

# What words to use when talking about health data: Secure Data Environments and Trusted Research Environments

## Final report

**Prepared for: Understanding Patient Data.**

**Written by: Research Works Ltd.**

### 1. Executive summary

---

Secure Data Environments (SDEs) and Trusted Research Environments (TREs) are well established methods of storing and accessing data in the UK, and, in particular, health data. Beyond general statements of public support for SDEs and TREs, there has been limited research and evidence on how the public and patients feel about specific aspects of them.<sup>1</sup> There is also a lack of independent, non-organisation-specific resources to describe these environments.

This research project was commissioned by Understanding Patient Data (UPD) to explore how the public feel about these environments and how best to communicate them to the public. In total 16 workshops were conducted with 50 members of the public from across the UK, alongside 8 interviews with experts in the use of health data. This findings document is designed to support the development of an umbrella resource describing SDEs and TREs as well as supporting other organisations when creating their own resources.

This research identified four core challenges in communicating SDEs and TREs to the public that should be considered.

- The first, is that an umbrella resource created to describe *all* SDEs and TREs will need to be mindful of the breadth of technical differences that may exist between individual SDEs and TREs, particularly as the definitions of SDEs and TREs themselves aren't universally agreed.
- This is an evolving and changing landscape and resources will likely need to evolve with it.
- Providing a balance between transparency and clarity for the public was a real challenge; several topics introduced that provided full transparency led to confusion amongst public participants.
- Public participants sought a reassuring tone, however some experts cautioned that language choices, especially for UPD's resources, should be neutral, and ensure any resource leaves the public to make an informed choice rather than seeking to influence.

Transparent and accurate reassurance, therefore, might be a topic for other organisations involved in the SDE and TRE ecosystem to consider with their resources.

There were six core topics that public participants wanted included in any resource describing SDEs or TREs. These were:

- The context of why large datasets are used for purposes other than direct care.
- The fact that data is de-identified prior to access.
- The fact that data in SDEs and TREs is accessed not shared.
- A transparent description of who is accessing the data.
- How the data is kept secure.
- The benefits of using SDEs and TREs to access large datasets.

Some of these broader topics are covered in existing UPD resources. Experts agreed with the topics chosen by public participants; however, some believed the resource should also mention data

---

<sup>1</sup> [What Words To Use - Rapid Review.pdf \(understandingpatientdata.org.uk\)](https://www.understandingpatientdata.org.uk/what-words-to-use-rapid-review.pdf)

outside of health and care, data linkage and the benefits of this, and how data is checked prior to release. These may be topics for a further stage of resource creation.

The level of interest in each of these six core topics differed based on demographic group and the level of concern participants held generally about the use of their patient data. The research consulted two groups of participants who were more likely to opt out of the use of their health data; Black ethnic minorities, and those aged 18-35 who are classified as A in terms of their socio-economic status. These groups were more interested in who is accessing the data and how it is kept secure. These differences, and others highlighted in the main report, could be considered by those developing SDE and TRE resources for the public to decide if tailoring information to different groups could be helpful.

Both public participants and experts agreed that an SDE/ TRE resource should contain layers of detail to meet the different information needs of the public. A resource should have a 1-line summary, the main resource containing the core messages above and links to further detail for those who seek it.

Public participants preferred an animated video with a voiceover as the best format for an umbrella resource describing SDEs and TREs. Experts believed an umbrella resource developed by UPD could be used by individual SDEs and TREs as a core resource supplemented by specific details relevant to the individual SDE or TRE.

## Report contents.

---

2.	Project context, objectives, and project methodology	4
3.	Considerations for an SDE/TRE resource	7
4.	The core topics the resource should cover and how to convey these to the public	11
	Introducing the use of health data	11
	The data flow and de-identification of the data	12
	Data access rather than data sharing	15
	Who uses the information and how they are approved	17
	The benefits of using SDEs and TREs to access large health data sets	19
	Security of the data	21
5.	The name Secure Data Environment and Trusted Research Environment and the use of both	24
6.	Providing layers of information – extra information for those who may wish to seek it	26
7.	The format of a resource describing SDEs and TREs and how it could be used	28
8.	Conclusions and recommendations	30
9.	Acknowledgements	32

## 2. Project context, objectives, and project methodology

Whilst many organisations that have Trusted Research Environments (TREs), Secure Data Environments (SDEs), and other similar environments have public-facing information about their product and/or service, 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 which risks inefficiencies if multiple resources are created to describe the same information each in a slightly different way.<sup>2</sup> There is also limited evidence on how people feel about these environments in general, and what they would like to know about them.

In this context, Research Works have undertaken research with the following objectives:

- Gain insight into members of the public's understanding of SDEs and TREs, 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 and their perspective on how they would best like them to be communicated.
- Co-develop specifications for public-facing resources based on the above, with the type of resource being recommended by the workshop participants.

The first stage of this research involved a rapid evidence review to synthesize existing evidence and provide context for the research, inform the stimulus materials to be used in the research and to identify gaps in knowledge to be explored with members of the public. This rapid evidence review has been published.<sup>3</sup>

This report outlines the findings from the qualitative research comprising 16 workshops with members of the public and 8 depth interviews with experts. In total, 50 members of the public were consulted and 8 experts.

The research took place in stages as outlined below in Table 1. A staged approach was chosen to enable participants to first engage with and understand these complex concepts before they began co-creating materials about them. The final stage was designed to test the co-created resource with fresh participants to ensure it was understood and engaging. Interviews with experts were included to refine the resource and ensure its accuracy.

Table 1: Stages of qualitative research

Stage A: 7 workshops with members of the public	Stage B: 7 follow-up workshops to co-create a resource	Stage C: tested and refined the resource in 2 workshops and 8 depth interviews
Explored understanding of SDEs and TREs and how people felt about these environments. Showed existing resources and identified what worked well and less well in explaining these concepts.	Re-convened workshops with the same participants as Stage A. Participants used preferred messages and language already identified to develop their own resource. They also discussed their preferred format for the resource.	In 2 workshops with new members of the public a draft resource developed from Stage B was tested for clarity. Interviews with 8 experts optimised the content to ensure it was accurate and transparent.

<sup>2</sup> Dare UK: UK Sensitive Data Research Infrastructure: a Landscape Review. [UK Sensitive Data Research Infrastructure: a Landscape Review \(dareuk.org.uk\)](https://dareuk.org.uk/)

<sup>3</sup> [What Words To Use - Rapid Review.pdf \(understandingpatientdata.org.uk\)](https://understandingpatientdata.org.uk/what-words-to-use-rapid-review.pdf)

All the workshops and interviews on this project took place between January 2024 and March 2024. Fieldwork was conducted mostly remotely using an online video conferencing platform except for the workshops with digitally excluded participants which were conducted face to face.

### Public participants research sample structure

The sample for the workshops in Stage A and B included a good spread in terms of demographic criteria (age, gender, socio-economic group, ethnicity, disability and long-term health conditions, UK-wide geographic mix). Three important groups were included based on previous research conducted in this area: those who are digitally excluded and those more likely to opt out from the use of their health data. Previous research highlighted two groups at higher risk of opting out: Black ethnic minorities, and those aged 18-35 who are classified as A in terms of their socio-economic status.<sup>4</sup>

Table 2: Public sample in Stage A and Stage B

Workshop 1 & 8	Female, 18-40, C2DE, Northern Ireland	<ul style="list-style-type: none"> <li>Mix of ethnic groups</li> <li>Mix of rural and urban</li> <li>Include some with disability and/or long-term conditions in each group</li> <li>Include some carers within the groups</li> </ul>
Workshop 2 & 9	Male, 41+, ABC1, Wales	
Workshop 3 & 10	Female, 41+, ABC1, England	
Workshop 4 & 11	Male, 18-40, C2DE, Scotland	
Workshop 5 & 12	More likely to opt out: Black ethnic minority, England	<ul style="list-style-type: none"> <li>Gender and age mix</li> </ul>
Workshop 6 & 13	More likely to opt out: 18-35 year old, AB, England	<ul style="list-style-type: none"> <li>Gender mix</li> </ul>
Workshop 7 & 14	Digitally excluded, England	<ul style="list-style-type: none"> <li>Gender and age mix</li> </ul>

For the final two workshops with the public in Stage C it was agreed that participants from groups that are more concerned with the use of their health data would be consulted. This was to ensure that the level of detail met the needs of these audiences.

