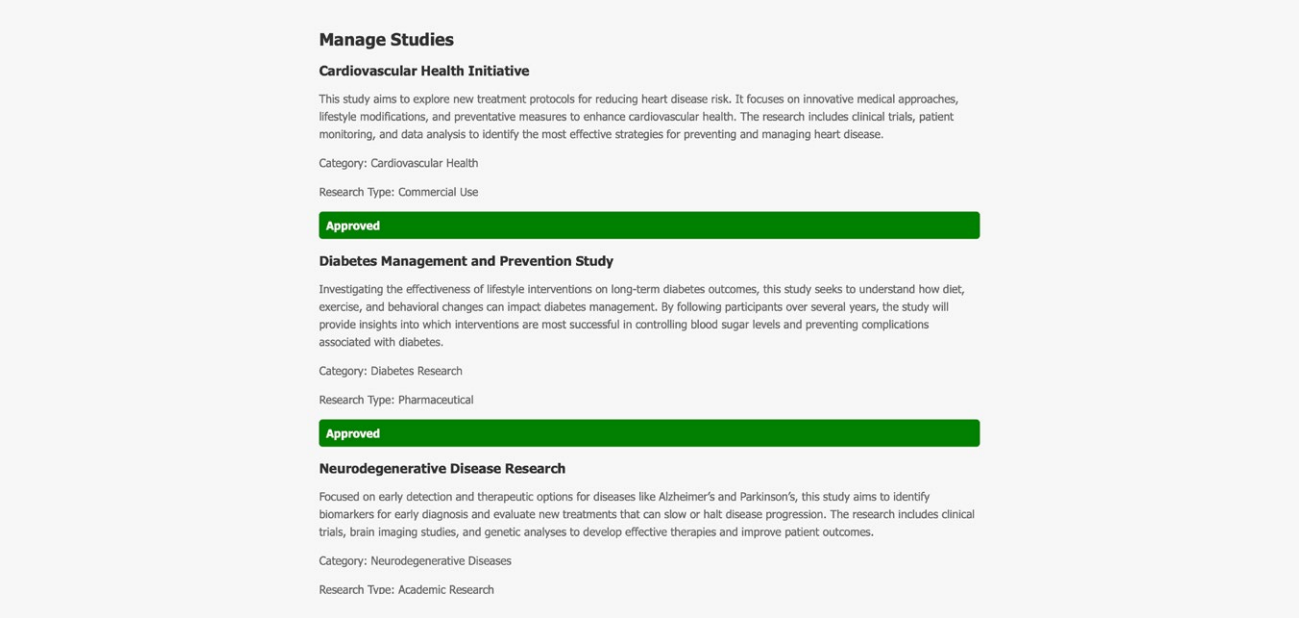


Figure 5: Example page setting out summaries of research to which health data is being provided

User Dashboard

The user dashboard is designed as a central hub for patients as members of the trust to review and manage their data preferences, monitor activities related to their data, and access information about the trust's governance and operations. The dashboard is structured into several key sections, each addressing a specific aspect of member engagement.

Your Data Preferences

This section allows members to set and adjust their consent preferences in a hierarchical manner. It is organized into four categories:

- **Meta Consent**: Members can select overarching preferences for the levels of consent they want to give, such as following governance decisions or managing consent on a case-by-case basis. This will determine the preferences they are then asked to provide.
- **Consent for Research Types**: Members can specify whether they consent to their data being used in publicly funded, pharmaceutical, or privately funded research.
- **Consent for Categories of Research**: Members can define preferences for specific research areas, such as mental health or epidemiology.
- **Consent for Specific Projects**: Members can approve or decline participation in individual projects requesting data access.

Real-time feedback is provided in this section, highlighting potential conflicts between preferences so that users can adjust their choices if necessary.

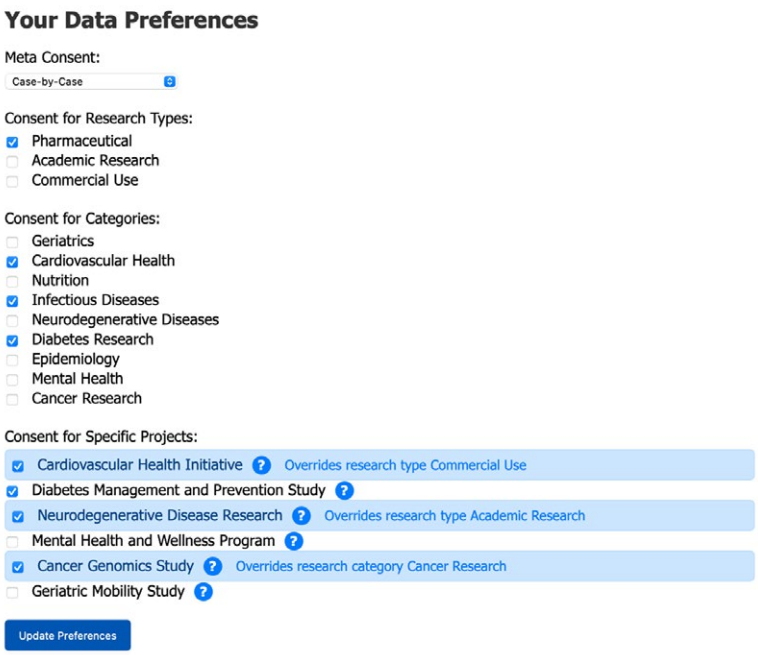


Figure 6: The Data Preferences section of the GPDT2 prototype portal

Recent Activity

This section provides a chronological log of activities associated with the member's account. It includes:

- Changes made to meta consent and consent preferences.
- Requests for data access by research projects, along with their approval or rejection status.
- Recent actions taken by the trust's governors, such as policy updates or project approvals.

