



Welcome to the use MY data webinar:

# Secret Garden or Open Landscape - Views on Trusted Research Environments

The webinar will begin shortly at 10:00

Please tweet with:

@useMYdata  
#datasaveslives

# An introduction to Trusted Research Environments

Secret Garden or Open Landscape – Views on Trusted Research Environments  
Use MY data Event

02/12/2020 | Tim Hubbard



## What is a Research Environment & why we use one?

- Computer system which is shared between multiple users
- Has commonly used tools pre-installed
- Has commonly used datasets pre-installed
- Generally can be accessed remotely
  
- Data may have been pre-processed, cleaned, structured, standardized
- Has mechanisms to import/install new tools and data
- May contain embedded High Performance Computing (HPC) facility

## Why use a Research Environment?

- Biomedical datasets have grown hugely in recent years
  - downloading expensive and difficult
  - costly to pay for multiple copies of data, ultimately cost for funders
- Developments in computing have made moving algorithms to data much easier
  - packaging of code as virtual machines, docker contains
  - use of code repositories
  - availability of virtualized computing infrastructure, cloud computing

## Why do we need trusted research environments?

- Health data research has historically been dominated by research cohorts and data distribution
  - Data is anonymised; volunteer research subjects accept risk of de-identification
  - Cohorts are small; tend not to represent the national population
  - Cohorts have narrow disease focus, limiting research on co-morbidities etc.
- Better to analysis whole populations from health systems, but public is nervous about risks
  - Well known example of distributed data being misused; individuals being reidentified, e.g. social media
  - TRE concept avoids data distribution; supported at OneLondon citizen's deliberation
- Holding copies of "personal data" requires strong controls on access and security
  - can afford to invest more in security at single TRE than local systems holding distributed data
  - data distribution liability to academic institutions and companies if breach: fines under GDPR; reputational risk

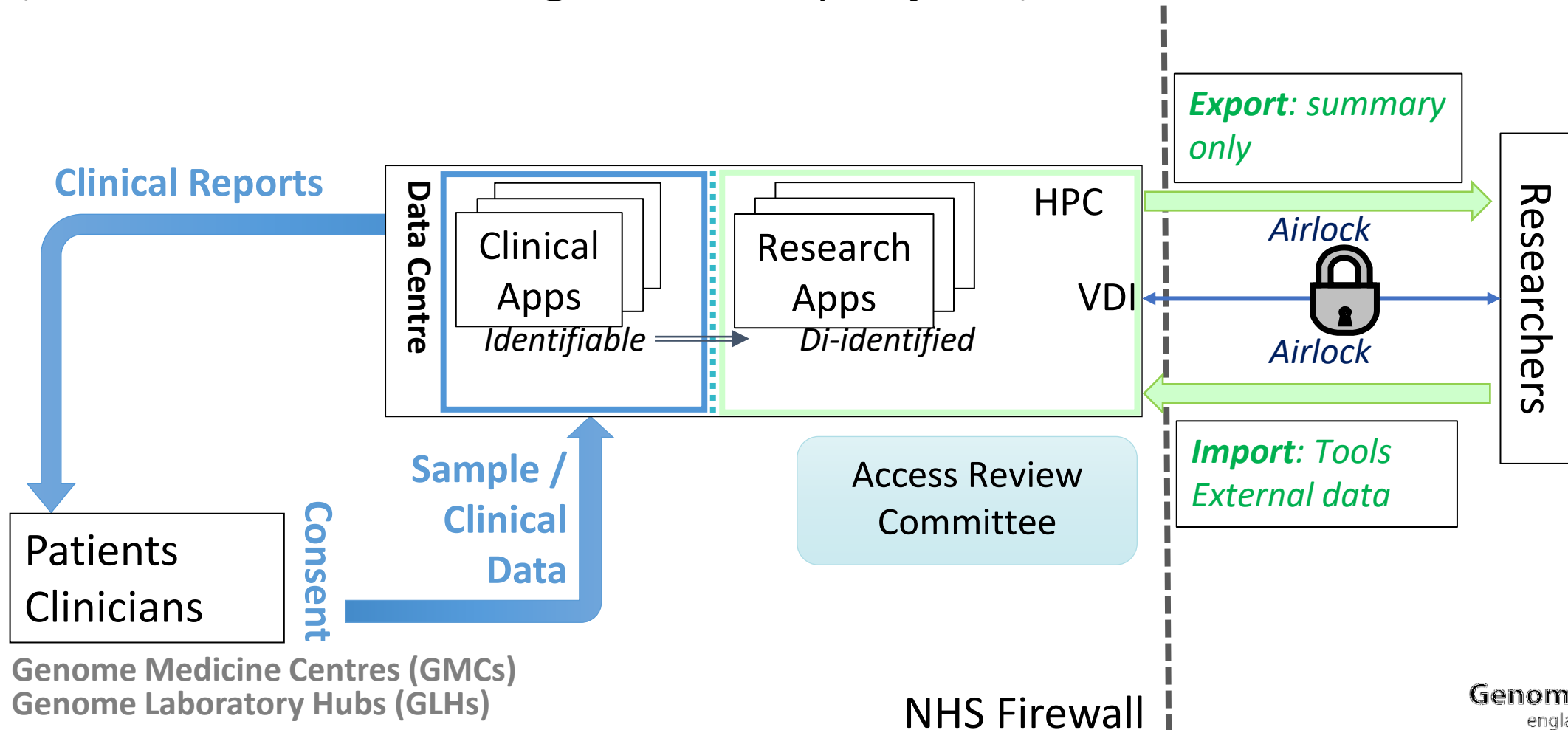
## What is a trusted research environment?