Table 3: Public sample in Stage C

Workshop 15	C1C2, 35-54, equal mix of Male and Female, a mix of different ethnic groups. England.	Urban and suburban Felt some level of concern about the usage of their health data.
Workshop 16	C2DE, 55+, equal mix of Male and Female, White British ethnicity. England.	Urban and suburban Felt unsure about how their health data is used.

### Expert research sample structure:

---

<sup>4</sup> [Public survey summary report - General Practice Data for Planning and Research \(GPDR\) - NHS England Digital](#)

The expert interviews consulted included those in advisory and communication roles as well as those working directly with health and care data both from a research perspective and a clinician perspective. The experts were identified with UPD and with the project steering group members.

Table 4: Expert sample

<b>Depth 1</b>	A stakeholder at NHS England involved in communicating data usage to patients.
<b>Depth 2</b>	A researcher working on a programme that involves communicating data usage to patients.
<b>Depth 3, 4 and 5</b>	<b>Clinicians</b> involved in the usage of patient data.
<b>Depth 6,7 and 8</b>	<b>Data 'owners', data advisors or researchers</b> involved in the usage of patient data.

### 3. Considerations for an SDE/ TRE resource

---

The aim of this research was to inform a resource to describe Secure Data Environments (SDEs) and Trusted Research Environments (TREs), which contain health data, to the public. Insight from this research identified several challenges to consider when developing a core Understanding Patient Data (UPD) resource and other resources to describe individual SDEs and TREs. This section will explore those four challenges.

#### The breadth of SDEs and TREs

A core UPD resource describing SDEs and TREs to members of the public will need to be sufficiently broad to cover technical differences that exist between individual SDEs and TREs. These technical differences include the data sources each individual SDE or TRE draws on, for example health data and/or types of data outside of healthcare, the differing processes that exist to approve access to the environment, different data processing and storage methods, different adoption of privacy-enhancing techniques, and the differing uses of the data including research, planning or population health management.

*“I think there’s a bit about being clear about what are the common attributes that are shared across these different environments, whilst recognizing that they might look and be set up differently.” [Expert]*

Due to the current lack of consensus on definitions for SDEs and TREs and to include as many environments as possible, we started with the broadest possible definition, which encompasses all use cases (direct care, planning and research), and identifiable, pseudonymised / de-identified data, and anonymous/aggregated data.

A significant difference between individual SDEs and TREs is whether the environment enables access to identifiable data, and whether it is for the purpose of direct care. For example, in the case of the NHS Federated Data Platform which is not currently planned to be used for research, just for direct care and planning, and we understand includes a [type of SDE](#) specifically for internal use. Both public participants and the experts consulted felt that this use case should be kept separate, and that focusing a core resource on secondary uses of data, not direct care, was simpler and clearer.

One of the most important points for public participants was whether they could be identified individually from the data. This was difficult to communicate, because identifiability is on a spectrum, and even with pseudonymised data, it is not correct to say that people can never be identified. However, when introduced to the concept of the use of SDEs with fully identifiable data for direct care purposes, public participants felt that this was a very distinct use case and should be explained separately. Experts were in agreement:

*“You would have to have 2 different types at the top level, one for identifiable information when you are getting treatment, and the other is just for health data.” [More likely to opt out, Black ethnic minority background, England]*

*“Data clearly needs to be identifiable for treatment. I assume that is always the case. Providing that example here is too much information, it’s confusing.” [Digitally excluded, England]*

*“Personally, I would just wipe out anything to do with the Federated Data Platform & I’d make it all about research”. [Expert]*

In the expert interviews, most experts concluded that a resource describing SDEs and TREs should be possible. These experts highlighted two commonalities that should be true across SDEs and TREs used for purposes other than direct care: first that individual level data does not leave the environment; and second that the data is de-identified before it is accessed. Some experts however pointed to the challenge of creating one resource that can cover all types of SDE and TRE given the complexity.

*“I think what you're talking about is a very secure, five safe environments where researchers can come and access data that they've put in a request for and they can't take it away.”*  
[Expert]

*“A simple flow of how data flows in and out of these environments will be different. That's why I struggle with how there's going to be a universal description of what makes something a secure data environment versus not.”* [Expert]

The recommended resource specification therefore focusses on the examples of SDEs and TREs used for secondary uses of data (not direct care) and attempts to be broad enough to cover the breadth of differences between SDEs and TREs as table 5 below highlights.

Table 5: Recommended types of data to cover in the overarching resource.

	Direct care	Planning	Research
Fully identifiable	No	No	No
Pseudonymised / de-identified	n/a	Yes	Yes
Aggregated / anonymous	n/a	Yes	Yes

### The changing landscape of SDEs and TREs

Another challenge faced in creating any SDE or TRE resource is that they are constantly evolving. There are still some unresolved nuances in the definitions, which the [UK TRE Community](#) has a specific [working group](#) dedicated to.

For example, there is ongoing work to create a set of standards which the NHS Secure Data Environment network should abide by, which could be referenced in the resource once complete. In addition, some experts reflected that SDEs and TREs may become more standardised in their technical infrastructure and merge, highlighting in particular Open SAFELY's model:

*“So, for example, we're moving to this Open Safely model and I'm sure that's something that will be adopted in a single TRE. With Open Safely, it's like this is the future, I think and it's going to be a very, very different way of doing things.”* [Expert]

The communications about SDEs and TREs are also evolving; the literature review undertaken as part of this study identified a lack of up-to-date literature on how to talk about SDEs and TREs<sup>5</sup>, and interviews with experts indicated that the sector is convening forums to aid discussion on this.

Any resource may therefore need to be adapted as SDEs and TREs evolve.

<sup>5</sup> [What Words To Use | Understanding patient data](#)



### A balance between transparency and clarity

In addition, there is a balance to be struck between transparency/comprehensiveness and ensuring information is clear and easy to understand for the public. Experts reflected that finding a balance between transparency and clarity is a very live source of debate and challenge in communicating health data messages to the public:

*"So, there's a real tension there between accessibility and accuracy and comprehensiveness. And we sometimes do struggle with NHS partners on just how much information to provide because they want to be legally watertight." [Expert]*

Some experts reflected that this desire for transparency is understandable in the context of historic instances where the NHS has been accused of a lack of transparency in the use of patient data. Experts were aware that being fully transparent can complicate communications, as there are exceptions to processes, policies still to be developed, and approvals yet to be sought:

*"You have this pure concept of something [TREs] and then there is a specific application which is under development and intense scrutiny, the NHS Secure Data Environment network. With that it is very easy to go into a rabbit hole" [Expert]*

In describing SDEs and TREs to the public, there were examples of technical details and exceptions which, whilst transparent and comprehensive, caused confusion and alarm for some general public participants. The following examples highlight this tension between transparency and clarity, and highlight some common misperceptions about 'what the public want to know':

- **Airlock systems:** In the early workshops, participants were shown diagrams to describe SDEs and TREs that included mention of an 'airlock' and descriptions of what data was imported and exported. Participants found this level of detail confusing, and it led to more questions:

*"That's too much detail. It's going to make me ask, who's approved that data centre? I feel like I have to keep going over and over it to understand it." [More likely to opt out, Black ethnic background, England]*

- **Central storage vs federated models:** Participants were also shown information about a storage solution model compared with a federated data model. Participants found the concept of federated data complex to grasp and some questioned whether this was information the public needed to know:

*"Normal everyday people don't need this information; this is too much. The researchers might care, but we don't care." [Digitally excluded, England]*

*"I don't think we need to know specifically whether it is stored in one place or lots of places. It is accessed securely on some sort of system anyway, the difference doesn't matter to me." [Male, 18-40, C2DE, Scotland]*

- **Data extraction:** In the third stage of workshops, participants were introduced to the idea that individual level de-identified data could be taken out of an SDE or TRE in exceptional circumstances. This was introduced because stakeholders felt there may be circumstances where this happens. The mention of exceptional circumstances caused concern for participants who wanted to know what these were and why data was leaving the system. Experts were equally concerned:

*“Have they got an example of exceptional circumstance, that worries me. We can guess or assume but it is ambiguous.” [Mixed gender, 35-44, C2DE, England]*

*“If this mentions exceptional circumstances without an example it is leaving the public with questions and no information to answer them.” [Expert]*

*“I have never come across an exceptional circumstance. This needs an example.” [Expert]*

The resource specification for a core UPD resource that accompanies this report has identified the messages that the public found most relevant and has sought to balance transparency with clarity. The recommendation is that some of the more detailed and technical messages (such as data extraction examples where these exist) sit in a third, more detailed, level of information separate from the overarching resource.

### **Providing a balanced tone of voice that does not reassure too much**

Public participants sought reassuring language and naturally preferred phrases like ‘secure’, ‘highly secure’, and ‘safe’. The preference amongst some experts on the other hand was to provide the public with balanced facts and careful language choices to enable them to make an informed decision themselves about the security and safety of their data:

*“I like the word ‘highly secure’ that sounds very reassuring.” [Female, 41+, ABC1, England]*

*“You would argue that Understanding Patient Data should be completely neutral. They should be patient-side. I would argue the message that you've shown there is very positive and reassuring, but is that the absolute truth?” [Expert]*

Any resource developed to describe SDEs and TREs should be mindful of this balance between being informative and being perceived to be too reassuring. The word ‘secure’ was deemed to get this balance right.

#### 4. The core topics the resource should cover and how to convey these to the public

---

As part of this research, public participants were shown information describing Secure Data Environments (SDEs) and Trusted Research Environments (TREs), and asked to determine which information was of relevance and importance to them. They felt a resource describing SDEs and TREs to members of the public should cover the following topics:

- The use of large health data sets for purposes other than direct care;
- The flow of data into and out of an SDE or TRE, specifically focusing on the fact that data is de-identified before it is accessed;
- The fact that data is accessed not shared;
- A description of who accesses the data;
- The benefits of using SDEs and TREs to access large health data sets;
- How the data is kept secure.

This section will take each of these topics in turn and outline why participants felt this was important and how they felt about this aspect of SDEs and TREs, what types of information they were looking for, and the terminology that worked well to convey each point. This section will also incorporate views from the experts consulted on how each point should be described.

##### Introducing the uses of health data

Knowledge of the uses of health data outside of direct care was limited amongst the public sample in this research. Some participants understood that their health data might be used for purposes outside of direct care including research, but this was only in very broad terms:

*“Health data is your personal health records; health professionals pass it around each other. I think it's all centralized, like any scans or anything. I imagine it is all on one system?”*  
[Female, 18-40, C2DE, Northern Ireland]

*“I imagine that statistics are shared with researchers or pharmaceutical companies to look at vaccinations and things like that.”* [Female, 41+, ABC1, England]

Describing how health data is used was of particular importance to some of those in the digitally excluded group who had not considered that their medical record would be stored or accessed anywhere other than their GP practice:

*“Medical data from appointments with doctors is for the doctors only. Why would anyone else want to or be allowed to see that? I imagine it is stored on a computer or in a file somewhere.”* [Digitally excluded, England]

Participants liked the introduction to the use of health data that UPD have already developed, ‘A guide to large datasets’.<sup>6</sup> They felt the language used was accessible and clear and that a summary of this content should be included within a SDE/ TRE resource. The inclusion of this content would ensure the resource could be accessed on its own without the need to seek out other information first:

*“I’m okay with this slide. It feels like it’s introducing you and giving you the background. It’s relating to things we’re familiar with.”* [Mixed gender, 35-54, C1C2, England]

---

<sup>6</sup> [What happens to health data: a guide to large datasets | Understanding patient data](#)

The experts interviewed felt that the content would need to be tailored to provide a summary specific to SDEs and TREs. For example, they suggested avoiding the mention of ‘small amounts of data’ because, in some cases, the data sets contained in SDEs and TREs can be large. Experts also felt the description could reference ‘planning’ more clearly. In addition, there was discussion about whether to mention data other than health data in this description, or whether this is too complicated and may alarm the public:

*“I think the resource needs to be transparent that health data is linked to other sources of data. This is particularly important when thinking about the Our Future Health study which will contain a variety of data types.” [Expert]*

*“I think for the purposes of introducing this concept, it's right to keep it focussed on health and then perhaps - once people have got their heads around the idea - then you could say, in some circumstances, we might also want to bring in data from other sources.” [Expert]*

When participants were introduced to the idea that health data could be linked to other data sources in these environments, this led to further questions and, for some, concerns. Public participants therefore believed that a focus on health data would be the best approach for the main resource:

*“Surely all data is linked together using whatever system they are using? I don't understand why that is different.” [Male, 18-40, C2DE, Scotland]*

*“When you start giving too much detail like this it brings in more questions for me. It sounds like my data is going to lots of places now and is being spread out.” [Mixed gender, 35-54, C1C2, England]*

The core UPD resource co-created with the public uses an edited version of the text from the UPD Guide to large data sets. Based on feedback from the public, this does not mention data outside of health and care currently.

**Whenever we go to a GP or hospital, information will be collected about us and our medical history. Only those who are directly involved in your care will be able to access your full patient record. But some of the information from your record may also be useful for specific purposes beyond your individual care.**

**If different bits of health and care data that cannot identify people from many patients are linked up, researchers and healthcare staff can look for patterns in the data, helping them develop new ways of predicting or diagnosing illness, and identify ways to improve clinical care.**

### **The data flow and de-identification of the data**

From public participants’ perspective, the most important aspect of SDEs and TREs was the fact that the data accessed cannot individually identify them. They did not feel comfortable with researchers or anyone who was not directly involved in their healthcare having access to their identifiable patient data:

*“I'd hope it's anonymised for research purposes and to offer more tailored care and support in different areas of the country.” [More likely to opt-out, 18-35, AB, England]*

*“If they don’t know who you are, what harm’s it going to do? I don’t see a problem if they can’t identify who you are.” [41+, Male, ABC1, Wales]*

In the first stage of research, participants were shown diagrams explaining the flow of data through a SDE or TRE and where de-identification takes place. Whilst some found this helpful, many found these diagrams too complex.<sup>7</sup> References to ‘importing’ and ‘exporting’ data caused confusion, and some struggled to understand the concept of the ‘airlock’. In later workshops, a simpler version of a diagram that still retained mention of an airlock was created but participants still felt that it was too complex:

*“Even if I wanted more information, this would blow my mind and I’d be worse off.” [Digitally Excluded, England]*

*“Busy and confusing, too much.” [Female, 41+, ABC1, England]*

Instead, participants sought a simplified description or visualisation of the de-identification process that conveys three points:

- Data entering the SDE/TRE;
- Identifiable information being removed and replaced with numbers;
- Results (but not the de-identified data itself) leaving the SDE/TRE in the form of a graph.