The activity log is designed to keep members informed about how their data is being used and how their preferences are being applied.

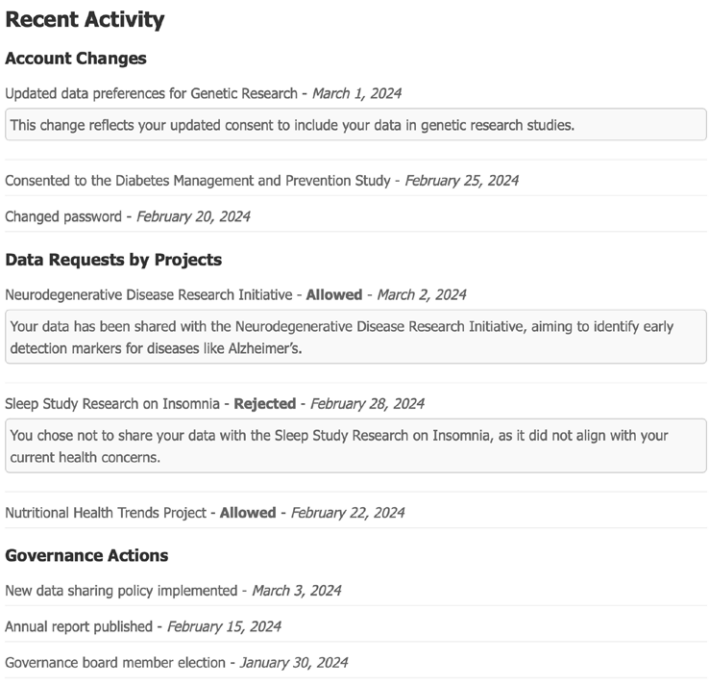


Figure 7: the Recent Activity page of the GPDT2 prototype portal

Governance Information

This section provides an overview of the trust's governance structure and recent actions taken by the trustee board. It includes:

- Profiles of board members.
- Summaries of recent governance decisions.
- Updates on policies and initiatives undertaken by the trust.

This section ensures transparency and keeps members informed about the trust's operations.

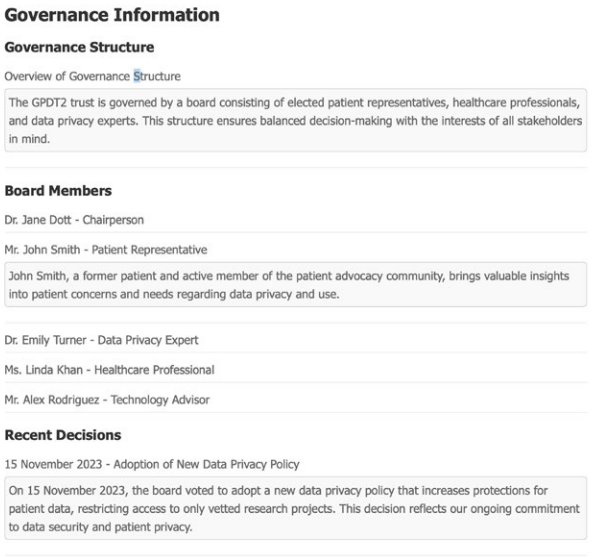


Figure 8: Governance Information page of the GPDT2 prototype portal

Data Audit

The data audit section provides members with a detailed log of how their data has been accessed or queried over time. It includes:

- Dates and purposes of data access.
- Information about the projects or researchers who accessed the data.
- Details of actions taken on the member's preferences.

This section offers an in-depth view of data usage, ensuring members have a clear understanding of how their data is being utilized.

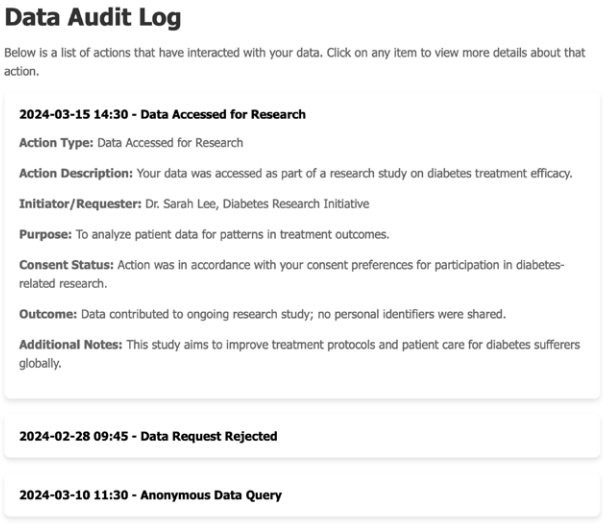


Figure 9: Data Audit log page of the GPDT2 prototype portal

Activity 3: Stakeholder Workshop

On 21 May 2024, in partnership with Prospect Brixham[46], we held a one-day Stakeholder Event, 'Developing Data Trusts in Practice'. This was an exploratory event which aimed to draw on the views, experiences and expertise of stakeholder participants in response to questions arising across research work on Data Trusts. The group examined questions relating to some of the practicalities around establishing Data Trusts using GPDT and Prospect Brixham as case studies to highlight differences and similarities across different types of data trusts and to enable reflection both on next steps in their development, and on the future for data trusts in a wider context.

Visual minutes for the event were created by Amber Anderson.