### A research environment that implements the 5 safes definition

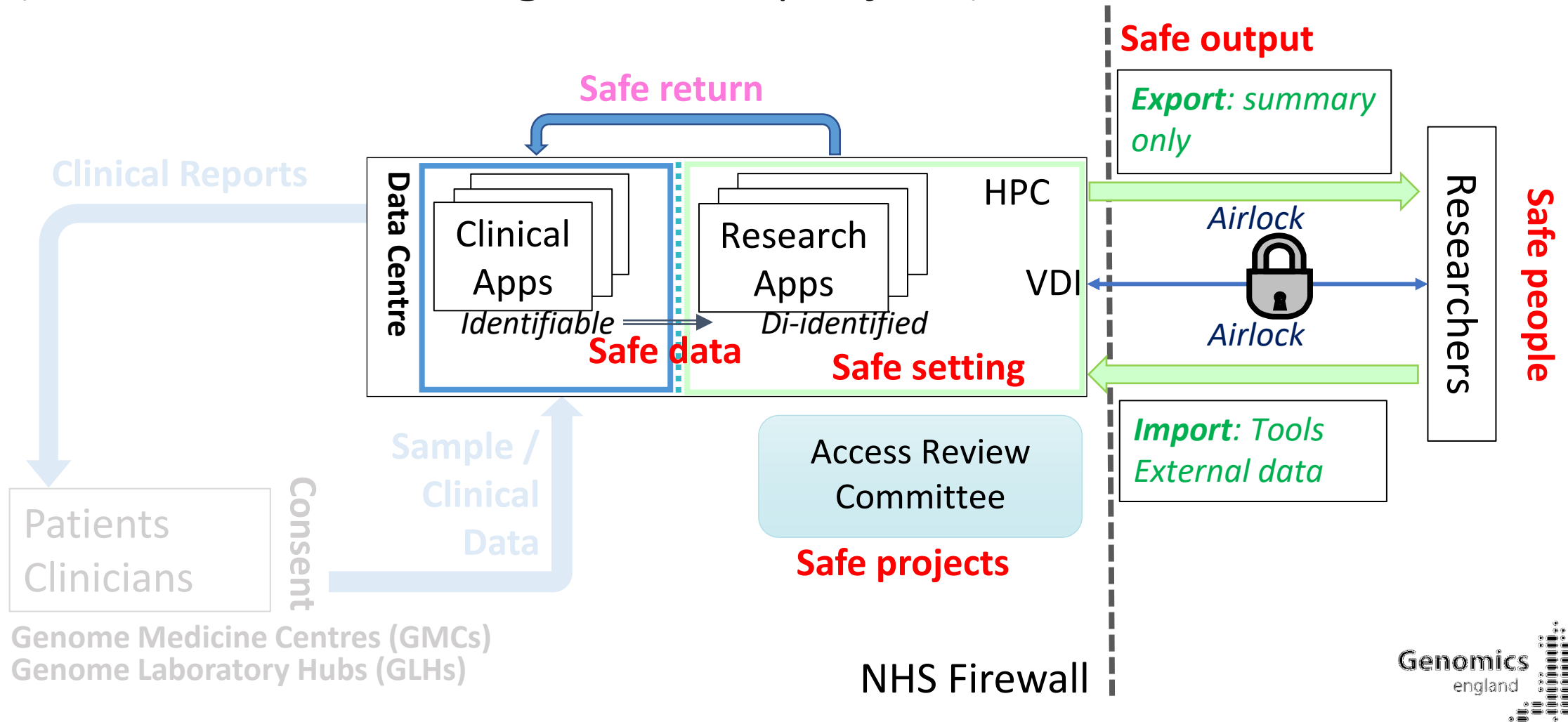
- Safe people
  - Technical skills to use the data; compliance with training requirements; signed confidentiality agreement
- Safe projects
  - Research project is appropriate, ethical, will benefit public; results will be published
- Safe data
  - Data within controlled environment is di-identified
- Safe settings
  - Researcher only as access to data within a controlled environment: no data distribution
- Safe outputs
  - Export of results for publication controlled to ensure confidentiality is maintained