To explain these points, participants liked the following terminology:

- Use of the phrase ‘data that is not identifiable’ or ‘data that cannot identify you’ rather than the term ‘de-identified data’. Participants found these phrases easier to grasp quickly.
- Experts recommended the phrase ‘data that is not identifiable’ would need to be checked by an IG expert as they thought that, in some cases, identifiable data might be viewed by researchers, and that de-identified data can often be reidentified (as referred to in the Goldacre Review).
- A visual representation of a line of information containing name, address and NHS number, and this being removed and replaced with a number. The visual included within the UPD ‘What happens to health data’ infographic conveyed this successfully.
  - Participants were introduced to the concept of de-identification and the fact that data could be re-identified. This was a challenging concept to convey quickly as participants assumed data was either identifiable or anonymous. This visualization of de-identification was better understood than a text description.
- Words such as ‘statistics’ and ‘graphs’ were more readily understood as outputs than the phrase ‘results’ or ‘analysis’. Participants struggled to understand what was different between ‘results’ and the data itself:

*‘Data is not taken out, only the results from the analysis are taken out’: “How can they use the data if it is not taken out?” “What is the difference between analysis and results?” [Mixed gender, 55+, C2DE, White British ethnic background, England]*

*“The fact that it explains that the data is in tables and figures and diagram form, rather than on an individual basis, just re-confirms the de-identified bit. I think that’s useful.” [Male, ABC1, 41+, Wales]*

---

<sup>7</sup> Two were shown: Genomics England diagram for 100,000 genomes project [Document library | Genomics England](#) and [Trusted Research Environments \(TREs\) – A Guide for Beginners - CF \(carnallfarrar.com\)](#).

Experts agreed about the importance of conveying the fact that data is de-identified. Some experts felt that the resource should go further and explain the possibility that data could, in theory, be re-identified:

*“Should this mention pseudonymisation so it is clear that they are preserving some method of re-identifying it?” [Expert]*

However, pseudonymisation was a complex topic for the public to grasp and one which public participants felt would be challenging to include in a resource designed to describe SDEs and TREs. When introduced to participants in the early stages of research, the UPD resource describing de-identification and the possibility of re-identification did cause concern, particularly amongst those who don't want to tolerate the uncertainty involved with de-identified data<sup>8</sup>:

*“So, is it anonymous? It's either anonymous or it's not. To make it this grey area of completely anonymous or not is creating too many questions and worries for me. You're either completely protected and anonymised or you're not.” [More likely to opt-out, 18-35, AB, England]*

*“What worries me is this bit about anonymous data not being confidential along with the tiny risk it could be re-identified.” [Digitally excluded, England]*

Public participants felt that a link to the UPD resource about de-identification could be provided for those who want to know more details. This would therefore reduce confusion or concern when the public first view the resource but provide transparency for those who seek it.

There was some discussion about whether the resource should make clear that, in a majority of SDEs and TREs, de-identified data can be used for certain purposes without consent. Members of the public who were more concerned about the use of their patient data felt this should be made clear, and some experts thought the resource should be fully transparent on this point:

*“So they can access our data without our knowing?” [Mixed gender, 55+, White British, England]*

*“Unless people feel they have some say in what their data is used for, how it is used and who is using it, there is a real danger that you lose public good will.” [Expert]*

The recommended resource text was co-created by public participants because it clearly conveys the point about personal health data being de-identified. Experts liked the transparency of explaining that personally identifiable information is replaced with a number (i.e. that it is de-identified, not anonymised).

**Before researchers can directly access the data in the highly secure environment, any identifiable information such as names and NHS numbers, are removed and replaced with numbers. This data is called de-identified data.**

### **Data access rather than data sharing**

Experts and public participants agreed that the fact that data is accessed, and not shared, is a core principle of SDEs and TREs that needs to be communicated. Participants broadly understood the

---

<sup>8</sup> UPD. What happens to health data, a guide to large datasets.

word ‘access’ to mean viewing the information and the word ‘sharing’ to involve actually sending it to someone. Participants could see the benefits of data access rather than data sharing because they believed this would be more secure:

*“The fact that it is accessed and not copied or shared is super important to me.” (Female, 41+, ABC1, England)*

*“Downloading means they have a permanent record of it, so it’s better if they can’t do that because they don’t own it and can’t store it.” [Mixed gender, 35-55, C1C2, England]*

Conveying this concept to the public in a way that made sense was more challenging. The following phrases caused confusion or concern within the workshops:

- **‘Can access it remotely’** – some participants were unsure what remote access meant and worried that the word ‘remote’ sounded less secure: firstly, because it could be accessed from anywhere and, by extension, anyone; and secondly, because they worried that researchers could remotely access all the data:

*“A secure room sounds really safe ... then you go, oh, remote connections, that’s not so safe” [Female, 41+, ABC1, England]*

*“The word remote access concerns me. That implies that they can access everything and might have access to the same information as doctors? How do they control what you can see?” [More likely to opt-out, 18-35, AB, England]*

- **‘It’s a bit like a reference library’** – while the core analogy was understood participants felt the comparison with a library implied a physical storage space and that you could access (and copy) the individual level data. That said, participants did like the warmer, less technical, terminology:

*“The library worries me. You can copy lots of things in a library? Can’t you take books out of a library too?” [Male, 41+, ABC1, England]*

*“The phrase ‘like a reference library’, it’s like someone having a conversation with you rather than reading something that’s dead formal. It feels less official, nicer.” [18-35, AB, England]*

- **‘Access to health data for analysis, without them needing to receive a copy’** – public participants were unsure what type of copy this referred to and it led to questions about the security:

*“What stops somebody receiving the information making a copy themselves?” [Male, 18-40, C2DE, Scotland]*

- **‘This means that multiple copies don’t need to be sent out or downloaded’** – public participants pointed out two challenges with this statement. Firstly, the idea that multiple copies were ever sent out or downloaded caused alarm; and secondly, the link between being given access and not needing to send out copies was not always understood. Participants preferred a more direct statement such as **‘copies can’t be sent out or downloaded’**.

*“Then it says ‘this means’, but that’s not obvious to me. It’s not obvious to me that multiple copies don’t need to be sent out, those two ideas don’t link for me.” [35-54, C1C2, England]*

- **‘They can only export analysis results and only after careful checks have been made’** – participants were divided in their opinions about this message. Some found the checks reassuring because it sounded thorough; others raised questions about why the data needed to be checked and whether the checks were relevant to the public or if these were more about quality control:

*“What are they checking for? I thought it wasn’t identifiable? Is this more of a message for the researchers so they can be sure it is good data?” (Male, 18-40, C2DE, Scotland)*

Some public participants wanted this description to outline the methods used in the past in order to clarify how this process is different. However, experts pointed out the challenges with doing this and thought that it could cause concern:

*“The public would be horrified that in the past this data has been sent out on disks with a promise that they would delete it once finished.” [Expert]*

The experts consulted agreed that this message is important but needs to be kept simple:

*“The fact that data is not shared is the whole point about SDEs and this needs to be made clearly.” [Expert]*

*“The point about multiple copies not needing to be taken out isn’t clear enough. Instead I would say it is stored in a central place and can’t be downloaded.” [Expert]*

Some experts believed that the fact results are checked is an important point to convey because it is an important part of the five safes principles designed to guard against individuals being indirectly identified. The public however, were less interested in this as outlined above. This point was challenging to convey without a detailed explanation and participants became quickly confused between data being de-identified then data being checked to remove any potentially identifiable information.

