

# Stimulus Materials

Workshops for Stage A

Final

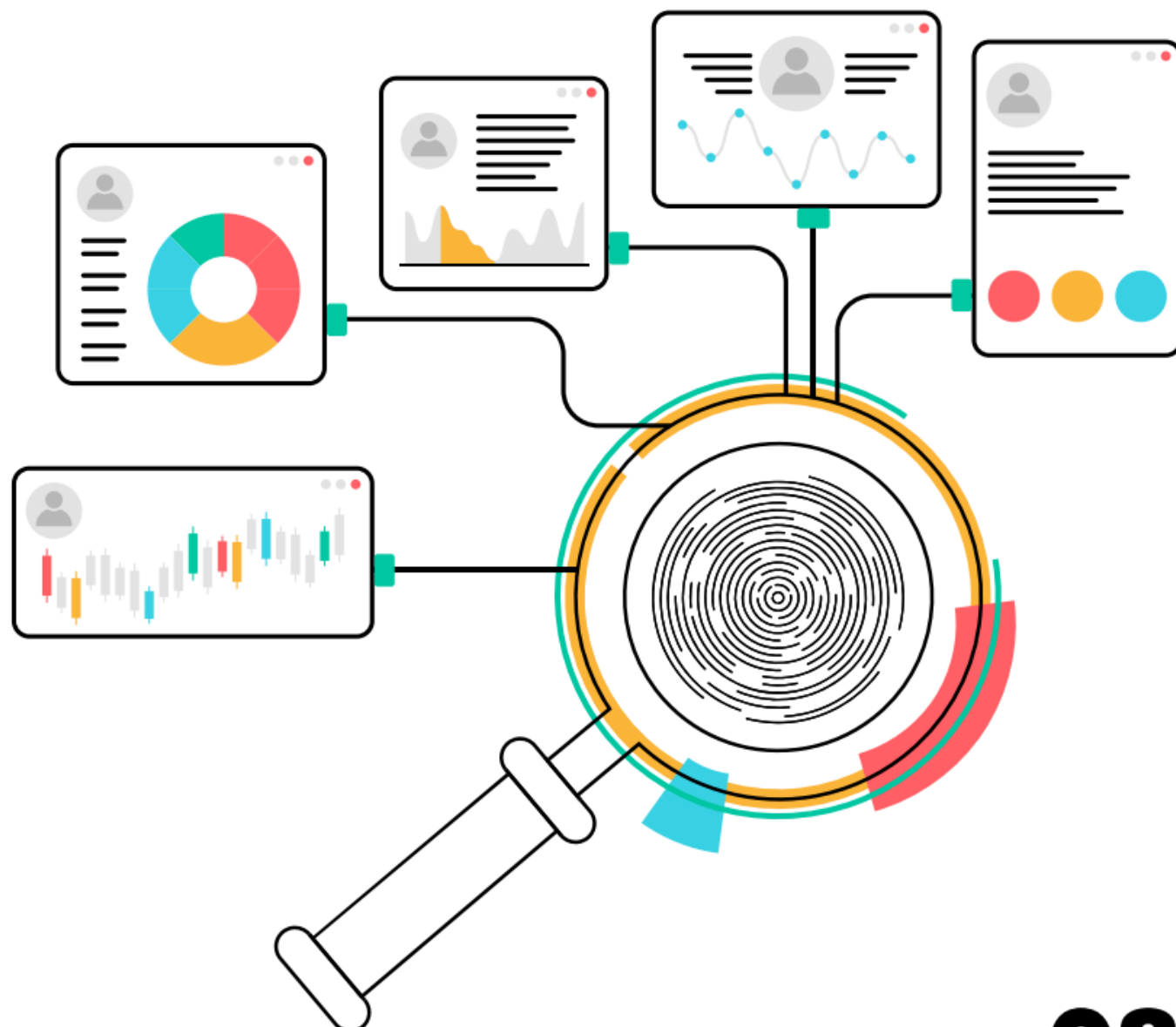
Amended after first 4 workshops

# WHAT ARE LARGE DATASETS?

Large datasets are created when the health data of many people is gathered together. When it is stored digitally and organised properly, huge amounts of data gathered from large numbers of patients can be analysed.