The event started with an overview of the Brixham and GPDT data trust projects:

46. https://prospectbrixham.org/

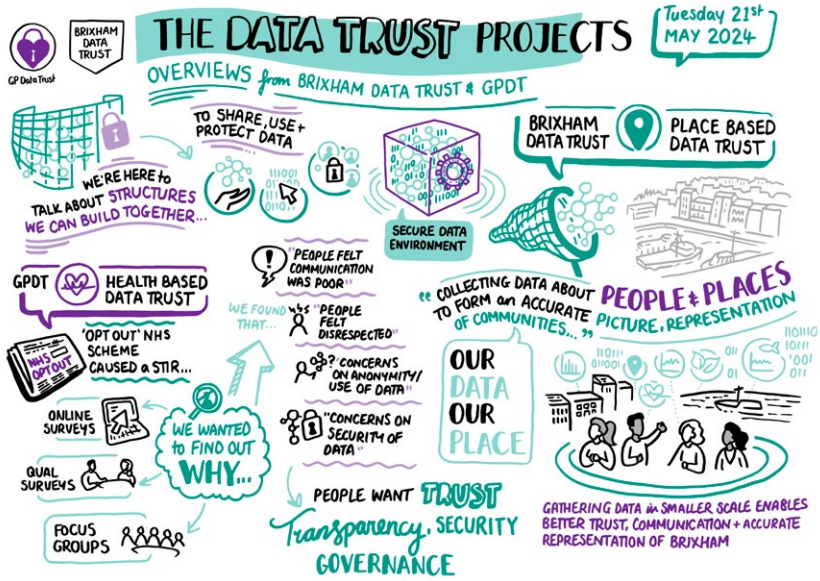


Figure 10: Visual minutes of overviews from the Brixham Data Trust and GP Data Trust project

Alex Krook then provided insights into her work on 'Digital Anthropology – research on data and place-making through the lens of the Brixham Data Trust pilot project':

Figure 11: Visual minutes of presentation on research on data and place making

In the presentation and subsequent roundtable discussion, the following themes emerged:

- The **pervasiveness** of data throughout people's lives, as well as its potential as a **collective asset** to improve people's lives.
- The fundamental importance of **data literacy** about data sharing's potential effects on people and populations arises out of this.
- The governance mechanism of a **one-off opt-out** for data sharing decisions is **not nuanced enough** to reflect people's choices and preferences;
- New governance models for both enabling **democratic participation** and **protecting the vulnerable** in data sharing are needed;
- These models need to be capable of **facilitating real choice**.

We then turned to considerations which arise if data trusts are to be established as charitable organisations. The potential benefits of using a charitable regulatory vehicle as a mechanism to establish and run such a trust were outlined by Caroline Redhead; Sylvie Delacroix discussed the implications of professionalising data trustees; and Stuart Wood provided the Charity Commission view on CIOs in the data trust context:

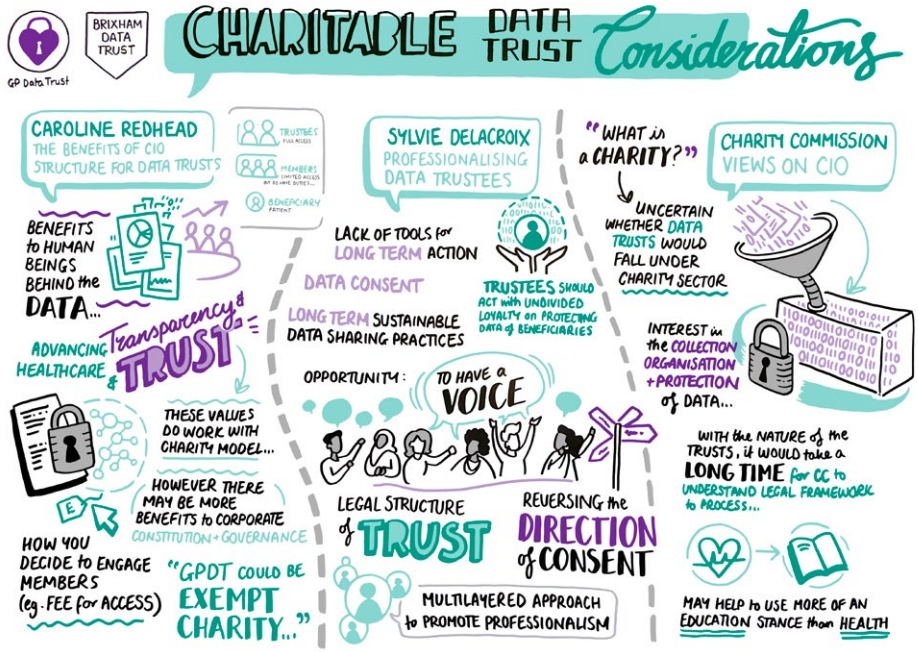


Figure 12: Visual minutes of session on considerations for charitable data trusts

Themes emerging from these presentations and the following discussion:

- **Long-term, sustainable data** sharing practices which empower data subjects are required;
- Regulatory structures which **protect the vulnerable and embody accountability and enforceability** are needed;
- How data trusts would sit within existing charitable regulatory provisions and approaches remains to be seen – this would be **a novel structure** which would have to be carefully scrutinised by the Charity Commission to establish whether it could be registered with them;
- Regulatory oversight needs to bear in mind the **conflicting needs** of data which might arise between its subjects and those who wish to use it for research;
- **Investment in data curation** is important.