<https://blog.ons.gov.uk/2017/01/27/the-five-safes-data-privacy-at-ons/>

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

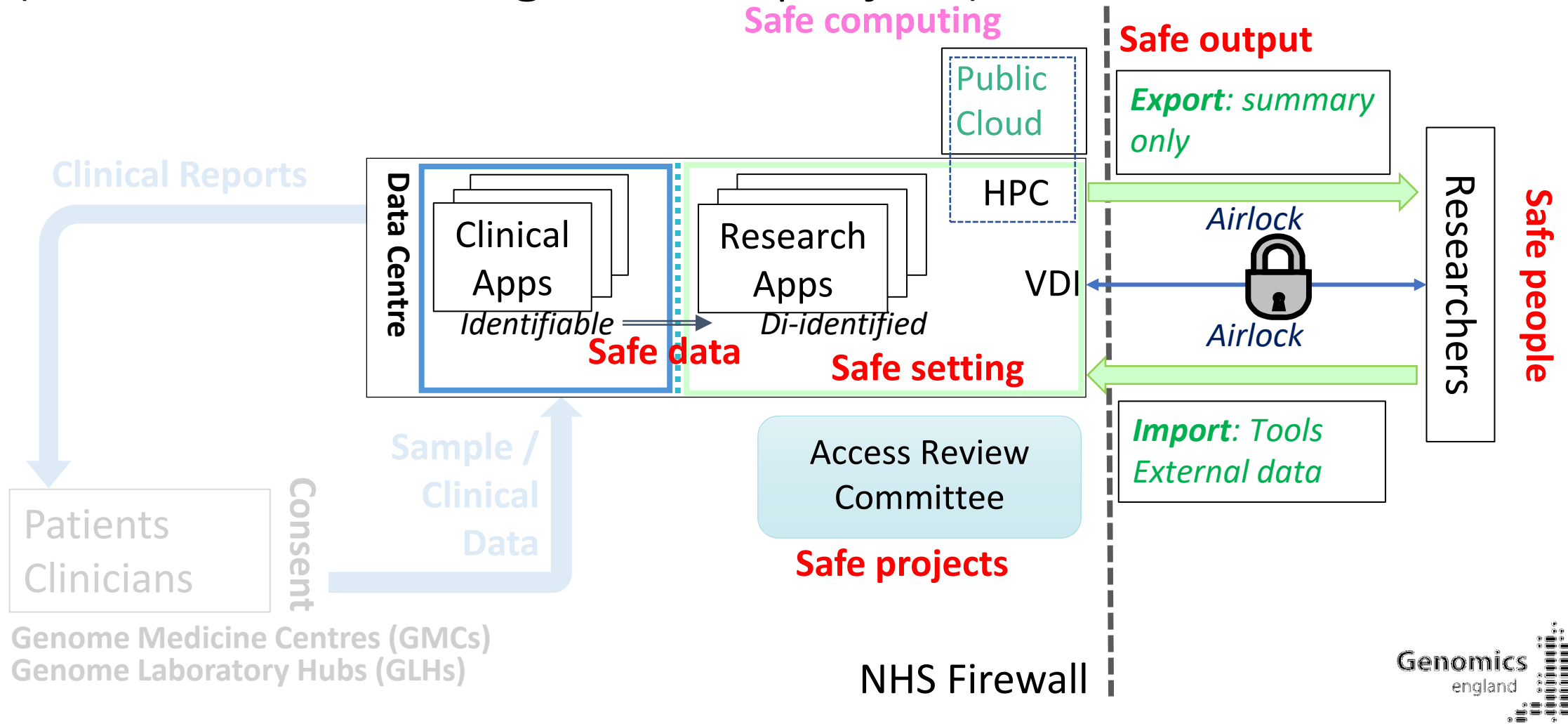


# Example Trusted Research Environment (Used for 100,000 genomes project)





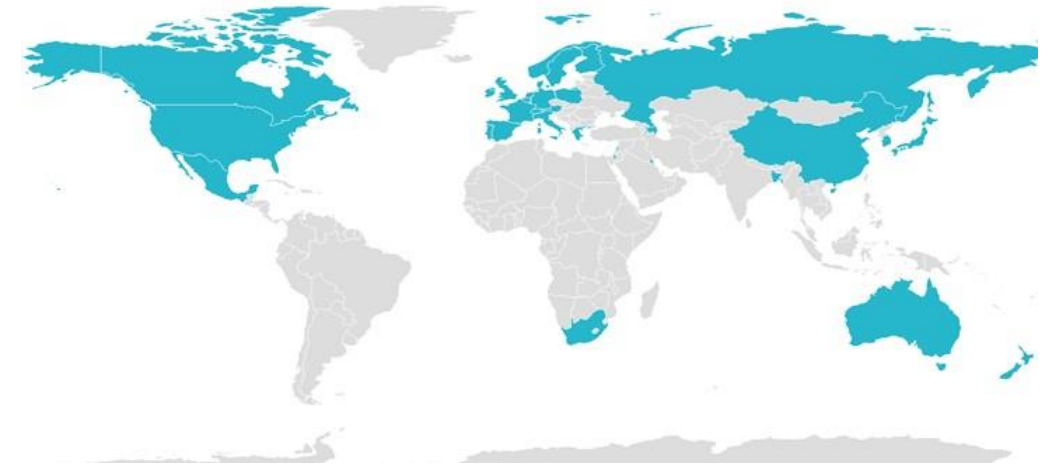
# Example Trusted Research Environment (Used for 100,000 genomes project)



# Genomics England Clinical Interpretation Partnership in numbers (As of 17<sup>th</sup> July 2020)



- 3,388 researchers world-wide
- 175 institutions with signed Participation Agreement
- 1,089 researchers with access to data
- 431 registered research projects  
<https://research.genomicsengland.co.uk/>
- 131 Project publications
- 24 Successful grant applications
- £46 million total awarded