This analysis can be used to improve the understanding of diseases and disability, and to develop new treatments and technologies. It can also be used to plan healthcare services for all our future needs.

When health data is used in this way – for research and planning – it is known as secondary or indirect use of data.



# WHY ARE LARGE DATASETS CREATED?

**The potential benefits of large datasets include:**



## ***Identifying better ways to predict and diagnose illness***

When medical staff have good evidence – which can be provided by analysing lots of data – they can use it to recommend the most effective care for patients.

- Having data about large numbers of patients means that better predictions can be made. For example, predictions about health risks for specific groups or even a whole population.
- Large datasets may also allow rarer diseases to be researched. Diseases that are very unlikely to occur in a small group of patients may be seen in data from large populations.



## ***Developing new treatments, and monitoring the safety of existing treatments***

Analysing large datasets can help researchers see both the short-term and the long-term effects of treatments. Some of this work will be done by academic researchers or scientists within the NHS, but some will be done by commercial organisations such as pharmaceutical companies.



## ***Planning services***

When the NHS knows how many people have a particular health condition, and what their care needs are over time, data can be used to plan which services they will need in the future.

Data on healthcare can also reveal how resources might be better used or to plan for extreme events such as pandemics. This information can improve responses and help reduce costs across the health service.



## ***Addressing health inequalities***

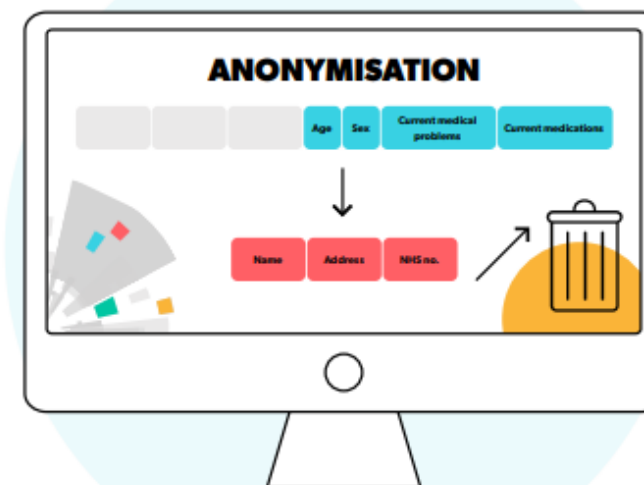
When health institutions and governments have data that covers the whole population, they can detect areas or groups of people that have worse health outcomes and target resources towards those who need them most.

## HOW IS MY PRIVACY PROTECTED?

### How de-identifying data can help preserve your privacy

De-identifying health data involves removing – or disguising – your personal information so that it is difficult to single you out of the dataset.

There are many ways of doing this. The method – or methods – chosen will dictate how well your private information is protected.



### Anonymisation

Data is considered to be 'anonymised' if all identifying information (such as your name, address or NHS number) is completely removed. This means that your identity cannot be re-linked to the data in a straightforward way.

Data that is treated in this way is no longer considered confidential and, generally speaking, does not fall within data protection laws.

It is important to note that even when health data is anonymised, it may still be possible to find ways of identifying your individual, personal information. However, it would likely require special circumstances or effort. This effort may involve using other sources of information to narrow down the number of individuals that the data may be referring to.

As a result, there is some debate about when data should be considered completely 'anonymous'.

# **What are Trusted Research Environments (TREs), also known as Secure Data Environments (SDEs) or Data Safe Havens?**

A secure environment that provides access to health data for approved researchers.

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.

Secure Data Environments give approved users access to health data for analysis, without them needing to receive a copy. The organisation providing the environment can control many factors, including;

Who can be a user;

The data that users can access;

What users can do in the environment;

The findings of analysis that users can remove.

Traditionally, researchers need to download a dataset onto their computer to be able to use it for their analysis. But in a TRE the data remains in a secure location, and approved researchers access it remotely. Researchers cannot take individual level data out of the TRE; they can only export analysis results (such as tables and figures), and only after careful checks have been made. It's a bit like a 'reference library' — approved researchers use the 'reference library' to access the data they need for a specific research study, but the data itself doesn't actually leave the security of the 'library'.