During the roundtable discussion of charitable data trusts, key requirements of transparency, balancing of power imbalances as well as risks and uncertainties were highlighted:

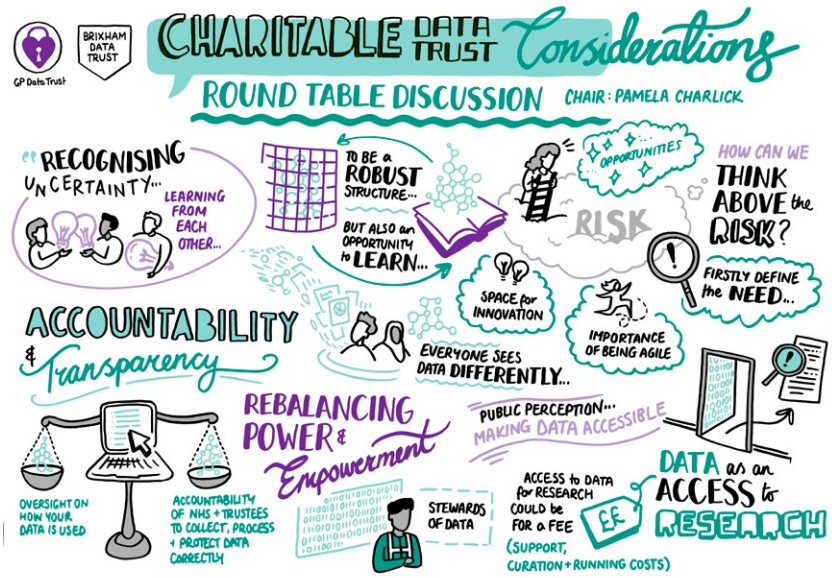


Figure 13: Visual minutes of round table discussion of charitable data trusts

Stakeholders then considered legal and technical challenges in the implementation of data trusts, with a demonstration of a potential interactive online portal for expressing data sharing preferences:

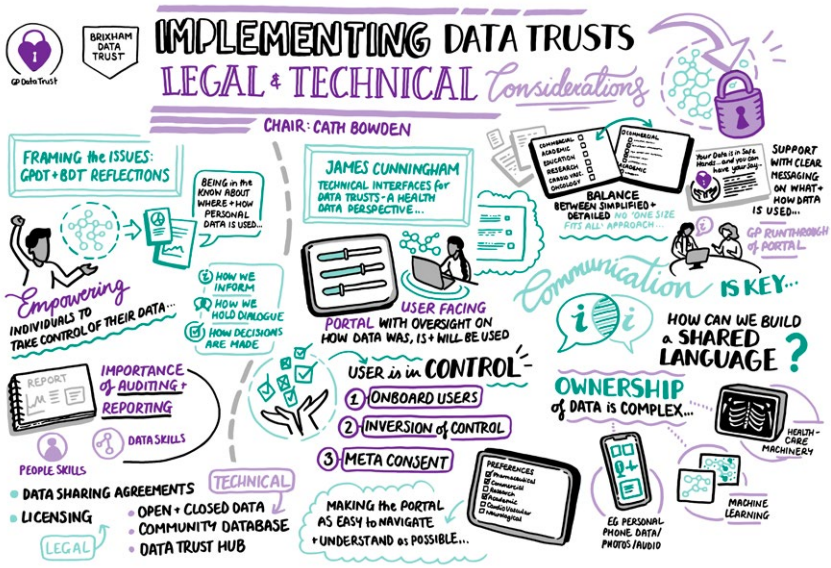


Figure 14: Visual minutes of legal and technical considerations for data trusts

Three key points emerged from the feedback on the demonstration and review of the portal:

- **Ease of Navigation and Understanding**: Participants emphasized the importance of ensuring the portal is intuitive and user-friendly. Simplified navigation and clear presentation of options are critical to encourage users to engage with and fully understand their data sharing preferences.
- **Complexity of Data Control**: The workshop highlighted the inherent complexity of data control and the diverse perspectives users bring to the portal. While some users prioritize simplicity and ease of use, others may require fine-grained control over how, when, and why their data is used. The portal must cater to this spectrum of needs.
- **Balancing Simplicity and Control**: A balance must be struck between simplified presenta-

tion for accessibility and the ability to configure detailed, fine-grained preferences. Striving for clarity without sacrificing the depth of control ensures users can make informed decisions that align with their personal values and priorities. This balance is also crucial for maintaining trust and transparency.

The discussions during the closing reflections session highlighted key areas of concern, potential, and further enquiry:

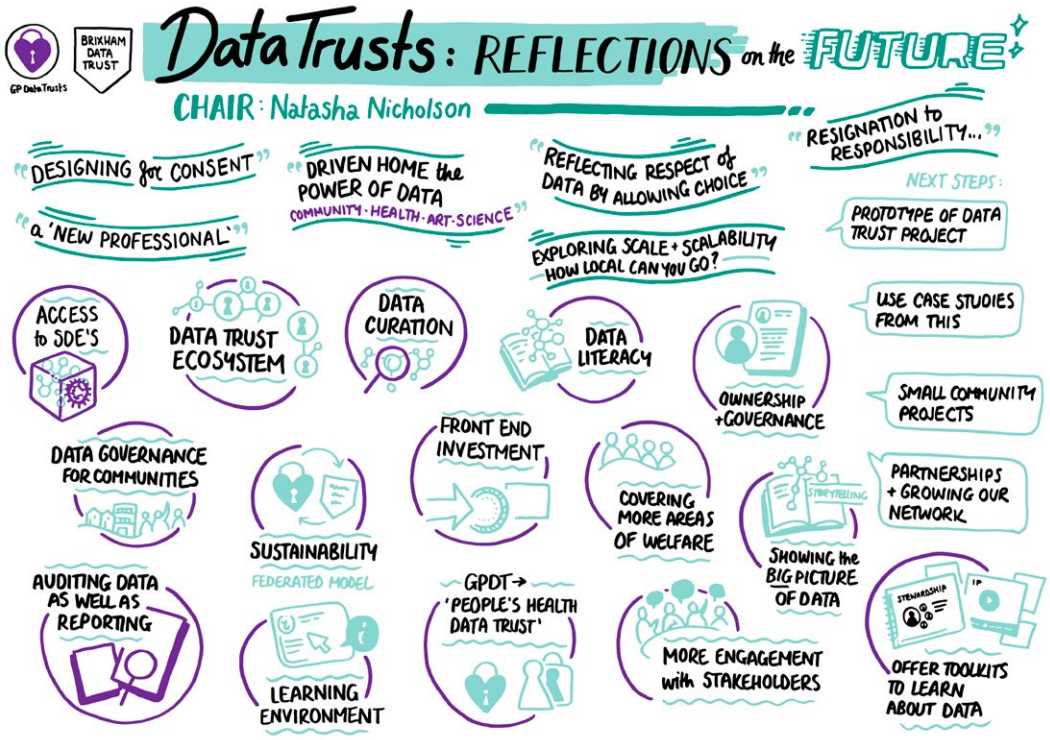


Figure 15: Visual minutes of closing session discussion of stakeholder workshop

Rather than harnessing the economic value of data, what stood out was the potential for discussions about data to act as a catalyst for bringing communities together to have their say about what is important to them, to shape projects that use their data, and to share in the benefits of those projects. It was felt that Data Trusts present some key challenges and opportunities including:

- **A clear, shared purpose:** Articulating a clear, shared need for collecting and using data is vital for trust and support. This is too often the need of the researcher rather than the need of the community.
- **Engagement across the community:** Specific places and projects can be used as a portal for engaging people in conversations about data.
- **Data literacy:** A Data Trust is an opportunity to, and has a responsibility to, inform and educate people about the use of their data.
- **Developing new habits of participation:** A Data Trust can only challenge existing power imbalances if people can participate in new ways. Expecting them to participate in existing ways will only reinforce those imbalances.
- **Transparency and trust:** One of the challenges is how to provide information in a way that is meaningful and not burdensome.
- **Enforcement:** To be effective, enforcement must have teeth; it must have sufficient powers to require the Trustees to improve. Requiring individuals to monitor what is done with their data and challenge those that do not fulfil their responsibilities, is unfair and unworkable.

Activity 4: Pilot user focus group

While the presentation of the portal was positively received at the stakeholder event, it was important to trial this with potential patient users. We tested out a focus group approach in a pilot focus group facilitated through the UseMYData group.

Recruitment and demographics

Participants were recruited via an open call in the UseMYData weekly members' newsletter.[47] Membership of UseMYData is made up of two categories:

- **Members** - patient advocates who are either patients, relatives or carers
- **Associate Members** – clinicians, researchers, charity workers, academics, public and commercial sector workers.

Although one Associate Member expressed an interest in taking part, we limited participation to Members only to maintain consistency across the groups and focus on the patient user perspective. All of the Members that responded to the call could be accommodated in the sessions. Although there was no selection based on demographic criteria, participants aged between 30 and 70, there were 3 men and 14 women, and a range of ethnicities.

Pilot Focus Group Design

The pilot focus group was designed to evaluate the usability and functionality of the prototype portal, with a particular focus on the user dashboard and its data preferences section. The session aimed to explore participants' perspectives on the hierarchical consent options offered by the portal, and how users might navigate the complexities of managing consent across multiple levels. By focusing on the data preferences section, the session sought to identify potential challenges in understanding and using these features, particularly for individuals with varying levels of familiarity with data governance.

The session included a demonstration of the portal's core functionality, emphasizing the hierarchical consent model. The data preferences section was highlighted to show how users could set consent at multiple levels, such as high-level preferences, research types, research categories, and specific projects. The demonstration was designed to illustrate how these levels interacted, including scenarios where preferences might overlap or conflict. For example, participants were shown cases where approving a research category could contradict a decision to reject specific projects within that category.

To guide the discussion, a series of questions and prompts were prepared to examine the practical and ethical implications of the consent mechanisms. Participants were asked to consider the appropriateness of the complexity presented and whether the options made sense within the broader context of a data trust. The discussion was intended to explore the balance between providing users with detailed control over their data with ensuring that the interface remained accessible and comprehensible. Attention was also given to the visual presentation of consent overrides and conflicts, encouraging participants to reflect on whether these features were effectively communicated.

The design of the session anticipated that feedback would provide insights into how users understand and engage with the data preferences section. The focus on hierarchical consent options was intended to identify areas for refinement in the portal's design, ensuring it met user needs while addressing the complexities inherent in managing data sharing preferences. Insights gathered from the pilot focus group were intended to inform subsequent iterations of the portal and shape the structure of the main focus group.

Pilot Focus Group Outcomes

The pilot focus group was designed to gather user feedback on the GPDT2 portal's proof-of-concept design, with an emphasis on understanding how users approached consent preferences and interacted with the system.

A central theme emerging from participant feedback in the pilot focus group was the importance of users fully understanding the implications of the decisions they were making. Participants expressed concerns about the clarity of the consent structure and stressed the need for transparency in how preferences were defined and applied. This highlighted the importance of

47 https://mailchi.mp/usemydata/patient-data-and-engagement-round-up-12729999?e=675d6abcf2 [mailchi.mp]

ensuring that the portal could effectively communicate complex concepts in an accessible way.

Discussions during the focus group frequently shifted toward the specific details of the consent preference structure rather than its overall utility or purpose. Participants engaged with the portal as if it were a finished product rather than as the proof-of-concept model that had been introduced. Consequently, much of the feedback focused on specific features and categorizations, such as how mental health data was grouped within the portal. For example, participants felt that the mental health category required greater nuance and a more accommodating structure to reflect the sensitive and varied nature of this data. While highly valuable, this feedback diverted attention from the broader objectives of the workshop i.e. to establish how any potential conflicts between preferences should be resolved.

