February 2024

# Understanding Patient Data: “What words to use when talking about health data” Rapid Evidence Review

## Contents

# Background

This project is one of the first undertaken by the new iteration of Understanding Patient Data, now hosted at the NHS Confederation. This project was launched in September 2023, with Research Works as the chosen supplier starting work in December 2023, following a competitive process. The purpose of the project as a whole is to:

- Gain insight into members of the public's understanding of Trusted Research Environments (TREs), Secure Data Environments (SDEs), and other similar data spaces that are used to access health data for research and analysis, and associated concepts (such as federation)
- Consider members of the public's feelings towards these topics
- Co-develop specifications for public-facing resources based on the above, with the type of resource being recommended by the workshop participants
- Produce the necessary resources

Whilst many organisations that have TREs, SDEs, and other similar environments have public-facing information about their product, there is a lack of independent, non-organisation-specific resources that can be used by anyone. This is important given that there are potentially over 68 of these environments in existence in the UK.[1] There is also limited evidence on how people feel about these environments in general, and what they would like to know about them.

This rapid evidence review is the first output of this project. Its objectives are to:

- Synthesize existing evidence and provide context for this research
- Inform research stimulus for co-creation sessions

Identify gaps to be explored in the co-creation workshops

This review has been conducted at a relatively high level, and often refers to 'the public'. UPD is keenly aware that there is no one 'public', but this term is used to highlight general prevailing views and attitudes. Through the workshops, and other projects, UPD considers the views of, and collaborates with, members of different communities in society.

We hope that this rapid evidence review will be useful for others working in this space at a local, regional, national and international level. More information about the project can be found here.

**For brevity, this report will refer to 'TREs' since this is the term used in most of the literature reviewed. However, its findings are still applicable to Secure Data Environments, Data Havens, Data Spaces, etc. This report includes a discussion about the different names used for these types of data environments, which refers to issues that have been identified with the word 'trusted' and 'research' and the reasons for different language choices.**

---

[1] DARE UK, 2023; *UK Sensitive Data Research Infrastructure: a Landscape Review*, https://dareuk.org.uk/wp-content/uploads/2023/11/2310_DARE_UK_DigInfraLandscapeReview_Final.pdf

# Summary

Below is a high-level summary of the findings and recommendations of the rapid review.

## Key findings about the understanding of TREs, SDEs and data spaces

- Early consultations about TREs suggested the public and patients are **broadly supportive of this direction of travel.** TREs were perceived to address some key concerns over health data security.
- At the same time, there were some areas where the public and patients felt **more work was needed to engender trustworthiness** in health data access and usage via TREs.
- Beyond general statements of public approval of TREs, there has been **limited research and evidence on how the public and patients feel about specific aspects of TREs.**
- Furthermore, **public and patient views on commercial use of health TREs have been mixed**. Better explanation of commercial data use has been identified as a way of providing transparency and support public understanding of controls.

➢ **Public and patient calls for improved lay information about TREs have been a running thread through consultations on this topic. The need for better communications has also been a staple part of all recommendations for future work on TREs.**

## Recommendations for explaining TREs, SDEs and data spaces

- The **variety of names** has been deemed confusing for public understanding of TREs and has given rise to debates over the best terms to be used.
- The debate is also striking in its focus on **speculative implications of different terms**, rather than exploring how these different terms were understood and perceived by different groups of the public and patients.
- The **Five Safes Framework** has been broadly accepted as the overarching conceptual basis for educating the public about TREs.
- Recommendations for explaining TREs to the public and patients have highlighted two overarching goals for such communications - **highlighting benefits of data use and addressing concerns over data security and confidentiality -** but also a range of other specific issues including data linkage and cloud-based storage.

➢ **The range of 'other' specific issues identified has highlighted the importance of thinking about communications in terms of different points in the data journey e.g. beginning and end.**

## Challenges in communicating about TREs, SDEs and data spaces

- Certain principles for communicating about TREs have been highlighted as potentially effective:
    - Not assuming any prior technical knowledge
    - Using visual representations to show data processes
      Demonstrating rather than explaining physical and logical controls by taking the public through the processes of data collection, access, handling and use.

- The examples found reflect some of the recommended principles for communicating about TREs, but there may be instances where some of the principles are more challenging to follow. For example:
    - Some of the case studies reviewed and also diagrams showing data flows may require prior knowledge.
    - More broadly, the lack of consistency in naming such data environments also fails to meet the need for consistent language that would aid public understanding.
    - Finally, some of the specific information needs identified in this review, for example, around measures to prevent identification through data linkage or technical controls on cloud platforms, have not been addressed in the TRE-related public communications identified in this research.

## Evidence Gaps

- There has been a focus on recommendations for effective communication on TREs, rather than exploring responses to existing materials and improving them. This is particularly of note regarding the debate around what they should be called.
- There is a lack of knowledge around what resources members of the public and patients want to see, beyond introductions to specific TREs and explanations of the Five Safes framework.
- There are difficulties in meeting information needs relating to technical processes and terminology and nuances of using data for different types of activities (i.e. where individuals are not satisfied by the basics).

## Next steps

The next steps involve a series of in-person and online workshops. The first set of workshops will include an introduction to the use of health data and key concepts, including TREs, and testing some existing resources as a stimulus. The second set of workshops will be more co-creation focused, developing recommendations for resources and identifying specifics like language choice, resource type, and style. These recommendations and specifications will then be tested further with members of the public, health and care professionals, and key stakeholders.

# Introduction

Over recent years, Trusted Research Environments (TREs) have become the dominant paradigm for health data access. TREs (also known as Secure Data Environments, Safe Havens, Secure Research Environments and by other names) allow approved researchers to access data and conduct analysis in safe computing environments, without data leaving them.[2] A number of recent papers have identified the key factors that have led to this development:[3]

- The Covid-19 pandemic underlined the need for greater and more timely access to health data. This accelerated the development of health data TREs being used to inform public health interventions, manage the pandemic and enable research to deliver public benefit.
- By comparison, data sharing models where data is disseminated to researchers have been perceived as higher risk. Once data leaves its original environment, data custodians cannot control access and use of this data.
- Data protection legislation introduced high penalties for failing to protect personal data, making data custodians potentially more risk averse about sharing data.
- The data dissemination model is comparatively costly and inefficient, particularly when large data sets need to be shared with multiple individual researchers.
- Recent advances in computing have facilitated the development of centralised data environments that support the data science capabilities required by researchers.
- Feedback from patients and the public suggested that they would be more comfortable with TREs as a model for health data access rather than data sharing where data is distributed to individual researchers and organisations.