*“The other thing to mention is that the results are checked. There is a process to check the results too and the data to make sure no one can be identified from the results.” [Expert]*

Public participants co-created the following description. They liked how this statement conveys that data does not leave the SDE/TRE and felt that the phrase ‘sent out or downloaded’ was sufficiently clear. They felt that this phrase should be linked to the description of de-identified data, so it is clear what is not being taken out of the SDE/TRE (i.e. individual-level data) and what is being taken out of the SDE/TRE (i.e. results).

**This data stays within the environment and cannot be sent out or downloaded.  
Only the results from the analysis are taken out.**

### **Who uses the information and how they are approved**

Information about **who could access the data held within an SDE/TRE** was of high importance to the public participants consulted. After establishing that data was de-identified, this was deemed the



next most important topic. This was because they sought reassurance, and because they wanted to know the data was being put to good use. They also sought reassurance that their data was not being used by organisations in which they had lower levels of trust (e.g. insurance companies):

*“So it’s not just NHS, it’s associated companies as well. These associated companies, as much as they have a contract with the NHS, are they allowed to do whatever they want with data?” [More likely to opt-out, Black Ethnic minority background, England]*

This is a topic on which participants sought more detail in the form of a transparent list of examples of organisations or individuals who could access the data. During the workshops the phrase ‘approved researchers’ or ‘approved users’ was tested as an umbrella term for those who could access the data. Participants did not feel that the use of this term provided enough information and it led to further concerns and subsequent questions:

*“Who are approved researchers? How are they approved? It needs more detail to take the questions out of it.” [Female, 18-40, C2DE, Northern Ireland]*

*“This is one case where it’s good to be specific. Does this include private companies and private individuals? That would make me more wary of it.” [Female, 41+, ABC1, England]*

While the phrase ‘private companies’ did prompt some questions, participants wanted transparency about who could access the data, and experts were in agreement:

*“The mention of private companies is a red flag but it’s an important inclusion for transparency I think. This resource must include it as an example.” [Expert]*

If an umbrella term is used to describe those who can access the data, participants preferred the use of ‘researchers’ rather than ‘users’ or ‘individuals’ because ‘researchers’ conveyed that they would have expertise in health data analysis:

*“Individuals implies it’s anybody. Researchers implies that it’s trusted people who are knowledgeable in the subject matter.” [More likely to opt-out, Black Ethnic minority background, England]*

Some members of the public were keen to understand if the NHS received a commercial benefit from private companies accessing the data. While they did not feel this was needed within the main resource, there may be an appetite to point participants in the direction of information on this topic elsewhere on the UPD website.

*“Who’s accessing it, why are they doing it and what is the benefit for the NHS? I’d really like to know commercially that they are getting a lot from it.” [Female, 41+, ABC1, England]*

Some participants felt this information should be supported with further details in the form of case studies (for example to explain why researchers from a charity would need to access it) and how approval is granted:

*“What’s the vetting process for these people getting access? In all 3 examples it doesn’t say.” [Male, 18-40, C2DE, Scotland]*

The level of interest in the approval process was highest for those who claimed that they were more likely to opt-out; namely participants from a Black Ethnic minority background and 18-35 year old

participants from an AB socio economic background. They sought information on who was approving access and what criteria individuals or organisations had to meet. Both groups shared concerns about commercial organisations having access to the data and a belief that they would be using it for profit. The research tested some examples of how to describe the approval process and the following phrases were positively received:

- **‘Signing an agreement to always protect the confidentiality of your data’** – participants liked the idea that those accessing the data had to agree, in writing, to use it appropriately;
- **‘Demonstrate that their proposal is an appropriate and ethical use of the data’** – the focus on ethics was liked as it conveys that the data is being treated seriously;
- **‘Demonstrate that it will deliver clear public benefits’** – participants liked this phrase as it brought the focus back to a benefit for them personally. Others felt the phrase sounded like ‘politician speak’ and did not trust the sentiment;
- **‘Organisations follow set processes and criteria, some of which are set out in law’** – the mention of law gave the approvals process a sense of gravity and seriousness;
- **‘Approved by an access review committee’** – participants liked the idea that a group of individuals would decide, though some did want more detail on who would sit on this committee. There were mixed responses to the idea the public would be represented on the committee. For some the idea that the public had a voice was reassuring; others found this idea confusing because they assumed that members of the public would lack specialist knowledge:

*“Do they have to have certain qualifications to be on the committee”? [55+, C2DE, White British, England]*

*“You could just say it’s the NHS Ethics Committee. A name is all. Then if I want to know I can Google that body, I can research that body and close that book. So, you do need this information, maybe condensed a little bit, but we just need the body’s name.” [More likely to opt-out, Black Ethnic minority background, England]*

However, experts believed that providing information within the main resource about how approval is provided was challenging. They outlined that each SDE or TRE may approach this slightly differently and therefore providing examples could be misleading:

*“Listing criteria could get you into a whole heap of trouble because you will immediately stumble upon things that certain ones don’t use.” [Expert]*

*“This resource needs to make clear that the criteria by which access is granted and who sits on the committee will differ.” [Expert]*

Experts also pointed to the challenge of assuring the public that the research approved would have a public benefit because early-stage research cannot guarantee this will be the case:

*“I worry about over promising there. You can say that the intent is to have that, but you can’t guarantee it. And research is, by its nature, not guaranteed to have the outcome that you anticipate.” [Expert]*

Public participants co-created the following description:

**To access this data, researchers go through an approval process and there are several criteria that they must meet before they receive approval. Researchers within NHS organisations, universities,**

charities and private companies such as pharmaceutical companies are examples of who could apply for approval to access the data needed for research.

### The benefits of using SDEs and TREs to access large health data sets

The research tested three ways of describing the benefits of accessing large health data sets within SDEs/TREs, each with increasing levels of detail:

- a) A one-sentence option - **‘to save and improve lives’**;
- b) A more detailed description - ‘Researchers can identify better ways to predict and diagnose illness, help develop new treatments and monitor the safety of existing treatments, support planning of health services, and help address health inequalities;’
- c) And case-study style examples such as ‘Providing academics with access to cardiovascular and cancer data for Covid-19 research.’

Participants preferred option b, which listed some examples of the benefits of secure health data analysis. The phrase ‘to save and improve lives’ was well received. Participants felt this was a positive introduction and quickly conveyed the benefits of giving researchers secure access to health data via SDEs and TREs. They suggested that including a few more examples of what the benefits of secure health data access are would be helpful, informative and qualify this statement:

*“It’s an impactful statement to say it saves and improves and lives.” [Male, 41+, ABC1, Wales]*

*“I think the examples are things you can relate to and it makes it easier to understand this sort of thing. I think they should be included.” [Female, 18-40, C2DE, Northern Ireland]*

While participants felt that examples or case studies could bring the benefits of secure access to health data to life (as in option c above), they found some of the examples provided a little ‘too academic’ (e.g. ‘analyse genomic and long-term clinical data from health records to gain insight into the nature of genetic changes that drive cancer evolution.’). They were interested in examples close to their own experience, e.g. planning where to put a new GP surgery:

*“I think the examples are good if you want more information. They should focus on projects that have a real benefit and back up the fact that they save lives.” [More likely to opt-out, Black Ethnic minority background, England]*

*“Both examples are of research. It would be good to have examples of how it’s helped plan our services, to make it more personal. These examples are very academic. Maybe examples of how it makes real people’s lives better.” [35-54, C1C2, England]*

Those aged over 41 and those from the digitally excluded group were the most interested in what health data was used for and the purpose of collecting it. These audiences wanted to know what benefit this would bring to the public and to the NHS more broadly:

*“It’s really good solid stuff that the research is being used. I’m thinking I’m not so concerned about my data now; I’m thinking about all this good research it’s being used for.” [Digitally excluded, England]*