The technical framing of the first focus group also limited the depth of discussion around the broader implications and use cases of the portal. Participants were less inclined to consider the ethical positioning and potential applications of the tool, focusing instead on immediate, tangible features. This outcome suggested that the format and framing of the workshop may have overly constrained the participants, leading to a narrower scope of discussion than intended. Although the feedback provided was relevant to the portal's design, it was less aligned with the overarching goal of evaluating its conceptual and ethical positioning.

Activity 5: Scenario development for future focus group

In response to the outcome of the pilot focus group, we reformulated the main focus group to centre around a series of scenarios. This approach aimed to provide a more structured and contextualized framework for participants to explore the portal's purpose and applications. Rather than presenting the technical design directly, scenarios were developed to illustrate a variety of choices individuals could make. This shifted the focus from the specific options presented to individuals, to how the Trust should deal with the levels of consent provided. This led to a broader discussion about the ethical, technical, and practical dimensions of the tool.

Developing the Focus Group Scenarios

We employed a loose, iterative methodology to develop three user scenarios to better understand the implications of granular consent options for data sharing within the GPDT. The scenarios were designed to illustrate real-world decision-making processes that individuals might go through when determining their preferences for sharing their health data for research purposes. The scenarios aimed to capture the complexity of consent decisions and provide a platform for discussing user attitudes and ethical considerations.

The scenarios allowed us to explore how users with different backgrounds, levels of expertise, and motivations may interact with consent options, and how these interactions align with the overall goals of the trust. This approach provided a structured way to identify potential challenges in user decision-making and assess the adequacy of the consent options in addressing diverse user needs and expectations.

Our approach began by reviewing the outcomes of the initial phase of the project, which included user feedback and early insights into the technical design of a platform designed to capture consent preferences, along with technical constraints identified during the development of the proof-of-concept portal. Using this information, we constructed scenarios that aligned with the intended use cases of the portal.

The scenarios were designed to reflect real-world situations where users would encounter complex decision points regarding data sharing preferences. We focused on scenarios that we identified as having the greatest potential for highlighting ethical considerations, technical issues, and the complexities involved in making granular consent decisions. Each scenario was tailored to explore different aspects of the consent framework, aiming to uncover areas where user decision-making might be challenged or where conflicting consent preferences could arise. The scenarios were designed to be both representative of potential user experiences and effective means of exploring the broader implications of the consent model in the GPDT.

Aim of the Scenarios

The primary aim of the scenarios was to investigate the decision-making processes of users when presented with the option to tailor their data sharing preferences. We sought to explore the practical and ethical implications of offering granular consent controls, examining how users might navigate these choices in different contexts. By presenting a variety of scenarios, we aimed to highlight the range of possible interactions with the consent model, including cases where users' preferences might conflict with default settings or where they may need to balance competing considerations.

Each scenario was constructed to address specific aspects of the consent framework:

1. **Ethical Dilemmas**: We aimed to identify situations where users' consent choices might lead to ethical tensions, such as opting out of certain categories of critical research while still engaging in with others. These scenarios were designed to explore how users perceive their role in the research process and the ethical responsibilities associated with data sharing.

2. **Technical and Usability Issues**: We used the scenarios to test the usability of the consent interface, particularly when users were required to make complex choices or override existing preferences. This allowed us to evaluate how intuitive the consent settings were and whether the system provided adequate feedback on the implications of users' choices.

3. **User Expectations and Validity of Choices**: The scenarios aimed to assess whether the choices made by users were aligned with their stated preferences and expectations. By examining cases where users might make inconsistent or conflicting decisions, we explored how well the portal's design supported users in making informed and valid consent choices.

Through this approach, the scenarios provided a structured way to evaluate different aspects of the consent framework, offering insights into potential areas for improvement in both the ethical design and technical implementation of the portal.

Scenario 1: Emily – explore the tensions between higher level and more granular consents

Emily is a 35-year-old healthcare professional who actively participates in research data sharing but exercises caution due to concerns about privacy and potential conflicts of interest. Emily chose to set her high-level consent preference to "Follow Governance Trustees' Decisions," indicating a general trust in the oversight and ethical governance of the data-sharing process within the GPDT2 Trust. However, she made specific exceptions, opting out of sharing data in certain categories and projects where she felt there was a higher risk of privacy issues or conflicts with her professional obligations.

Emily's primary concerns centered on the sensitive nature of data related to mental health and infectious diseases. She excluded both of these categories from her consent preferences, reflecting her discomfort with the potential misuse of sensitive health information, especially when it might carry stigmatizing implications or be used outside of her intended scope. Additionally, Emily opted out of participating in the "Local Diabetes Research Project," as it involved her own workplace, and she wished to avoid any potential conflict of interest.

Reasons for Constructing the Scenario:

This scenario was constructed to explore the complexities and tensions between broad, general trust in a governance framework and the desire for granular control over specific data-sharing decisions. By choosing to follow the trustees' decisions while simultaneously excluding sensitive categories and projects, Emily's scenario illustrates the conflict that can arise between high-level consent choices and detailed, context-specific privacy concerns.

The inclusion of publicly funded open research and pharmaceutical research, while opting out of private research and certain specific projects, highlights the nuanced attitudes a healthcare professional might have towards data sharing. It demonstrates a willingness to contribute to public and pharmaceutical research that aligns with Emily's professional values, while avoiding privately funded research, which she perceives as having less transparency and greater potential for commercial misuse.

This scenario also tested the system's capacity to manage and communicate these nuanced preferences, particularly when overarching consent choices (e.g., following governance decisions) are overridden by more specific user selections. The emphasis on categories like mental health and infectious diseases, which Emily excluded, aligns with known areas of heightened privacy sensitivity, allowing us to evaluate how the system handles and highlights these exceptions.