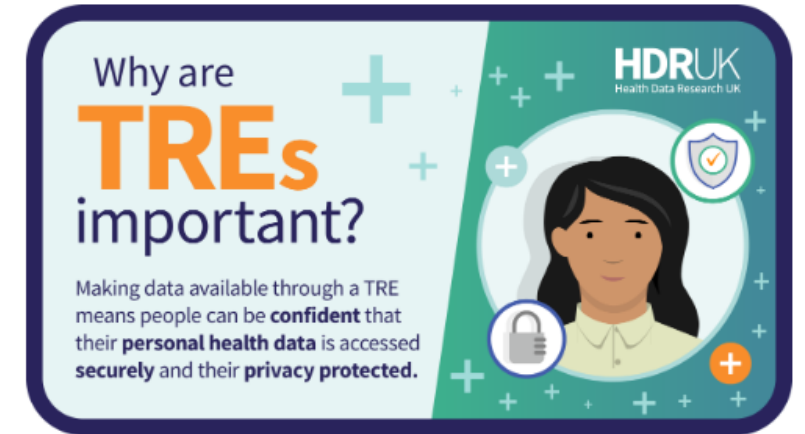
## TREs make research safer.



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.



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.



Making data available through a TRE means that people can be confident that their personal health data is or will be accessed securely and their privacy is always protected.

## STIMULUS 3b

Secure Data Environments will address the drawbacks of data sharing. They will improve:

**Patient privacy.** Secure Data Environments must apply techniques to remove personal details. This makes sure that patient information remains confidential.

**Security.** As NHS data will only be hosted on systems that can prove they have high levels of protection.

**Efficiency.** Secure Data Environments enable many different sources of data to be linked. This means that researchers and analysts can access bigger sets of data faster. This will speed up how quickly the NHS can make decisions and the discovery of new treatments.

## STIMULUS 0

Data Federation is like being able to read multiple books at the same time without needing to stack them all into one big pile. In healthcare, it means you can look at different types of health information from various places, all through a single "window," without needing to copy all that data into one place. That's what Data Federation does.

Some Secure Data Environments/ Trusted Research Environments use data federation.

## TREs operate in line with the 'Five Safes'

**Safe people:** Only approved researchers are granted access to data.

**Safe projects:** Data is only made accessible for projects in the public interest.

**Safe output:** All research findings are checked by TRE staff, with any potentially identifying information removed.

**Safe settings:** Data is accessed by researchers in a secure room or via an approved, remote connection to one.

**Safe data:** Data is de-identified, and only data that's really needed for the project is made accessible.



Link to further information about the five safes:

[The 'Five Safes' – Data Privacy at ONS | National Statistical](#)

# 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.

Learn more about TREs



## How is my data safeguarded?

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



## STIMULUS 5b

Video describing a Secure Data Environment:

<https://www.youtube.com/watch?v=TGf4ymZuoSs>

Trusted Research Environments (TREs)