Most of the experts consulted agreed that this was an appropriate level of information to include about the benefits of providing secure access to health data:

*“This type of information should be up-front for the public. Finding new treatments, this is why we do it. I worry that leading with more technical detail could worry them.” [Expert]*

Experts felt, however, that there was a fine line to be navigated between providing information about how the data is used and appearing to be ‘selling’ the benefits:

*“I would argue the message that you’ve shown there is very positive and reassuring but is that the absolute truth?” [Expert]*

*“Should the wider uses this data could be put to be included for transparency or would that backfire and make the public more concerned?” [Expert]*

Some experts felt that the resource should include more information about the importance of large datasets in enabling researchers to see patterns. This concept was introduced in the early public workshops but participants felt this was more of a benefit for a technical audience:

*“The main point here is it is about the ability to look at data at scale to identify patterns that would not be possible looking at smaller amounts of data.” [Expert]*

*“I would imagine that the data being linked together and looking at big data sets is common sense. I’m not sure why that needs to be in there.” [Male, 41+, ABC1, Wales]*

*“I find this confusing. I’m not clear why this is a benefit? It doesn’t feel like it’s speaking to me as a member of the public.” [More likely to opt-out, 18-35, AB, England]*

The public had some interest in the phrase ‘help make research more efficient, cost-effective and collaborative’ because they felt that the NHS needed solutions that were cost effective and efficient. Others found the reference to cost-effectiveness and efficiency confusing because they did not know how an SDE or TRE would deliver these benefits. To appreciate these benefits, they felt that a comparison would need to be made between the way in which SDEs/TREs work and how this process has been managed previously:

*“Efficiency has got to be a good thing because traditionally vaccines and things like that can take decades.” [Female, 41+, ABC1, England]*

*“How does it make it more efficient and cost-effective? I don’t get that.” [Male, 18-40, C2DE, Scotland]*

Public participants co-created the following text:

**Secure Data Environments and Trusted Research Environments provide access to information that can identify better ways to predict and diagnose illness, help develop new treatments and monitor the safety of existing treatments, support planning of health services, and help address health inequalities.**  
**They also help make research cost-effective and collaborative.**

### **Security of the data**

The security of the SDE/TRE system was of importance to public participants but the degree of information desired about this topic differed based on demographics. For some, the word ‘secure’

communicated enough about the security of the data, although some felt that care should be taken not to over-use the word throughout the resource as too much repetition could prompt them to question it:

*“When you repeat the word ‘highly secure’ like they have done, and you emphasis it too much, it feels like they’re saying ‘it is secure I promise, it really is secure’. They need to back it up.” [Mixed gender, 35-54, C1C2, England]*

Those who were happy with this level of detail were more likely to be aged over 41 and from a C2DE socioeconomic group:

*“It’s good knowing that this organisation is out there protecting our data.” [Digitally excluded, England]*

Those who wanted to know more detail about data security included those aged 18-40, and those who claimed to be more likely to opt out (namely participants from an AB socio-economic group background and those from a Black ethnic minority background).

These participants wanted the word ‘secure’ to be qualified with more detailed information. However, when the public thought about security, they were thinking about cyber-security and how their information will be kept safe from hackers. This is a topic that the public participants consulted were worried about, and these risks were implicitly understood. They did not ask for the resource to acknowledge this was a risk, instead they sought information about how the risk is mitigated.

*“I’m concerned about whenever data is stored, you see stories about data being leaked and hacked” [Mixed gender, 35-54, C1C2, England]*

*“Most things are digital now, but can someone hack in and get my information?” [More likely to opt out, Black Ethnic minority background, England]*

None of the information tested in this research quite addressed this concern. Cyber security is not specifically covered in the five safes and one example describing cyber security measures (below) was deemed to be too vague:

- ‘NHS data will only be hosted on systems that can prove they have high levels of protection.’<sup>9</sup> Participants were unsure what ‘high levels of protection’ meant:

*“This doesn’t feel like a very secure data environment. What do they mean ‘prove they have high levels of protection’? I’d rather it said they ‘have’ high levels of protection.” [Male, 18-40, ABC1, Wales]*

*“I want some reassurance about how it’s protected, what measures are in place, how they’re keeping it safe. You hear a lot about data leaks and breaches, even with very big companies. It’s a personal matter when it’s your health.” [Mixed gender, 35-54, C1C2, England]*

*“I want to know about how strong their firewalls are. How safe is it from outside interference” [Male, C2DE, 18-40, Scotland]*

---

<sup>9</sup> [Secure Data Environments \(SDEs\) - Data saves lives - NHS Transformation Directorate \(england.nhs.uk\)](https://www.england.nhs.uk/secure-data-environments/)

Some participants felt a line could be included within the main resource on this topic, with links to further detail. Other participants felt this should just be a link to further information. Experts however raised the point that each environment will have different security measures in place so providing an explanation that is true of all SDEs and TREs could be a challenge.

As part of the broader conversation about data security, the Five Safes were introduced to public participants.<sup>10</sup> These were felt to give a good overview of the importance placed on using health data safely. Participants understood ‘safe people and safe data and felt these were most relevant to them. Safe projects, safe outputs and safe settings were less well understood (and participants felt that none provided enough information about cyber-security):

*“The word safe is there. So that’s what it’s making you feel like, everything is safe and above board and all bases have been covered. I think that’s really worth trying to reassure the public about.” [Female, 18-40, C2DE, Northern Ireland]*

- **Safe people** – participants felt it was important to convey that those accessing the data had to be approved and to provide examples of who they are within the main resource. This content is discussed in the section above titled ‘Who uses the data and how are they approved’.
- **Safe data** – the fact that the data is de-identified was of high importance for participants. This content is discussed in the section above titled ‘The data flow and de-identification of the data’.
- **Safe projects** – The fact that data is made available for projects in the public interest was interpreted as more of a benefit than a reassurance about security:

*“I like that the data is only made available for projects that are in the public interest. That’s a positive benefit to this.” [Female, 41+, ABC1, England]*

This content is discussed in the section above titled ‘The benefits of using SDEs and TREs to access large health data sets’.

- **Safe output** – the sentence ‘research findings are checked by TRE staff, with any potentially identifying information removed’ caused some confusion and concern. Participants believed, based on what they had been told about the de-identification process, that data was de-identified before researchers could access it and therefore became confused about the idea that research ‘findings’ could ever contain identifiable information:

*“This gives you the fear that identifying information could be given out. It shouldn’t even be in there at this point.” [Male, 18-40, ABC1, Scotland]*

Participants did not feel this information was needed within the main resource; but felt that if the five safes was provided as further information then it would need clarification.

---

<sup>10</sup> [https://dareuk.org.uk/wp-content/uploads/2022/04/DARE\\_UK\\_Public\\_Dialogue-What\\_is\\_a\\_trusted\\_research\\_environment.pdf](https://dareuk.org.uk/wp-content/uploads/2022/04/DARE_UK_Public_Dialogue-What_is_a_trusted_research_environment.pdf)

- **Safe settings** The reference to a secure room or an approved, remote connection caused concern for some who felt the phrase 'remote connection' implied less security. Others wanted more information about cyber security instead. A few participants were interested to know where the data in each specific SDE or TRE is stored:

*"As soon as I read it I thought, 'oh well I can screenshot it'. They should say how that isn't possible before I can even think it." [Male, 18-40, ABC1, Scotland]*

*"I assumed that the NHS held all the information, so who actually holds the information and where are these trusted environments?" [Mixed gender, 55+, White British, England]*

The insights gleaned from sharing the Five Safes as stimulus have been incorporated into the recommended resource; however, this research would suggest that a reference to the five safes should be provided as further information rather than within the main resource.