While a number of individual TREs have been created and operated effectively over the last decade, this recent shift promoted TREs to becoming the main model for accessing and using NHS health data. The 2022 Goldacre Review commissioned by the UK's Department of Health and Social Care (DHSC) recommended TREs as a 'clear path forward' to creating trustworthy and secure health data systems.[4] Following the review, the 2022 UK government 'Data Saves Lives' strategy committed to the implementation of TREs in order to facilitate safe and efficient health data usage and research for

---

[2] For brevity, this report will refer to 'TREs' since this is the term used in most of the literature reviewed. This report includes a discussion about the different names used for these types of data environments, which refers to issues that have been identified with the word 'trusted' and 'research' and the reasons for different language choices.

[3] Graham M, Milne R, Fitzsimmons P, et al., 2023, "Trust and the Goldacre Review: why trusted research environments are not about trust", *Journal of Medical Ethics*, 49:670–673; UK Health Data Research Alliance (HDRA), 2020, *Trusted Research Environments (TRE): A strategy to build public trust and meet changing health data needs.*

[4] 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

public benefit.[5] The strategy is backed by a £200 million investment to enable more secure and effective use of NHS data through TREs.[6]

Work to develop strategy, principles and standardised specifications for health data TREs has often involved public and patient engagement. These consultations suggest that there is broad public support for the direction of travel i.e. for using TREs to access health data. However, there is limited understanding of public awareness, understanding, perceptions of, and attitudes towards different aspects of TREs in the context of health data. In addition, there are a wide range of terms used to refer to such data environments and a lack of agreement about the most suitable terminology to use. There is no consensus on how to explain TREs to the public in a way that could relate to different individual TREs.

In this context, there is a need for a more targeted exploration of these questions to improve understanding of how key data concepts and aspects of health data access through TREs should be explained to the public. This research has been commissioned by Understanding Patient Data (UPD) to address this gap in knowledge and help to inform resources that can be used by different organisations to explain how health data TREs enable secure and effective health data access and use for the benefit of the public. As previous research suggests, increasing public understanding of how health data is used and protected through TREs is likely to be critical to engendering trustworthiness and support for these developments.

## Research Objectives

The purpose of this rapid evidence review was to **synthesize existing evidence** about key research questions to **inform the next stage of co-creation research**. The review aimed to:

- **Synthesize existing evidence and provide context for this research** by establishing what is already known about the level of awareness, understanding, perceptions and attitudes to relevant concepts (including TREs/SDEs/data environments) amongst the general public and patients.

- **Inform research stimulus for co-creation sessions** by identifying available resources that explain key concepts. Examples will be shown to co-creation participants to test understanding/engagement and prompt discussion about optimal ways to explain these concepts.

- **Identify gaps to be explored in the co-creation workshops** by finding gaps in public understanding and engagement with these topics and identifying what language is used to

---

[5] UK Government Department of Health and Social Care, 2022; *Data Saves Lives: reshaping health and social care with data,* 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; Brophy, R., Bellavia, E., Groot Bluemink, M., Evans, K., Hashimi, M., Macaulay, Y., McNamara, E., Noble, A., Quattroni, P., Rudczenko, A., Morris, A. D., Smith, C. and Boyd, A., 2023, "Towards a standardised cross-sectoral data access agreement template for research: a core set of principles for data access within trusted research environments", *International Journal of Population Data Science*, 8(4). doi: 10.23889/ijpds.v8i4.2169.

[6] Jones MC, Stone T, Mason SM, et al., 2023, *Navigating data governance associated with real-world data for public benefit: an overview in the UK and future considerations*. BMJ Open, 13:e069925. doi:10.1136/ bmjopen-2022-069925

explain these concepts as well as the extent to which this includes jargon (i.e. unnecessarily complicated phrases and terms).

## Methodology

Evidence searches were conducted using the following **inclusion/exclusion criteria** to select the relevant sources:

- **Subject matter:**
  - Changes to infrastructure, specifically, SDEs, TREs, data environments, data federation/federated data platforms
  - Terms used to describe what patient data is used for i.e.
    - Direct care
    - Secondary uses:
      - Planning
      - Research
      - Clinical trials
      - Population health
      - Health surveillance
  - Terms used to describe how patient data is shared
    - Data access
    - Data dissemination
    - Data linkage
- **Date of research**: Only studies published after 2018 were included to reflect more recent findings.
- **Language of publication:** Only sources in English were included as the timescale and budget did not allow for translation.
- **Research methods**: Studies using both primary and secondary research were considered, as well as those using diverse methods, including qualitative, quantitative and mixed methods. Both academic and grey literature were considered.
- **Geographic origin of publication:** UK studies were prioritised, but studies from other countries were reviewed where relevant.

Searches were conducted using a mix of approaches which included: reviewing sources suggested by UPD and the project steering group; searching relevant websites (NHS England, the Department of Health and Social Care, and other health-related public sector and third sector websites); and conducting broader searches online and on databases.

All potentially relevant sources were assessed for quality and relevance based on agreed inclusion and exclusion criteria and quality of methodology. Once a list of relevant sources was agreed, full texts were reviewed and key findings were summarised for each of the different research questions.

This report presents key findings from the rapid evidence review, which were discussed at a workshop with the project steering group. The purpose of the workshop was to discuss and agree:

- The issues the report raises that should be the focus of the co-creation stage.
- How any existing resources could be used to inform research stimulus that would be shown to research participants in the co-creation phase.

# Public awareness, perceptions and attitudes to key health data concepts

## General awareness and attitudes regarding health data

The UK public is broadly supportive of data collection and usage where it delivers benefits for individuals and society. In particular, there is support for collecting and using data for public benefit or to create and improve products and services that benefit individuals. The main concerns about data collection and usage relate to data security (e.g. data loss, hacking), commercialisation (e.g. data sale for profit) and anonymity.[7]

Health and the economy have been identified as key sectors where data can be used for the public good.[8] Specifically, 20% of people in one survey thought that health represented the biggest opportunity to use data for public benefit, which was higher than for any other issue, including the economy.[9] Other studies have found that the public and patients are broadly supportive of using health data for a range of purposes, including: direct care, service planning and provision, as well as research and innovation.[10]

This support is not unconditional. Many members of the public are concerned that certain conditions are met for health data use. Firstly, the public's willingness to share data is higher if they can see evidence of public benefit resulting from the use of it.[11] In the health context, these benefits are often perceived transactionally, with individuals hoping to see personal benefits, for example, to gain or retain access to services or make them more convenient. Accordingly, public support for data collection and usage is strengthened if data user organisations provide evidence of how data use will benefit the public and specific communities. For example, the Office of National Statistics (ONS) found that people wanted organisations to share examples of the positive impact resulting from their use of data.[12]

Secondly, the public's willingness to share data is influenced by whether it is identifiable or not. The ONS survey suggested that whether or not individuals could be identified from data had the biggest impact on their willingness to share it.[13] Understanding that they would not be identifiable made people more comfortable with the idea of sharing their data.[14]

Thirdly, trust in individual organisations collecting, storing and using data affects how willing the public are to allow their data to be used. Typically, surveys have found higher levels of trust in NHS-

---

[7] The Centre for Data Ethics and Innovation (CDEI), 2022, *Public attitudes to data and AI: Tracker Survey Wave 2*.