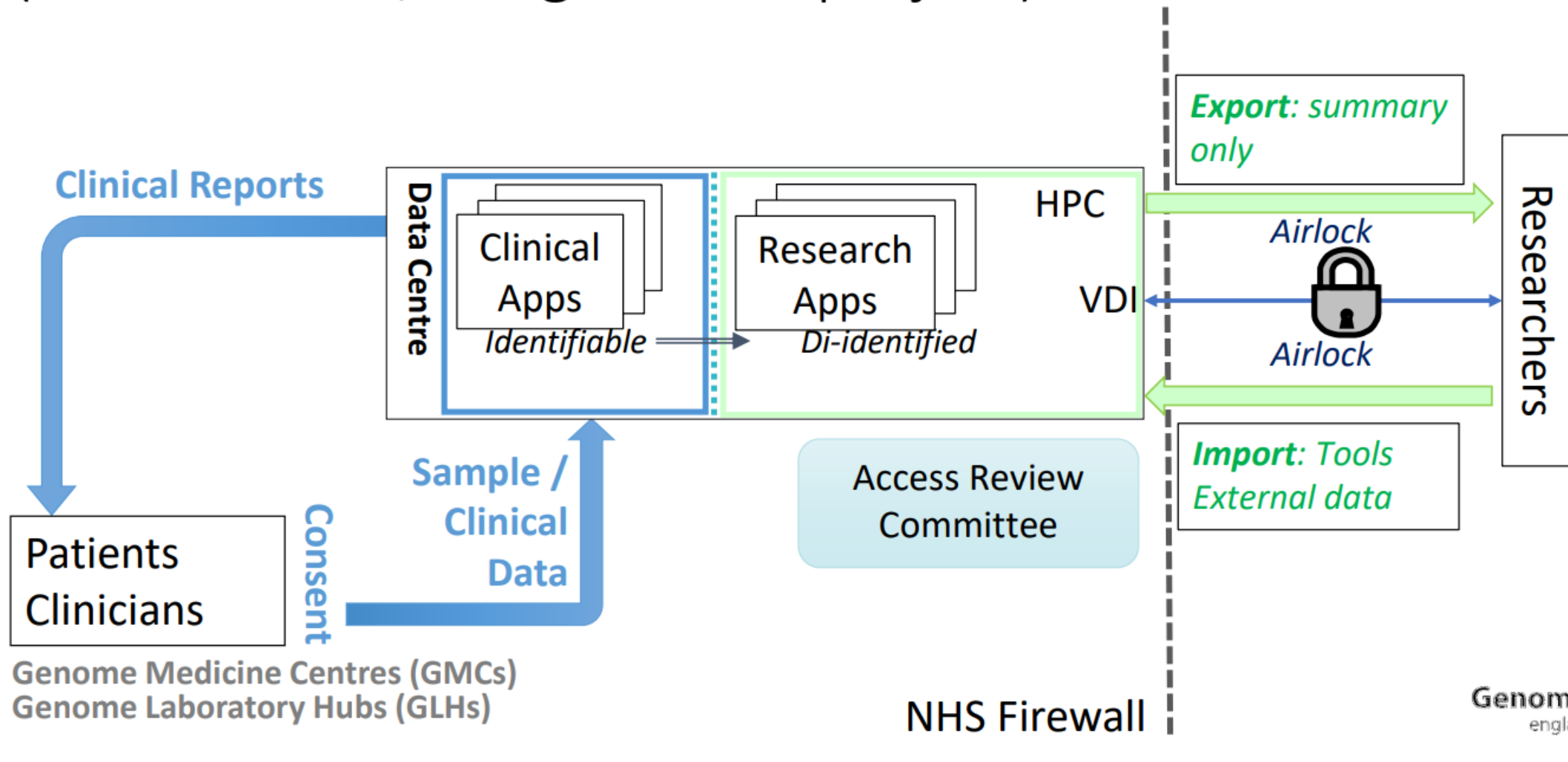
Secure Data Environments (SDEs)