None of the information tested met the need of describing security of SDEs and TREs in a way that public participants wanted although this topic was important to some demographic groups. The research indicates that more information about security should be included both within a resource developed by UPD and any resource specific to individual SDEs or TREs.

**The data stored in Trusted Research Environments and Secure Data Environments is kept secure by ...**

## 5. The name Secure Data Environment and Trusted Research Environment and the use of both

---

Public participants were comfortable with the use of both Secure Data Environment (SDE) and Trusted Research Environment (TRE) if the language makes clear they are the same thing. The phrase ‘Secure Data Environments, sometimes called Trusted Research Environments’ conveyed this well. In later sessions, participants were shown the phrase ‘Secure Data Environments and Trusted Research Environments’ which caused more confusion as participants were unsure if they were different and if so, why they were both being grouped together:

*“The use of the word ‘and’ in-between Secure Data Environments and Trusted Research Environments does imply that they are different.” [Mixed gender, 35-54, C1C2, England]*

Participants were shown information about the differences between SDEs and TREs (namely that SDEs are run by the NHS) in the final workshops. A majority felt this was too much detail for a member of the public to comprehend and that this additional detail created further confusion:

*“I don’t see why we need to know this. Take that out. I can’t see this being interesting or useful to members of the public.” [Mixed gender, 35-54, C1C2, England]*

There was discussion in the steering group and in the expert interviews about whether there were differences between SDEs and TREs, and whether these were material differences when it comes to engaging with the public. Overall, experts consulted did not feel that any differences between SDEs and TREs should be included as part of the resource. (Some experts consulted did not feel there were any technical differences between the two):

*“I did not think there was a difference between them other than the accreditation programme for SDEs, so I wouldn’t include that.” [Expert]*

*“I struggle to see a benefit in trying to describe to the public a topic which experts have struggled to describe.” [Expert]*

For contextual awareness, participants were asked to discuss their views about the term ‘Secure Data Environment’ and ‘Trusted Research Environment’. Overall SDE was preferred. Participants felt that the word ‘secure’ conveyed an important point about the environments and sounded more factual. Conversely ‘trust’ was perceived to be a looser concept. Participants felt it was their decision whether they trusted these environments or not:

*“I think secure is to generate the trust element. There are certain protocols that have been met first, certain firewalls and parameters set around it. The word secure suggests being monitored as well, so there’s oversight there.” [Male, 18-40, ABC1, Wales]*

*“Something can be trusted, but not necessarily secure. The main thing for me is that it’s secure and anonymous and private, so the second one for me [SDE] signals that better than the top one [TRE].” [More likely to opt-out, 18-35, AB, England]*

*“I prefer the word secure because trust is a word that’s thrown about too much. If somebody tells you to trust them, I generally don’t trust them.” [Female, 18-40, C2DE, Northern Ireland]*



When first introduced to the names, participants interpreted 'TRE' as a place that only trusted researchers could access, and 'SDE' as a place where data is stored securely. Once introduced to various uses of the data, for example for planning, some participants highlighted that the inclusion of the word 'research' in TRE gave the impression that TREs was only used by scientific researchers:

*"I'd imagine a Trusted Research Environment is a lab that does research on people or like a research trial." [More likely to opt out, Black Ethnic minority background, England]*

For the purposes of the core resource, both SDE and TRE have been used in the description because these are both terms the public may come across:

**Secure Data Environment, sometimes known as a Trusted Research Environments.**

## 6. Providing layers of information – extra information for those who may wish to seek it

Public participants suggested that there are three levels of information required in a resource describing Secure Data Environments (SDEs) and Trusted Research Environments (TREs) to satisfy differing information needs:

- a) A one-line summary of what an SDE/TRE is - some public participants felt this would provide sufficient information. This view was most common amongst participants aged over 40, most male participants aged 41+ in Wales as well as digitally excluded participants:

*“When it comes to information, you want it simplified, to quickly understand it and get on with your day. Most people don’t want to read too much, they like one-liners.” [Male, 41+, ABC1, Wales]*

This one line summary should include those topics which participants felt were most important, namely that the data is not identifiable, why it is being used, that it is being accessed and who by.

- b) The main resource which covers the topics outlined in section 4 -most public participants felt this was the level of information they would find useful.

The main resource should introduce the broader context and cover the topics that were felt to be of relevance to members of the public.

- c) A series of links to further detail on specific topics for those who are interested. Participants who were most likely to seek out further information were those who claimed that they were likely to opt out (namely, those from a Black ethnic minority background and those aged 18-35 from an AB socio economic group in England).

*‘I’d like to know a little bit more. Just to make you feel reassured, and for you to understand exactly what’s going on. If it was just a few words like, you’re safe, we’re doing the right thing, that wouldn’t be enough.’ [More likely to opt-out, Black Ethnic minority background, England]*

In providing these layers of information the resource can provide access to full transparency and depth for those who wish to seek it out, but without overwhelming the public with too much detail within the main resource which risks causing confusion.

Table 6 below outlines these three levels of information and summarises what topics participants would like to see covered at each level.

Table 6: The layers of information desired by the public participants in this research.

Cover in 1 line summary	Cover in the main resource	Cover in ‘further information’
A way to convey why this is being introduced e.g. it is a method of storing/ accessing data.	Introduction to the use of large data sets.	<ul style="list-style-type: none"> <li>• Link to existing UPD resource.</li> <li>• Further details about the different uses of patient data, outside of research.</li> </ul>
The data that is accessed is not identifiable.	The data that is accessed is not identifiable.	<ul style="list-style-type: none"> <li>• UPD resource about de-identification.</li> </ul>

This is about data access, not data sharing.	This is about data access, and data is not shared/downloaded.	<ul style="list-style-type: none"> <li>• How the data is kept secure and safe from hackers.</li> <li>• Where the data is held.</li> </ul>
Who is accessing the data.	The data cannot be taken out of the environment.	<ul style="list-style-type: none"> <li>• The five safes.</li> </ul>
	Who is accessing the data.	<ul style="list-style-type: none"> <li>• What data sources can be drawn into a SDE/TRE e.g. local authority data.</li> </ul>
	Access is controlled and only those who are approved get access.	<ul style="list-style-type: none"> <li>• Details about how approval is granted.</li> </ul>
	What the data is being accessed for; what is the benefit.	<ul style="list-style-type: none"> <li>• Examples or case studies illustrating where large data sets stored in SDEs/TREs have been used.</li> </ul>

Experts agreed that providing layers of information would be helpful because it provides sufficient and relevant information for different levels of interest and chimes with what they have heard from members of the public in their own consultations:

*“The conclusion I have come to is that we need to communicate with people on levels. So, for example, some people might have a deep interest in cyber and they need to be able to get down to the level that they need to make them comfortable. But most people are at the top level, and that top level will deal with 80% of people and for the other 15% you produce other layers.” [Expert]*

*“That's an interesting point about a simpler description. There will be some though that want the details, so I guess it's whether you can have two levels of detail. The easy and the deep.” [Expert]*

Both public participants and experts felt the layers of information could work well both for the UPD resource and information but also for communications about a specific SDE or TRE. In these cases the UPD resource could be used as an introduction and the links to further detail could provide specifics on how it works in each case.

## 7. The format of a resource describing SDEs and TREs and how it could be used.

---

Public participants preferred an animated video with a voiceover as the format for the main resource to describe Secure Data Environments (SDEs) and Trusted Research Environments (TREs). They felt that a combination of visuals and a voiceover would be beneficial in order to explain and visualise this complex topic. A voiceover was preferred to text layered over an animation because participants felt that it would be difficult to engage with text and images simultaneously. An animation was seen to be more engaging than an infographic.