[8] Ibid.

[9] Ibid.

[10] BCG, 2023, *Towards a Healthier, Wealthier UK: Unlocking the Value of Healthcare Data;* CurvedThinking, UPD and OneLondon, 2019, *Understanding Public Expectations of the Use of Health and Care Data;* Hopkins Van Mil on behalf of the National Data Guardian for Health and Social Care (NDG), UPD and UK Research and Innovation (UKRI) ScienceWise programme, 2021, *Putting Good into Practice: A public dialogue on making public benefit assessments when using health and care data;* OneLondon, Ipsos and The King's Fund, 2020, *Public Deliberation in the Use of Health and Care Data;* The Patients Association, 2023, *Developing a Data Pact: The relationship between the public, their data, and the health and care system.*

[11] This has been a consistent finding across much of the literature on health data use cited in this report.

[12] The Office of Statistics (OFS), 2023, *What we know from engaging with the public on data: a summary of people's attitudes to data*.

[13] CDEI, 2022, *Public attitudes to data and AI: Tracker Survey Wave 2.*

[14] ONS, 2023, *What we know from engaging with the public on data.*

based health data use compared to other organisations and users. For example, a 2022 survey found that 73% of UK adults thought that NHS data use would be safe, effective, transparent and responsible. By comparison, 61% trusted academic researchers and 57% trusted pharmaceutical researchers.[15] Trust in other types of organisations accessing health data can be even lower. A 2023 YouGov survey found that the public were less willing to share their data on the NHS Federated Data Platform if it was run by a commercial company rather than by the NHS.[16] A London-based consultation about health data use found lower support for commercial companies being able to access health data. The idea of access being given to insurance companies was rejected.[17]

In order to support health data use, there are conditions the public, generally, want to be met. These include:[18]

- Having a clear medical purpose for health data usage;
- Restricting data access to approved purposes;
- Ensuring data quality and accuracy;
- Imposing legal penalties for misuse of data;
- Adhering to agreed frameworks for safe data access and use, and reviewing these frameworks periodically;
- Being transparent about how and why health data is used and protected;
- Sharing information about the outcomes of health data use;
- Sharing any commercial benefits and medical advances with the NHS and across the NHS to avoid health inequalities.

Despite support for health data use under the conditions cited above, there is still limited public and patient understanding and awareness of health data.[19] For example, a recent study that involved 112 members of the general public and patients found that a quarter had never thought about health data use beyond their direct care, whilst over a half thought only about this occasionally.[20] In this context, it is not surprising that most research with the public on this topic is of deliberative nature, where information about key health data concepts is shared with research participants so they can arrive at informed opinions via discussion.

Part of the challenge in raising public awareness and understanding of health data use is the technical nature of many key concepts and the inconsistent use of language to describe them. For example, one study stresses that the language used to explain uses of patient data in care, treatment and research is difficult, complex and confusing, making it difficult to build understanding and trust.[21] Others have highlighted similar issues with data terminology. ONS notes that “terms such as ‘aggregated’, ‘anonymised’ and ‘de-identified’ can be confusing and often raise further questions.”[22] Another study found that 42% of the public could not identify the meaning of the term ‘pseudonymised data’ correctly.[23] However, in the same study, 63% of nationally representative

---

[15] CDEI, 2022, *Public attitudes to data and AI: Tracker Survey Wave 2.*

[16] YouGov NHS FDP Poll Results, 2023.

[17] OneLondon, Ipsos and The King’s Fund, 2020, *Public Deliberation in the Use of Health and Care Data;*

[18] Global Counsel, 2023, *Polling Digest: Overview of UK Public Opinion, April 2023;* OneLondon, Ipsos and The King’s Fund, 2020, *Public Deliberation in the Use of Health and Care Data*

[19] Public Experience Library, 2021, *Public Perceptions of NHS Data Use.*

[20] Hopkins Van Mil on behalf of NDG, UPD and UKRI ScienceWise programme, 2021, *Putting Good into Practice.*

[21] Ibid.

[22] Ibid.; HDRA & NHSX, 2021, *Building Trusted Research Environments – Principles and Best Practices: Towards TRE ecosystems (1.0). Zenodo.* https://doi.org/10.5281/zenodo.5767586

[23] NHS Digital, 2022, *GPDPR Research Report* produced by Research Works Limited.

sample correctly identified the meaning of 'de-identified' data, suggesting that understanding of at least some terms may be increasing.

Aside from not understanding data terminology, researchers have also found that the public may have a better understanding of certain ways in which health data is used than others.[24] For example, a qualitative study about public views about health data found that people generally understood how patient data is used for direct care. There was less spontaneous awareness and understanding of secondary uses of health data. However, with the help of specific examples, participants in that study could understand the benefits of using health data for service planning or population health. More detail about public views on particular uses of health data is provided below.

## Direct care

Research suggests that the public supports the use of health data as part of direct patient care.[25] Patients can see the benefits of different health professionals being able to access their health records across the NHS and many assume this is already in place. Key benefits for patients include: not having to repeat the same information to different professionals; speedier access to information in emergencies; better understanding of the individual's health situation; and avoidance of unnecessary repeat tests.

Discussions about using health data for direct care on behalf of The Patients Association revealed a demand for patients to have access to their own health records and some control over what is shared and how. [26] This research also identified patient interest in knowing what information is shared about themselves. Concerns were also expressed about data quality and whether inaccuracies could lead to the wrong decisions and actions being taken. In this research, certain kinds of health data were also perceived to be more sensitive and were therefore perceived to require greater caution in terms of data access and use. This included data concerning sexual health, mental health and substance misuse.

The findings above are relevant to the present discussions of the NHS England Federated Data Platform (FDP) as they show support for enabling access to health records across NHS settings (which would be enabled through the platform). However, no research directly exploring public views about the NHS England FDP has been identified by this rapid evidence review. This is understandable as the contract for FDP software has only recently been awarded.[27] Discussions about this contract, and the involvement of commercial companies in running the NHS England FDP more generally, have highlighted public concerns about patient data privacy and commercial interests. Specifically, patients have been concerned about whether commercial companies running the NHS England FDP would have access to their data and how this would be prevented. UPD has suggested that more and clearer information is needed about how patient data privacy will be maintained in order for the public to feel confident about how their data will be used via the FDP.[28]

---

[24] The Patients Association, 2023, *New report on the relationship between the public, their data and the health and care system*. See: https://www.patients-association.org.uk/blog/new-report-on-the-relationship-between-the-public-their-data-and-the-health-and-care-system

[25] CurvedThinking, UPD and OneLondon, 2019, *Understanding Public Expectations;* The Patients Association, 2023, *Developing a Data Pact.*

[26] The Patients Association, 2023, *Developing a Data Pact.*

[27] NHS England, December 2023, *Federated data platform update*.