Data Safe Havens

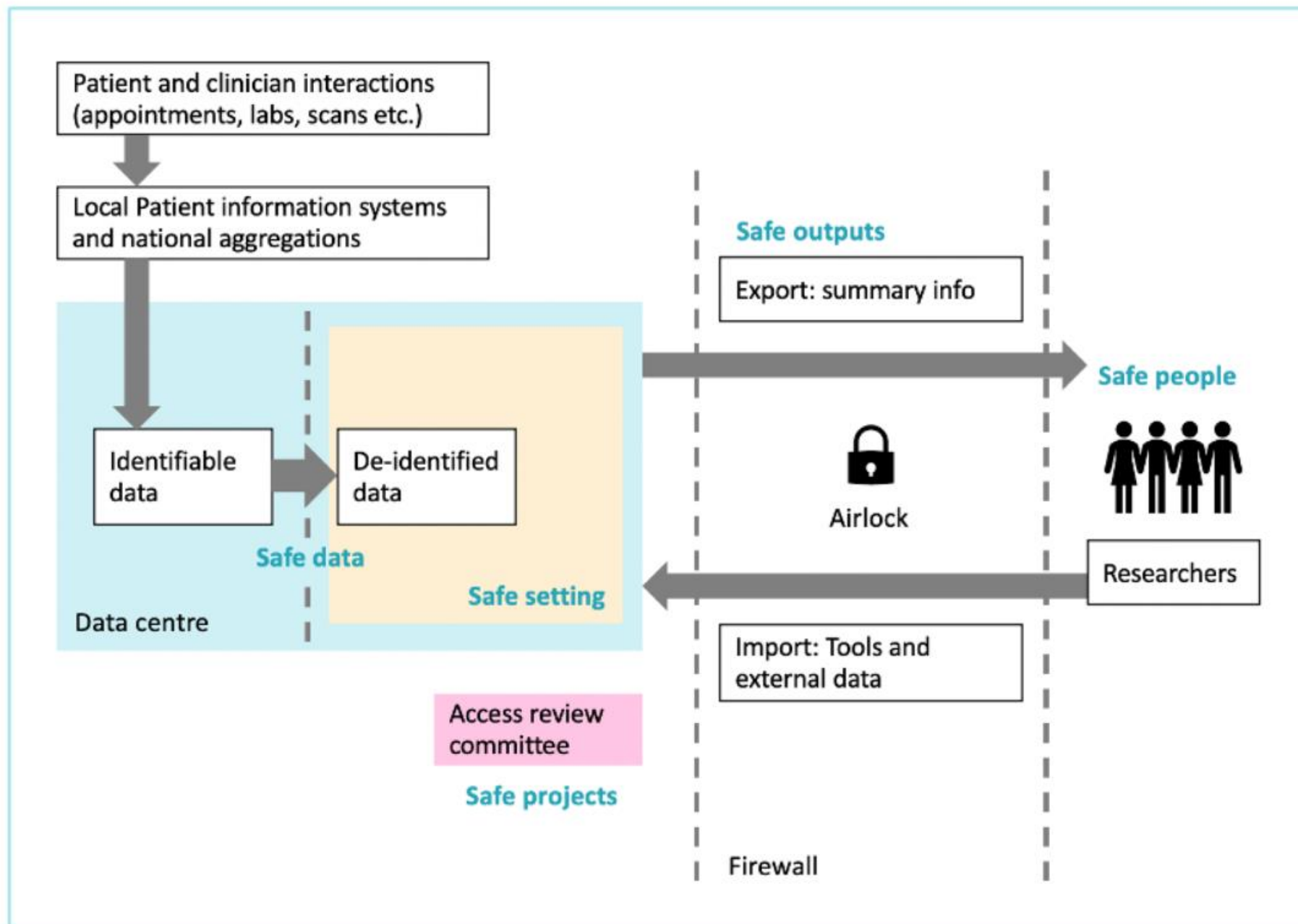
Controlled access data environments



# Example Trusted Research Environment (Used for 100,000 genomes project)



# STIMULUS 7c



STIMULUS 8a

Case studies

Name	Purpose	Can users see fully identifiable data?	Used by researchers outside of the NHS?
NHS England SDE.	Example: providing academics with access to cardiovascular and cancer data for Covid-19 research.	No	Yes
Public Health Scotland’s National Safe Haven	Provides access to health and administrative data.	No	Yes
HSC Northern Irelands’ Honest Broker Service TRE	Provides researchers with health and social care data on the population of Northern Ireland.	No	Yes
SAIL Databank (Wales)	Provides researchers with secure remote access to datasets with anonymised person-based health and social care data for the population of Wales.	No	Yes
ONS Secure Research Service	Gives accredited or approved researchers secure access to de-identified, unpublished data (including the Census) for research projects for the public good.	No	Yes
Genomics England Research Environment	Has over 2,000 researchers carrying out analysis with a range of tools in its high-performance computing environment, with genome data from the 1000,000 genomes project.	No	Yes

STIMULUS 8b

Case studies

Name	Purpose	Can users see fully identifiable data?	Used by researchers outside of the NHS?
OpenSAFELY	Secure analytics platform for NHS electronic health records of over 58 million people from England.	No	Yes
CO-CONNECT	Standardises antibody data collection across the UK and provides access to data for research on immunity to Covid-19.	No	Yes
UK Biobank	Starting to provide data on half a million study participants via the new research analysis platform developed by DNAnexus and Amazon Web Services.	Yes – participants have opted in	Yes
NHS Covid-19 Store	Sits on a Microsoft Azure platform under contract with NHS England and NHS Improvement. Within that secure cloud processing environment, Palantir manage their Foundry platform from which data and code do not leave.	No	No
NHS England Federated Data Platform	A ‘data platform’ which will enable NHS organisations to bring together operational data – currently stored in separate systems – to support staff to access the information they need in one safe and secure environment.	Yes, for purposes of direct care only	No