*“With the voiceover, the key information was being spoken about and you could watch the animation as well so it was easier to understand.” [Female, 41+, ABC1, England]*

*“I like not having to read, I like that they’re talking at me. You can get bored reading but a video makes it more understandable.” [More likely to opt out, 18-35, AB, England]*

*“There’s a lot of reading. You kind of zone out and don’t really read it. There’s more chance of me listening to a video than reading a visual thing like this.” [Female, 18-40, C2DE, Northern Ireland]*

Participants felt that it was important for a video to convey its points quickly without becoming too lengthy, and to retain a sense of authority and seriousness despite being an animation:

*“Maybe it’s because of my age, but that was way too long for me. I feel like I’m so used to TikTok, that was too slow. My instinct was to go to the fast forward button. What I did like was the diagrams. It really helped with understanding. I felt a lot more secure in that.” [More likely to opt out: Black Ethnic minority background, England]*

Participants liked the idea of a researcher or expert appearing in the video, as illustrated by both the Open SAFELY video and the video describing SDEs.<sup>11</sup> They felt that this approach gave the topic a human face and a sense of credibility. Some participants also preferred voiceovers which addressed them personally i.e. by using phrases such as ‘your data’ and ‘you’:

*“I liked the speaker in this one. He was informative and authoritative stood in a laboratory.” [Mixed gender, 35-54, C1C2, England]*

Some experts could see the resource being used by individual TREs or the NHS SDE network as a core resource to describe the principles of TREs and SDEs. This could then be accompanied by case studies for specific environments, or specific lists of those have accessed specific SDEs/TREs and the reasons why. Experts pointed to some SDEs where this is done currently:

*“This could be used by any researcher looking to communicate a TRE to the public. It would need some awareness raising though as I’m not sure everyone is aware of UPD.” [Expert]*

*“This is a resource that’s kind of the foundation and then you have people who can do sort of talking head pieces on top of it that talk about the specifics.” [Expert]*

Another suggestion included providing a link to the resource on the NHS app for patients to access if invited to participate in research or clinical trials where an SDE may store their data:

---

<sup>11</sup> Open Safely: [OpenSAFELY - YouTube](#) SDEs: [The Data Jigsaw: Secure Data Environments \(youtube.com\)](#).

*“In England, most people have the NHS app so you could provide a link. There is a lot of promotion of being involved in research on there.” [Expert]*

Experts felt that the resource, once produced, could have three core benefits:

- Reducing the workload of individual TREs and SDEs who would not then need to produce their own resource or could focus on specific content relevant to them.
- Consistency across SDEs and TREs in terms of language and messages used to communicate them to the public:

*“This will just create a level of consistency and cohesion in the system that is currently completely lacking.” [Expert]*

- A publicly available resource that could be used where public consultations are taking place on SDEs and TREs:

*“I've used their (UPD) material before when I've been doing things like patient participation forums. I'm hoping it would be another resource that's on their website that people like me can download.” [Expert]*

## 8. Conclusions and recommendations

---

The following recommendations are based on insight gleaned from this research and have been divided into two groups; recommendations relevant only for the umbrella resource created by UPD, and recommendations that would apply to any resource describing a Secure Data Environment (SDE) or a Trusted Research Environment (TRE).

### Conclusions and recommendations specific for an umbrella resource designed by UPD:

- A core UPD resource describing SDEs and TREs to members of the public will need to be sufficiently broad to cover technical differences that exist between each SDE and TRE. This research suggests the resource should focus on SDEs and TREs for secondary uses only (rather than direct care) and include two core points; that the data is de-identified before it is accessed, and that it is accessed and not shared.
- The landscape of SDEs and TREs is evolving as is the communication around these environments. Any umbrella resource may need to evolve over time.
- The tone of voice should strike a balance between informing the public without seeking to reassure too much.
- The resource could mention SDEs and TREs if the language makes clear they are the same e.g. 'Secure Data Environments, sometimes known as Trusted Research Environments'. If they appear to be different this creates confusion for the public.
- An animated video with a voiceover is the preferred format for the main resource. The combination of visuals and a voiceover is likely to be beneficial to explain and visualise this complex topic.
- Experts suggested this core, umbrella, resource could be used by individual SDEs or TREs to describe the principles, complemented by specific detail individual to that TRE/SDE.

### Conclusions and recommendations relevant for anyone describing SDEs and TREs to the public:

- Providing clear, simple, information that is tailored to the public is likely to be better received than offering complex, detailed, or technical descriptions. Finding a balance between transparency and clarity for the public is challenging.
- Introducing the use of large datasets is likely to be beneficial to set the context for public participants who may not have considered the use of their data outside of direct care.
- The most important aspect of SDEs and TREs for public participants was the fact that the data accessed cannot individually identify them, so this should be made clear in any resource, where it is correct. Participants sought a simplified visual representation of data entering the environment, identifiable information being replaced with a number, and graphs and statistics coming out.
- Conveying the point about data access, not data sharing, was challenging and participants became confused by many of the phrases used. A simple description such as 'This data stays within the environment and cannot be sent out or downloaded' would be welcomed.
- The topic of who accesses the data was of great interest to participants and a transparent list of examples of organisations or individuals who can access the data is likely to be well

received. Detail on how approval to access the data is granted could be provided as extra information.

- Public participants tended to want to hear what the 'public' benefit of accessing the data in a SDE or TRE is. Listing possible use cases worked well here such as 'identifying better ways to diagnose and treat illnesses'. Experts however cautioned that this does not stray into appearing to 'sell' the benefits; particularly for a UPD umbrella resource.
- This research suggested some confusion over two of the five safes: 'Safe outputs' where participants were confused why are they checking for identifiable information when that has been removed, and 'safe settings' where the phrase 'remote connection' was not always understood and caused concern.
- Public participants sought information on cyber security in SDEs and TREs. The public tended to think about the security of their data in terms of the security of the storage facility, rather than in terms of the flow of the data and access controls.
- Providing information in layers including a 1-line description, a main resource, and extra links to further information will ensure that a resource meets the varying information needs of the public.
- The name Secure Data Environment was preferred to Trusted Research Environment. Participants felt that the word 'secure' conveyed an important point about the environments and sounded more factual.

## 9. Acknowledgements

---

The authors would like to thank all the members of the public who participated in this research as well as the experts who gave up their valuable time to contribute.

We would like to acknowledge contributions from the following experts and stakeholders, in alphabetical order. These experts and stakeholders have given their consent to be named individually as contributors to this research.

### Named individuals.

Dr Alison Knight, Data and Privacy Specialist, Health Research Authority.

Amy Darlington, Executive Director, Imperial College Health Partners.

Dr Catherine Kelly, Clinical Advisor, Digital Clinical Informatics, NHS England and Chief Clinical Information Officer, NHS Borders.

Doreen Tembo, Head of Public Involvement and Engagement at Health Data Research UK (HDR UK)

Emily Jesper-Mir and Rebecca Asher, Partnerships & Engagement Managers, Digital Equity, Wellcome.

Joe Watts, Head of Data Access and Linking Policy, Department of Health and Social Care / NHS Transformation Directorate.

Lorna Branton, working in NHS data strategy and communications, NHS England.

Lucy Knight, Policy Associate, NHS Confederation.

Dr Natalie Banner, Director of Ethics, Genomics England.

Natalie Fitzpatrick, Principal Research Fellow, UCL Institute of Health Informatics.

Rebecca Cosgriff, Deputy Director: Data for R&D, NHS Transformation Directorate, NHS England and NHS Improvement.

Dr Richard Lee, The Royal Marsden.

William Pett, Head of Policy, Public Affairs & Research, Healthwatch England.

We would also like to acknowledge contributions from the remaining experts and stakeholders who chose to remain anonymous.