List of Consents:

- *High-Level Consent:*
- *Follow Governance Trustees' Decisions*
- *Types of Research:*
- *Publicly Funded Open Research: Included*
- *Pharmaceutical Research: Included*
- *Privately Funded Research: Excluded*
- *Specific Categories:*
- *Mental Health: Excluded*
- *Infectious Diseases: Excluded*
- *All other categories: Included*
- *Specific Projects:*
- *Local Diabetes Research Project: Excluded*
- *All other approved projects: Included*

Scenario 2: John — Balancing Privacy with Selective Public Health Support

Scenario Description:

John is a 45-year-old high school teacher who initially chose to withdraw all consent for data sharing, reflecting his strong prioritization of personal privacy. However, he decided to make specific exceptions for research projects that, in his view, offer a clear and direct public health benefit. John's approach illustrates a careful balance between maintaining his own privacy and supporting selected research initiatives that align with his values, particularly when these projects are transparent about their data usage.

John's decision to opt out of all general types of research and specific categories, but to selectively consent to individual projects, indicates a high level of scrutiny towards data sharing practices. He has chosen to participate in the "Long Term Effects of Childhood Autism" project and the "University Diabetes Study," both of which have well-defined public health goals and clear, transparent policies on data use. These choices reflect his support for projects that have a strong potential for societal benefit, while minimizing his involvement in broader data sharing categories that he perceives as riskier or less transparent.

Reasons for Constructing the Scenario:

This scenario was developed to investigate the attitudes of users who take a conservative approach to data sharing, opting out broadly but making exceptions based on perceived public health value. John's scenario highlights the complexities faced by individuals who value privacy highly but are still willing to contribute to select research endeavours that meet their ethical standards. It explores how users navigate the tension between a strong preference for privacy and a desire to support beneficial research when transparency is provided.

By opting out of all categories and types of research, except for specific projects with clear objectives, the scenario allows us to evaluate the system's capacity to manage granular, project-specific consents that override broader preferences. This approach also tests the ability of the system to clearly communicate these exceptions and ensure that users understand the implications of their choices. The inclusion of projects related to childhood autism and diabetes research provides examples of studies with evident public health benefits, making the scenario realistic and relevant for ethical discussions.

Explicit List of Consents:

- *High-Level Consent:*
- *Withdraw All Consent*
- *Types of Research:*
- *Opted out of all types of research*
- *Specific Categories:*
- *Opted out of all categories*
- *Specific Projects:*
- *Long Term Effects of Childhood Autism: Included*

- *University Diabetes Study: Included*
- *All other projects: Excluded*
- *High-Level Consent*

Scenario 3: Lisa — Selective Opt-In with Concerns About Private Profit Motives

Scenario Description:

Lisa is a 28-year-old graphic designer who has decided to approach her data sharing preferences on a case-by-case basis. She has generally chosen to opt in for most research types and categories, reflecting her belief in the value of contributing to research for the public good. However, she has explicitly opted out of projects funded by private companies due to concerns about potential profit motives. Lisa's selective consent approach indicates a cautious yet supportive stance towards data sharing, driven by her personal values and scepticism towards the involvement of commercial entities in research.

Lisa's choices reflect her willingness to participate in publicly funded open research and pharmaceutical research projects, where she perceives a clearer alignment with public health objectives. She has opted out of all privately funded research as well as specific projects that she believes may prioritize profit over the well-being of participants, such as the "Private Healthcare Usage Research Project." Her approach illustrates a user who is engaged and deliberate about their consent choices, willing to contribute broadly while maintaining reservations about certain types of research.

Reasons for Constructing the Scenario:

This scenario was created to explore the attitudes of users who, while supportive of the broader goals of research, remain wary of commercial interests and their potential impact on ethical practices. Lisa's case-by-case approach highlights the complexities of making granular consent decisions and provides insight into how users balance general support for research with specific concerns about funding sources.

The scenario examines the system's ability to handle nuanced, individualized consent decisions, especially when users choose to engage selectively rather than making broad, all-encompassing choices. By opting in to publicly funded and pharmaceutical research but opting out of privately funded projects, the scenario tests the system's capacity to accurately reflect user preferences and handle exceptions effectively. It also raises questions about transparency, trust, and the role of profit motives in research, providing a basis for ethical discussion during the focus group session.

Explicit List of Consents:

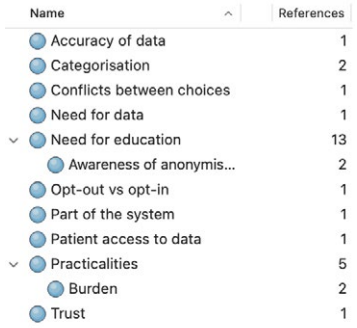
- *High-Level Consent:*
- *Case-by-Case (explicitly reviewing each decision)*
- *Types of Research:*
- *Publicly Funded Open Research: Included*
- *Pharmaceutical Research: Included*
- *Privately Funded Research: Excluded*
- *Specific Categories:*
- *Opted in for all categories*
- *Specific Projects:*
- *Private Healthcare Usage Research Project: Excluded*
- *All other projects: Included*

Activity 6: Main Focus Group

The patient-facing portal was tested in a focus group using the scenarios set out above. The scenarios were presented to participants, who were asked to comment on the efficacy of the portal in facilitating expression of consent to health data sharing.

Analysis

A thematic analysis was undertaken using NVivo software of notes taken in the first session and the transcript of the second session. A number of themes were identified, with the Need for Education and Practicalities being the most common.



Name	References
Accuracy of data	1
Categorisation	2
Conflicts between choices	1
Need for data	1
Need for education	13
Awareness of anonymis...	2
Opt-out vs opt-in	1
Part of the system	1
Patient access to data	1
Practicalities	5
Burden	2
Trust	1

Focus Groups Summary of Findings

In these sessions, participants shared their views on the portal as a way for individuals to express their preferences regarding whether, who by, and for what purposes, their health data could be used for research.