[28] UPD, November 2023, *Making better use of NHS data: Where we're at with the Federated Data Platform.*

## Secondary uses

Most commonly, the secondary use of health data has been discussed with regards to research and less so in relation to service planning and improvement.[29] Other secondary uses, for example, population health management or health surveillance, have rarely been discussed explicitly or as separate topics with the public and patients. Research has identified very low understanding of why and how non-NHS organisations and individuals would access and use health data.[30]

However, once specific examples of different secondary uses of health data have been shared with the public and patients through public and patient engagement, consultations or research, there has been broad support for using data in these ways. Clarity about specific secondary data uses and tangible examples have therefore been highlighted as critical to public understanding of, and support for, secondary uses of health data. As before, clear public interest and patient anonymity further consolidates support.

- **Service planning and improvement:** Recent research commissioned by NHS Digital (now NHS England) found that around two thirds of the public agreed with the use of health data for planning NHS services, treatment and prevention.[31] Other research found that support was higher when people could see the benefits to themselves, for example, improved local health service provision. Conversely, support was lower when data was used to inform national planning and improvement because the public felt less sure whether their local area would benefit from it.[32] As previously mentioned, using tangible examples of how health data is used to improve service planning will be important for building public understanding and support for this type of health data use.

- **Research and innovation:** A narrative review of public and patient views on health data sharing for research purposes found that public support for this use of health data was high.[33] The reasons for accepting this use of health data were: wanting to contribute to future healthcare advancements; 'giving back' (in cases where patients have benefitted from healthcare services); and hoping to benefit from improved future patient care.

  Other research supports these findings. For example, previous UPD research found that public interest in health research and belief in its importance is high.[34] However, as with health data use in general, public awareness and understanding of how data is used for health research, its complexity and the different stakeholders involved, was found to be limited. As before, public benefit and patient anonymity increased public support for this type of health data use. Conversely, commercial access to health data for the purposes of health research tended to raise questions about potential public benefits and, as a result, decreased support.[35]

- **Population health:** As with service planning, research has found little spontaneous awareness of this kind of health data use but sharing examples of specific data uses has helped to create

---

[29] CurvedThinking, UPD and OneLondon, 2019, *Understanding Public Expectations.*

[30] Ibid.

[31] NHS Digital, 2022, *GPDPR Research Report* produced by Research Works Limited.

[32] CurvedThinking, UPD and OneLondon, 2019, *Understanding Public Expectations;*

[33] Kalkman S, van Delden J, Banerjee A, et al., 2022, "Patients' and public views and attitudes towards the sharing of health data for research: a narrative review of the empirical evidence. *Journal of Medical Ethics*, 48:3–13. See: https://jme.bmj.com/content/48/1/3

[34] CurvedThinking, UPD and OneLondon, 2019, *Understanding Public Expectations;*

[35] Ibid.

understanding and support. For example, the Patients Association research found that the public supported GPs using health data to target patients at risk of certain conditions. Public engagement was seen as critical for raising awareness and support for this type of data use, including the co-production of information and hyper-local focus of communications.[36]

- **Health surveillance:** Research about public views on health data use for health surveillance has been identified as a gap in evidence by this rapid evidence review. Some reports that were reviewed identified public support for using health data for managing the Covid-19 pandemic, as well as similar future emergencies.[37] However, any detailed exploration of public views about this type of health data use was lacking.

## Public awareness of, and attitudes to, TREs / SDEs / data environments

Early consultations about TREs suggest that the public and patients are broadly supportive of this direction of travel.[38] TREs have been perceived to address some key concerns about health data security via their controls and governance structures. This, and other consultations and research, point to some key aspects of TREs that have helped to secure public confidence.

Firstly, the fact that data will not be leaving a controlled and secure data environment has helped to reassure the public about data security and confidentiality. For example, a recent survey found that 86% of respondents felt more comfortable with data access than with a data sharing / data dissemination approach.[39] Secondly, controls over access and vetting of projects further reassured the public that there are mechanisms in place to ensure that data is used for public benefit. Thirdly, monitoring of data use by host organisations, contractual agreements signed as part of the data access process, and consequences for data misuse and breaches helped increase public confidence in TREs.[40]

At the same time, there are some areas where the public and patients feel more work is needed to shore up public trust in health data access and usage via TREs. For example, one consultation highlighted the need for improved public-facing communications, so that the design and functioning of TREs is better explained in lay terms. In addition, this consultation stressed the importance of involving public and patient representatives in the data access management and decision-making process to ensure transparency of data use, outcomes and impact.[41]

Beyond general statements of public approval of TREs, there has been limited research and evidence gathered about how the public and patients feel about specific aspects of TREs. An exception to this is the work conducted to develop standardised data access agreements (DAAs) and related guidance for TREs, which has involved public and patient representatives as part of the TRE Legal Toolkit Action

---

[36] The Patients Association, 2023, *Developing a Data Pact.*
[37] BCG, 2023, *Towards a Healthier, Wealthier UK.*
[38] HDRA, 2020, *Trusted Research Environments (TRE): A strategy to build public trust and meet changing health data needs.*
[39] BCG, 2023, *Towards a Healthier, Wealthier UK.*
[40] HDRA, 2020, *Trusted Research Environments (TRE): A strategy to build public trust and meet changing health data needs.*
[41] Ibid.

Force.[42] A paper reporting the results from this work has discussed public and patient responses to different aspects of managing data access in TREs, providing some more specific findings about public attitudes to particular features of TREs.

Overall, the members of the public and patients consulted approved of the initiative to develop standardised DAAs and some thought this could promote efficiency and consistency in how data access was managed in different TREs. However, respondents raised issues with some aspects of the proposals for standardising DAAs. Specifically, some stressed the importance of ensuring that individual researchers accessing TREs were aware of their responsibilities and promoted good practice. To help ensure this, respondents supported the idea of researcher training for anyone granted access to TREs.

Other research has identified similar concerns about researcher compliance. For example, a consultation by DARE UK (Data and Analytics Research Environments UK) found that the public did not take individual researcher compliance for granted. Reflecting on these concerns, participants felt that the safest way of accessing data would be from a safe room in the TRE where the data was held. Some participants accepted other ways of accessing data, for example from a safe room at an approved university or from the researcher's laptop, provided relevant governance and safety processes were in place and adhered to. Some still remained concerned about the potential for data breaches through individual researcher non-compliance.[43]

To date, public and patient views about the commercial use of health TREs have been mixed. The DAA study showed some public unease about commercial access, particularly with regard to issues about intellectual property and future commercialisation of results derived from public health data. At the same time, some respondents accepted commercial access to health data for projects that would deliver public benefit. To ensure public benefit, respondents wanted public and patient representatives to be included in decision-making about data access. Better explanation of commercial data use was recommended to provide transparency and support public understanding. This study also found that the public were more likely to support commercial involvement if some of the value generated was re-invested into the health system.[44]

Similar concerns about third-party access to data have been identified in other research, suggesting that the public are worried that NHS data may be exploited. A study about public views on this matter on behalf of UPD suggested that NHS bodies would need guidance to ensure public benefit is prioritised over commercial interests. It was envisaged that such guidance would provide a set of rules for managing third party access to NHS health data.[45] Other recommendations to help the NHS manage third party data access included: introducing a governance system to monitor such partnerships; and ensuring that partnerships were driven by public benefit, based on a set of shared principles, as well as being transparent and accountable.[46]

Another issue patients and public representatives raised in this study concerned the perceived need for a standardised response to breaches of DAAs and non-compliance. Handling breaches and

---

[42] Brophy, R., Bellavia, E., Groot Bluemink, M., Evans, K., Hashimi, M., Macaulay, Y., McNamara, E., Noble, A., Quattroni, P., Rudczenko, A., Morris, A. D., Smith, C. and Boyd, A., 2023, "Towards a standardised cross-sectoral data access agreement template for research.

[43] DARE UK, 2022, *Building a trustworthy national data research infrastructure: A UK-wide public dialogue.*

[44] BCG, 2023, *Towards a Healthier, Wealthier UK.*

[45] Understanding Patient Data (UPD), 2020, *Foundations of Fairness: Where next for NHS health data partnerships?*

[46] UPD, 2020, *Foundations of Fairness: Views on uses of NHS patient data and NHS operational data.*

ensuring accountability was seen as critical for earning public trust. For this reason, some thought that there should be a template for the monitoring of TRE use. The need for reassurance about monitoring was also evident in concerns expressed by some, who questioned whether it would be possible to monitor a great number of individual researchers accessing the same TRE.

In addition, respondents from this study argued that any future changes to standards in DAAs and data access processes should be only made after consulting with the public and patient representatives. They also suggested that all DAA-related documents should be produced in accessible, easy read versions to increase transparency and confidence in data access processes.

## How are key health data concepts explained to the public

As evident from the studies cited here, public and patient calls for improved lay information about TREs have been a running thread through consultations on this topic. The need for better communications has also been a staple part of all recommendations for future work on TREs. For example, the UK Health Data Research Alliance (HDR UK) has highlighted improved lay explanations about the design and functioning of TREs as one of the key areas where further work is needed in order to develop a productive TRE ecosystem.[47] Much of the discussion about public-facing information and communications about TREs has been concerned with *what, when and how TREs should be explained and communicated*, rather than exploring public and patient feedback about existing communications. The section below outlines some key themes central to discussions about public-facing communications about TREs, followed by an overview of communication formats that have been used to date.

*What is in the name?*

At present, there is an array of different names and terms used to refer to TREs. Even a cursory look at the names of some existing TREs demonstrates this. For example, UK TREs include Scotland's Data Safe Haven programme (now under Research Data Scotland), UK Secure eResearch Platform in Wales, Genomics England's Research Environment, UK Data Service Secure Lab, and NHS Digital TRE for England (now the NHS England Secure Data Environment and sub-national Secure Data Environments).[48] The names of TREs vary in whether they refer to 'trusted' or 'secure' environments and whether they are named as 'research' or 'data' environments. Another variation is using the term 'data haven' rather than TRE or SDE.

This variety of names has been considered confusing for the public and has given rise to debates about the best terms to use. Medical ethicists and others concerned with TREs have considered the potential implications of different terms for public understanding and expectations of TREs. Some have argued that TREs remove the need for 'trust', as governance structures and security mechanisms prevent inappropriate uses and data breaches.[49] In their view, focusing on trust could have unwanted implications. For example, it has been suggested that the public may feel vulnerable about giving access to their health data on the basis of trust or may not expect TREs to give access to

---

[47] HDRA, 2020, *Trusted Research Environments (TRE): A strategy to build public trust and meet changing health data needs.*

[48] HDRA & NHSX, 2021, *Building Trusted Research Environments – Principles and Best Practices: Towards TRE ecosystems (1.0). Zenodo.* https://doi.org/10.5281/zenodo.5767586

[49] Graham M, Milne R, Fitzsimmons P, et al., 2023, "Trust and the Goldacre Review: why trusted research environments are not about trust", *Journal of Medical Ethics*, 49:670–673;

commercial companies. Instead, it has been suggested that the term 'Secure Research Environment' may be more appropriate.

Others have, however, pointed out that governance structures and security mechanisms do not fully remove the need for trust. From this point of view, data security measures still leave an open question about trust in both organisations holding and organisations using health data.[50] As an example, the authors cited recent concerns about a US technology corporation's involvement with the NHS England Federated Data Platform, showing that trust in organisations involved in TREs mattered greatly.[51]

The potential implications of other terms have also been debated. Medical ethicists and others concerned with TREs have suggested that a focus on the word 'research' may imply that TREs are used for research only, whilst they may also be used for other purposes (service planning, auditing etc). In addition, they have suggested that a focus on the word 'security' may imply that data security and privacy risks have been fully resolved, which they did not think was entirely the case. In their view, the use of the term 'data haven', also raised other potential issues, as they felt that the term could have unhelpful connotations and be associated with tax havens. Given these issues, they argued for a descriptive term to be used, suggesting 'Controlled Access Data Environments' (CADE).

The discussion above underlines a lack of agreement at the most basic level of communication i.e. what such systems should be called. The debate is also striking in its focus on the speculative implications of different terms, rather than being based on research evidence about how these different terms are understood and perceived by the public and patients.

This report is not suggesting that any organisations should change the name of their environment, as we understand that such changes can be costly and confusing in themselves, but it is worth noting the impact of different naming conventions.

### *The Five Safes framework*

Discussions about how TREs should be explained to the public have consistently called for the explanation to be based on the Five Safes framework.[52] This framework was originally created by ONS in 2002 to describe data access in a secure environment. It has since become widely used as a more general framework for designing and describing the characteristics of safe and efficient data access systems. The example overview below outlines key questions for each of the Five Safes that secure data systems need to resolve satisfactorily:

**"The Fives Safes are:**

- Safe People — are the researchers using the data appropriately trained and aware of their role in data protection?
- Safe Projects — does the project make good use of the data? Is it lawful and, particularly for NHS data, in the public interest?
- Safe Data — what is the potential for individuals to be identified in the data?
- Safe Settings — are there technical controls on access to the data?

---

[50] Affleck P, Westaway J, Smith M *et al*, 2023, "Trusted research environments are definitely about trust", *Journal of Medical Ethics,* 49:656-657.

[51] Paul Affleck, Jenny Westaway, Maurice Smith and Geoff Schrecker, 2021, Trusted Research Environment – a name to trust?, *BMJ Blogs*.