Key themes arising:

» **Central role for a data steward:** Overall, participants were supportive of the idea of an intermediary or Trustee making decisions about who would be granted access to health data. Concern was expressed about the burden that a system offering this level of granular consent would place on individuals. It was felt that the process of expressing preferences could be overwhelming and this might cause more disengagement, reducing the data available for research. Being able to engage with, or follow the choices and recommendations of the Health Data Trust trustees would help here.

» **Support for the facilitation of preference expression**: Participants liked the idea of being able to express preferences over why and by whom their data would be used.

» **Preference expression should include choices about types of research health data is to be shared with:** A number of participants told us that although they were keen for their data to be used for research, there are some types of research that they would not want their data used for, for example, military research or research involving animals. One participant expressed this as wanting data to be 'used with a conscience'.

» **Health data literacy is crucial:** The primary concern expressed by the participants in both sessions was the need for education around health data and health data research. It was felt by some participants that the public would not have sufficient understanding of how health data is used in research, the importance of health data research, and the safeguards that are in place, to enable them to make meaningful decisions. This was expressed as a concern that misinformation and misunderstandings would lead to opt-outs, and the view that we should not be accommodating choices based on misunderstanding. Calls were made for public health education campaigns to be run on health data literacy. This need for education was also expressed in relation to the safeguards preventing data misuse, identification of data subjects and what data anonymisation means.

» **Sustainability of health data trusts:** Some participants thought that the practicalities of administering a Data Trust offering individuals this level of control, would be unworkable and require a "vast and permanent bureaucracy".

» **Research classifications:** Another key area of concern was the way in which research would be classified in the portal. One participant thought for example that having multiple categories of research into physical health, but only one for research into mental health, indicated a lack of parity in how mental and physical health are perceived.

It is evident from what our participants told us that there are significant challenges in offering individuals this level of individual control over their health data, notably the challenges in categorisation and the level of knowledge and understanding required to make meaningful choices over how, why, and by whom your health data is used.

Therefore, it may be that delegated decision-making should play a greater role in a Patient Health Data Trust and indeed, the fiduciary duty placed on the Trustees may require this. To achieve fulfilment of such a duty, an education programme would be needed to enable individuals to exercise agency (in setting some level of individual preferences, setting the terms of the Trust, choosing between Trusts, and playing a role in collective decision-making within the Trust).

Chapter 4: Recommendation

Aims and Objectives

By the end of our pilot stage project, it was apparent that the next key areas of enquiry lay in technical and governance areas.

Our core aims were:

- From a technical perspective, to develop and trial a patient-facing dashboard to facilitate the expression of choices and preferences in relation to sharing health data for research purposes.
- From the governance perspective, to explore the detail of the type of legal structure appropriate for this task, and the elements such a structure would need to contain in order to maintain the trust of people sharing their data for research purposes.

In order to answer these questions we:

1. Developed a series of 5 Use Cases to facilitate exploration of the needs a Health Data Trust technical interface would need to meet;
2. Designed the GPDT2 Portal Prototype;
3. Tested our technical and governance approaches at a stakeholder workshop and two focus groups with members of the public.

Recommendations

Health Data Literacy

Data is pervasive throughout people's lives. Health data has enormous potential as a collective asset to improve people's lives. Health data literacy is fundamentally important so that people can understand about data sharing's potential effects on people and populations.

1. **A national public education campaign on health data and its use in research is vital. This needs to include information about the types of organisations which undertake research using health data, and how they work together. Improvements in health data literacy across the population would support health data trustees in fulfilling their fiduciary duty to represent the informed wishes of beneficiaries.**

Mechanisms for making choices about health data

The governance mechanism of a one-off opt-out for data sharing decisions is not nuanced enough to reflect people's choices and preferences, or to correct current power imbalances within the data sharing landscape.[48]

2. **New, sustainable governance models for both enabling democratic participation and protecting the vulnerable in data sharing are needed. These models need to be user-driven and capable of facilitating real choice, and ensuring accountability and enforceability if things go wrong.**
3. **The question of how data trusts would sit within existing charitable regulatory provisions and approaches needs to be tested through an application to set up a charitable Health Data Trust with the Charity Commission.**

The Health Data Trust

4. **A Health Data Trust or Trusts should be co-created with patients and other stakeholders. This co-creation would need to determine whether one or several Health Data Trusts are required to represent the expression of values and facilitate choice. The Health Data Trust/s should act as a mechanism to facilitate patients' choices about the sharing of their health data for research. This should be:**
    - » Intuitive to use, simple to navigate and present clear options;
    - » Capable of facilitating a wide range of preferences. Where preferences cannot be followed, for example because they conflict with other choices made, the reasons for this should be clearly explained;
    - » Based on a clearly articulated shared patient and researcher need for collecting and using data.
    - » Overseen by a board of trustees which is committed to the definition of data stewardship set out above and representative of the patient voice.
    - » Based on clear terms, co-produced by patients and stakeholders, with clear and effective enforcement mechanisms.
    - » Committed to providing information back to data providers on the ways in which their data has been used, and what the research using it has led to.

48 Alexandra Giannopoulou, Jef Ausloos, Sylvie Delacroix, Heleen Janssen, Intermediating data rights exercises: the role of legal mandates, International Data Privacy Law, Volume 12, Issue 4, November 2022, Pages 316–331, https://doi.org/10.1093/idpl/ipac017

Implementing Health Data Trusts

General Practice Data Trust Project Report

Sarah Devaney, Cath Bowden and James Cunningham

March 2025

www.socialsciences.manchester.ac.uk/csep/research/projects/gpdt-pilot-study