[52] HDRA & NHSX, 2021, *Building Trusted Research Environments – Principles and Best Practices: Towards TRE ecosystems (1.0). Zenodo.* https://doi.org/10.5281/zenodo.5767586

- Safe Outputs — is there any residual risk in outputs being released from the secure environment?" [OpenSafely, 2023]

More recently, the Safe Settings aspect of the framework has acquired another dimension. Increasingly, computing settings used for safe data access systems are not hosted internally by individual organisations but on cloud platforms operated by contracted commercial providers. This has created additional security requirements to ensure that third party commercial providers cannot access and use the data which is being hosted on their platforms.[53]

The Five Safes framework has been broadly accepted as the overarching conceptual basis for educating the public about TREs. It has, therefore, been used in public-facing communications to explain how data is kept safe and secure in TREs, for example, in HDR UK's one-page introduction to TREs:

**What is a TRE?**

A TRE is a **Trusted Research Environment.** Also known as 'Data Safe Havens', TREs are highly secure computing environments that provide remote access to health data for approved researchers to use in research that can save and improve lives.

**Why are they important?**

**TREs make research safer.** Making data available through a TRE means that people can be **confident** that their personal health data is accessed **securely** and their **privacy protected.**

TREs help make **research efficient**, **collaborative** and **cost effective**, providing rich data that enables **deep insights** which will go on to improve healthcare and **save lives.**

TREs provide approved researchers with a **single location** to access valuable datasets. The data and analytical tools are all in **one place**, a bit like a **secure reference library.**

**Learn more about TREs and discover examples of how TREs are being used to enable life-saving health research.**

**How is my data safeguarded?**

Health data should always be kept safe and secure, and used responsibly to ensure privacy. Heath Data Research UK ensures these high standards are met by promoting the use of the 'Five Safes' model across all TREs.

**Safe People**
Only trained and specifically accredited researchers can access the data

**Safe Projects**
Data is only used for ethical, approved research with the potential for clear public benefit

**Safe Settings**
Access to data is only possible using secure technology systems – the data never leaves the TRE

**Safe Data**
Researchers only use data that have been de-identifed to protect privacy

**Safe Outputs**
All research outputs are checked to ensure they cannot be used to identify subjects

*<u>What to say about TREs and how to say it</u>*

Recommendations for explaining TREs to the public and patients have highlighted two overarching goals for such communications:[54]

- **Highlighting benefits of data use** – this has been recognised as a key component of communication about TREs, since ensuring that data is used for public benefit has been identified as critical to building public trust and confidence in health data access systems.
- **Addressing concerns about data security and confidentiality** – showing how security is managed in TREs to ensure that data is safe and individuals cannot be identified has been identified as another key aim for public-facing communications.

---

[53] HDRA, 2020, *Trusted Research Environments (TRE): A strategy to build public trust and meet changing health data needs.*

[54] HDRA & NHSX, 2021, *Building Trusted Research Environments – Principles and Best Practices: Towards TRE ecosystems (1.0). Zenodo.* https://doi.org/10.5281/zenodo.5767586

In addition to these overarching objectives, a range of very specific information needs have been identified. The list below may not be exhaustive but shows the range of different topics that have been identified as important to communicate to the public and patients:[55]

- Keeping the public up to date regarding the present TRE situation and proposed developments;
- Explaining (in lay terms) processes for data handling within TREs, for example, how data linkage is managed to maintain privacy;
- Providing lay summaries of research projects which have been granted access, with clear explanations of public benefit;[56]
- Considering how beneficial health data use is to different groups and who benefits from it;[57]
- Ensuring that there are mechanisms to enable researchers to report back the results of their projects to participants;
- Explaining technical controls that protect data held in public cloud so that it cannot be accessed and used by the hosting organisation, i.e. commercial providers of cloud platforms;
- Providing full technical documentation that is transparent about security design and implementation, including independent assessment reports and allowing for review by technical press and experts.

The examples of specific information needs above also highlight the importance of thinking about communications in terms of different points in the data journey. For example, while lay summaries of projects and their public benefit may be relevant at the project and researcher vetting stage, feeding back research results to participants comes at the end of the process and can be potentially used to highlight the benefits of TREs and health data usage. Other topics may be less attached to specific stages in the data journey and more generally impact on public willingness to share their health data, for example, explanations of measures used to protect data.

Certain principles for communicating information about TREs have been highlighted as potentially effective:[58]

- Ensuring that the language used is consistent and clear;
- Not assuming any prior technical knowledge;
- Using visual representations to show data processes;
- Demonstrating rather than explaining physical and logical controls by taking the public through the processes of data collection, access, handling and use.

Particular channels for communicating information about TREs to the public and patients have also been suggested and discussed by HDR UK as critical to building and earning public trust:[59]

- **Public and patient involvement** (PPIE) has been advised at various stages of health data use, from co-design of governance structures and processes, through involvement in decision-making around the approval of research projects, to ongoing communication of research results. HDR UK's strategy document provides examples of PPIE activity, including participant-led forums

---

[55] Ibid.
[56] HDR UK, 2020, *Trusted Research Environments (TRE): A strategy to build public trust and meet changing health data needs.*
[57] DARE UK, 2022, *Building a trustworthy national data research infrastructure: A UK-wide public dialogue.*
[58] HDR UK, 2020, *Trusted Research Environments (TRE): A strategy to build public trust and meet changing health data needs.*
[59] Ibid.

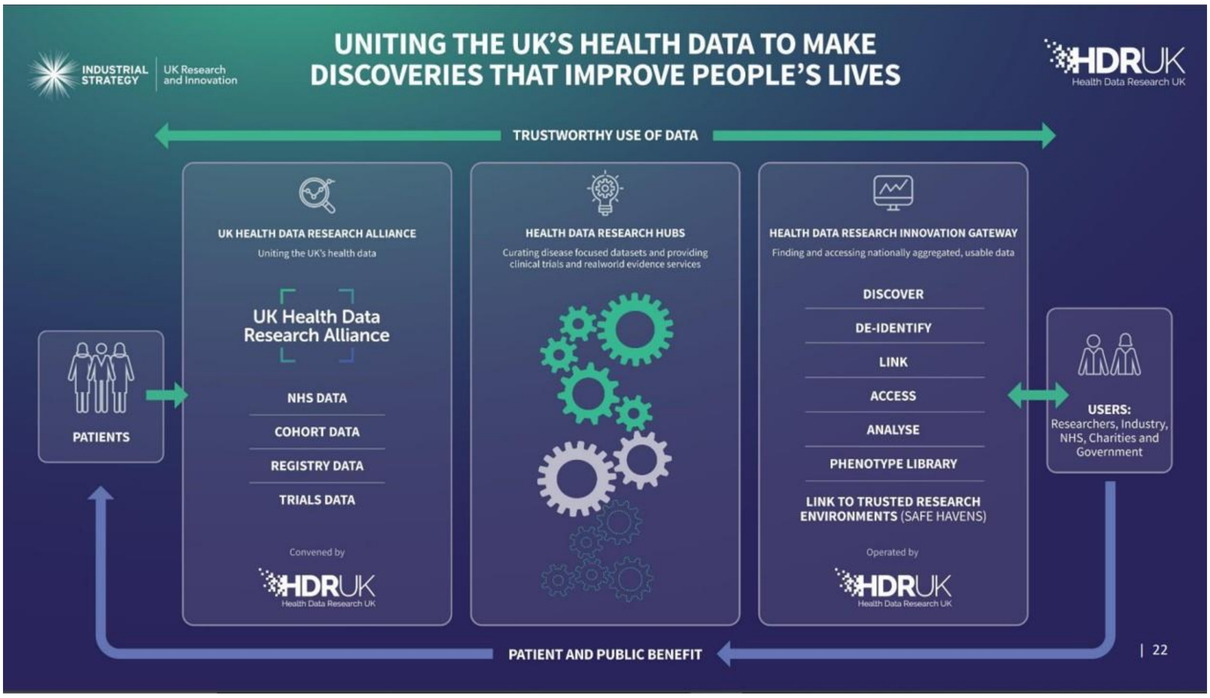
(e.g. Innovative Medicines Initiative programme), engagement activities with different disease-specific patient communities in local areas, citizen juries (e.g. Connected Health Cities).

- **Routine lay information** has been identified as important for communicating information about different aspects of TRE processes, for example, providing lay summaries of approved projects and reporting research results.
- **Stakeholder communications** have also been identified as another important channel for increasing awareness and understanding of TREs, their design and functioning, and their impact in terms of public benefit. Key stakeholders listed in the strategy document include: HDR UK, Use My Data, Understanding Patient Data, National Institute for Health and Care Research, Association of Medical Research Charities and individual research charities, and existing TRE operators.
- **Healthcare settings** have been suggested as another opportunity to showcase the public benefits of research made possible by TREs, for example, through posters in hospital wards and GP practices.
- **Media and opinion formers** have also been suggested as a potential information channel, for example, disseminating information about the benefits of research on TREs through the NHS website and mainstream media channels, as well as by champions, journalists, and vloggers.

TRE information formats

The rapid evidence review process identified a limited number of resources that have aimed to explain TREs in lay terms, rather than addressing professionals and researchers involved with TREs. Some of these materials were designed to be used in public and patient consultation activities, so were not developed for wider use. Nevertheless, they are still instructive in showing a range of ways to explain TREs to the public. Examples of these different formats are shown below:

- **Infographic lay overviews:** The HDR UK one-page infographic overview explains how TREs work and how data is protected through the Five Safes framework. Another infographic has been produced to outline the flow of data from patients to researchers and the principles of public benefit underpinning health data use.



UNITING THE UK'S HEALTH DATA TO MAKE DISCOVERIES THAT IMPROVE PEOPLE'S LIVES

TRUSTWORTHY USE OF DATA

PATIENTS

UK HEALTH DATA RESEARCH ALLIANCE
Uniting the UK's health data

UK Health Data Research Alliance

NHS DATA
COHORT DATA
REGISTRY DATA
TRIALS DATA

Convened by HDRUK Health Data Research UK

HEALTH DATA RESEARCH HUBS
Curating disease focused datasets and providing clinical trials and realworld evidence services

HEALTH DATA RESEARCH INNOVATION GATEWAY
Finding and accessing nationally aggregated, usable data

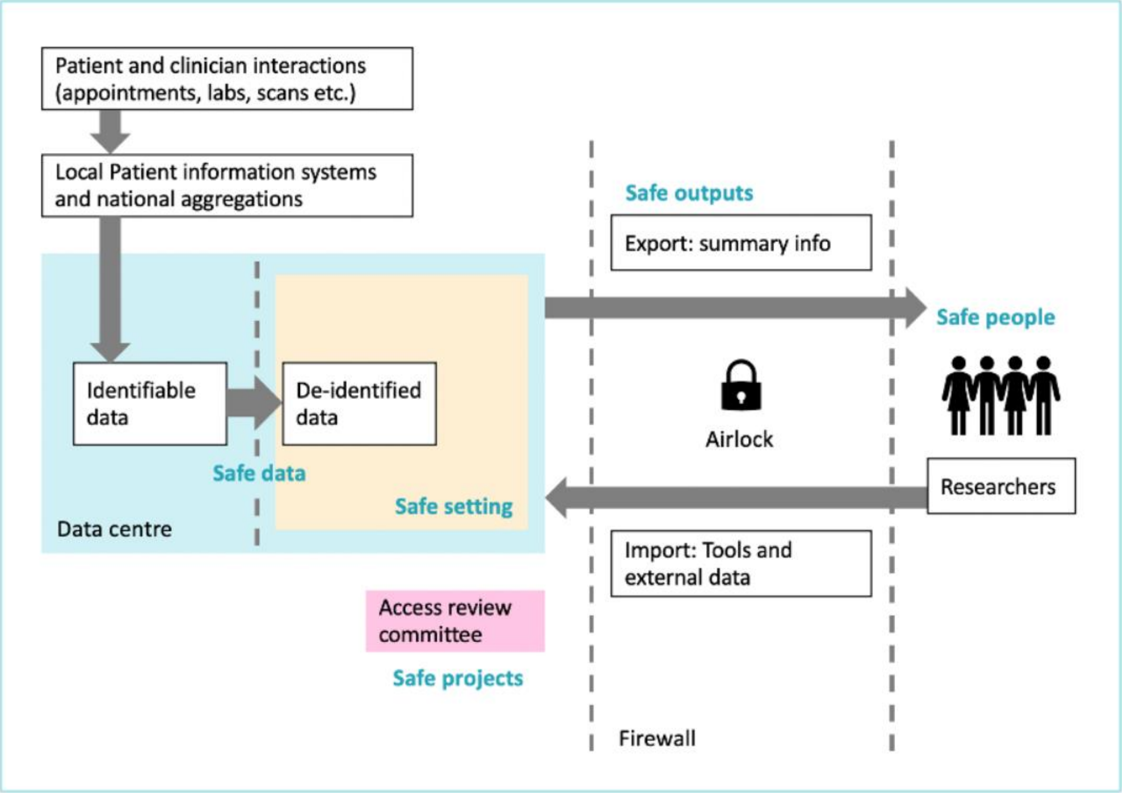
DISCOVER
DE-IDENTIFY
LINK
ACCESS
ANALYSE
PHENOTYPE LIBRARY
**LINK TO TRUSTED RESEARCH ENVIRONMENTS** (SAFE HAVENS)

Operated by HDRUK Health Data Research UK

**USERS:** Researchers, Industry, NHS, Charities and Government

PATIENT AND PUBLIC BENEFIT

- **Diagrams:** Visual representations have been used to show data flows from data collection, through TREs where data is held and protected, to data usage and analysis.



- **Data journey examples:** Another format used to explain the flow of health data, the protections in place, its uses and benefits, has involved data journey examples, for example, the data journey in a project exploring the relationship between diabetes and education results.

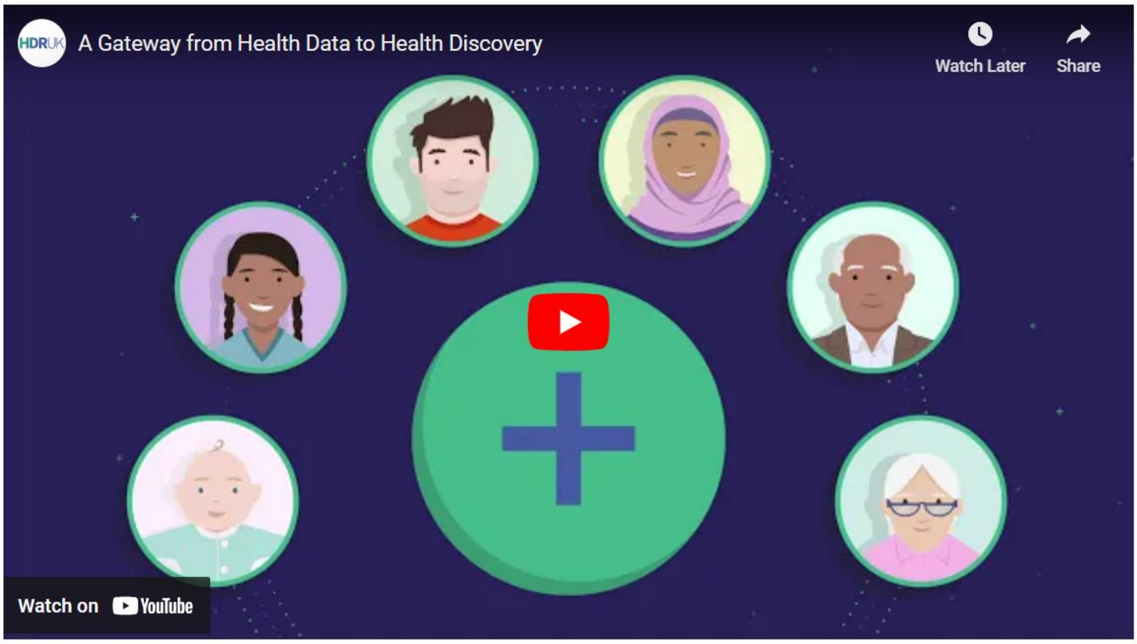
The data journey, from collection to use in research

Diabetes and education data research

*DARE UK Public Dialogue workshops, January 2022*

I'm now going to show you an example of what the data journey looks like from its original collection to it's use in research. This is a hypothetical example, but the journey reflects that taken for real life projects.

- **Videos:** ONS have produced a video to explain the Five Safes framework and the design and functioning of TREs. However, the content is not specific to health data and discusses both data sharing and TRE models of data access and use. Another video has been produced by HDR UK to explain the current developments in uniting UK health data through the Innovation Gateway, which would provide a single point for accessing health data held by different TREs.



- **Case studies:** There are multiple case studies about health data and research stakeholder websites which introduce projects using health data held in TREs. However, most use technical language so are targeting professional audiences rather than the public.

The examples above reflect some of the recommended principles for communicating about TREs, for example, to use visual representations and demonstrate data flows, protections and uses. However, there may be instances where some of the principles are more challenging to follow. For example, some of the case studies reviewed and also the diagrams showing data flows may require prior knowledge, so therefore may not meet the requirement not to assume any prior knowledge. More broadly, the lack of consistency in the names of these data environments also fails to meet the need for consistent language that would aid public understanding. Finally, some of the specific information needs identified in this review, for example, around measures to prevent identification through data linkage or technical controls on cloud platforms, have not been addressed in the TRE-related public communications identified in this research.

## Conclusions and next steps

The rapid evidence review has found that communications about TREs are recognised to be critical to public understanding and trust in health data use. One striking feature of many discussions about how to explain TREs to the public has been the focus on recommendations for effective communication rather than exploring public and patient response to existing materials to help improve them. Admittedly, some of the recommendations arose from work that involved consultations with the public and patients, but without explicitly explaining how any recommended principles were based on public feedback. A case in point is the debate about naming these data environments which reflected on the potential implications of different names without testing them with the public.

➔ **This research will proceed to address this gap by exploring public and patient understanding and responses to different names, as well as existing communication materials used to explain TREs.**

The literature review suggests that there are effective means to explain the basic design and functioning of TREs and different health data uses. Specifically, the Five Safes framework provides a sound conceptual basis for explaining the factors that make TREs safe. Other formats that can help further explain TREs include diagrams of data flows and data journeys. Case studies can be helpful to promote benefits of health data use. Existing evidence also suggests that the public and patients can understand different uses of health data with the help of examples and case studies.

While work on producing effective communications about TREs is underway, there are certain areas where information needs may not be met or may be more difficult to meet. Prior research suggests that the public want their concerns to be addressed and benefits to be explained. However, explaining some of these may concern very technical processes and terminology, which is a major gap in current communications. Examples of technical or specialist topics and terms that may be challenging to explain and which have not been sufficiently addressed include: measures to prevent identification through data linkage, controls on public clouds so commercial providers cannot access the data, managing commercial uses of health data, or using health data for health surveillance. Also, many case studies explaining the outcomes and benefits tend to be technical and may need further work to adapt them for the general public and patient audiences.

➔ **This research will proceed to test existing resources to capture public understanding of, and views about TREs, explained with reference to the Five Safes framework, to expand the evidence base around using this framework and inform future materials.**

- ➔ **This research will also explore the examples required to explain different health data uses to the public and patients.**

- ➔ **It will also capture key questions that the public and patients have about various issues and information needs highlighted in the rapid evidence review, as well as gauge if there are further information needs that have not been previously identified.**

- ➔ **The research will also gather evidence about the public's understanding of different terms and their response to different words and phrases, to identify the optimal words and phrases to use when explaining different aspects of TREs and different health data uses.**

- ➔ **The different names used for TREs, and examples of different communication formats that have been used to explain TREs, will be included in research stimulus and shared with respondents in the co-creation stage.**

## Acknowledgements

This project is supported by the Wellcome Trust [227162/Z/23/Z]; Medical Research Council [MC_PC_23017]; the National Institute for Health and Care Research; and NHS England.

It is also supported by a project steering group, consisting of individuals from the following organisations:

- Wellcome
- Healthwatch
- NHS Confederation
- NHS England
- use MY data
- Department of Health and Social Care
- Health Research Authority
- Health Data Research UK