A screenshot of the Genomics England Research Portal website. The page has a dark blue header with navigation links: Home, Research Registry, Useful links, News, Help, and Sign in. Below the header is a teal banner with the Genomics England logo and the text "Research Portal". The main content area features a large heading "Welcome to the Genomics England Research Portal" and a paragraph: "From here you can track your application to join the GeCIP or access the 100,000 Genomes Project data. As a GeCIP or Discovery Forum member you will be able to browse and submit projects in the Research". To the right, there is a "Quick links" section with three items: "Apply to join GeCIP", "Research Environment user guide", and "Change/Reset your password".

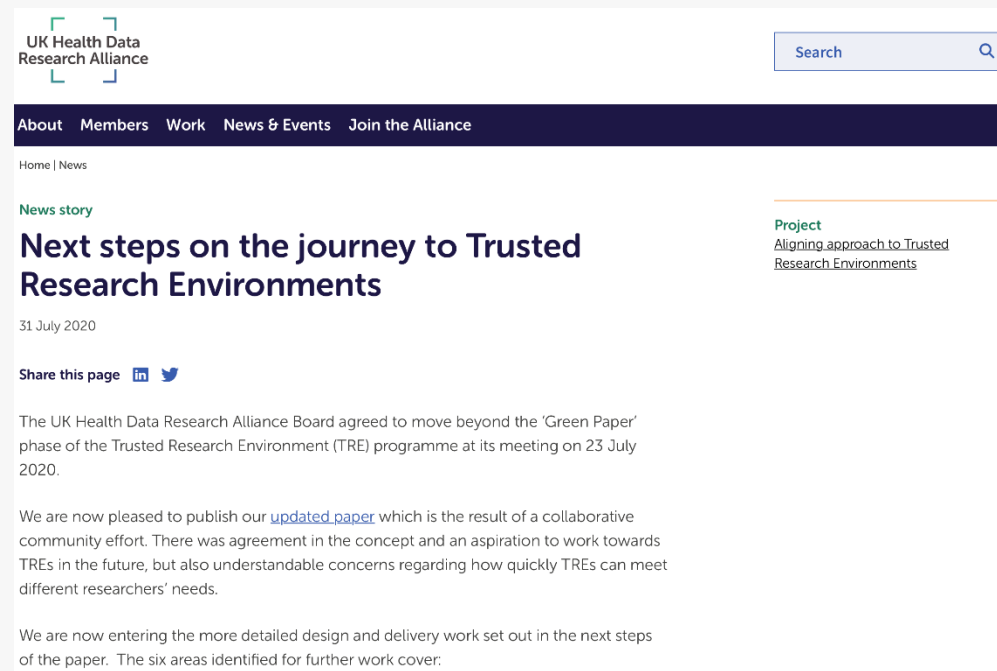
# Trusted Research Environment (TRE) Green Paper

Draft updated following consultation with stakeholders

UK HDR Alliance board approved next phase of work

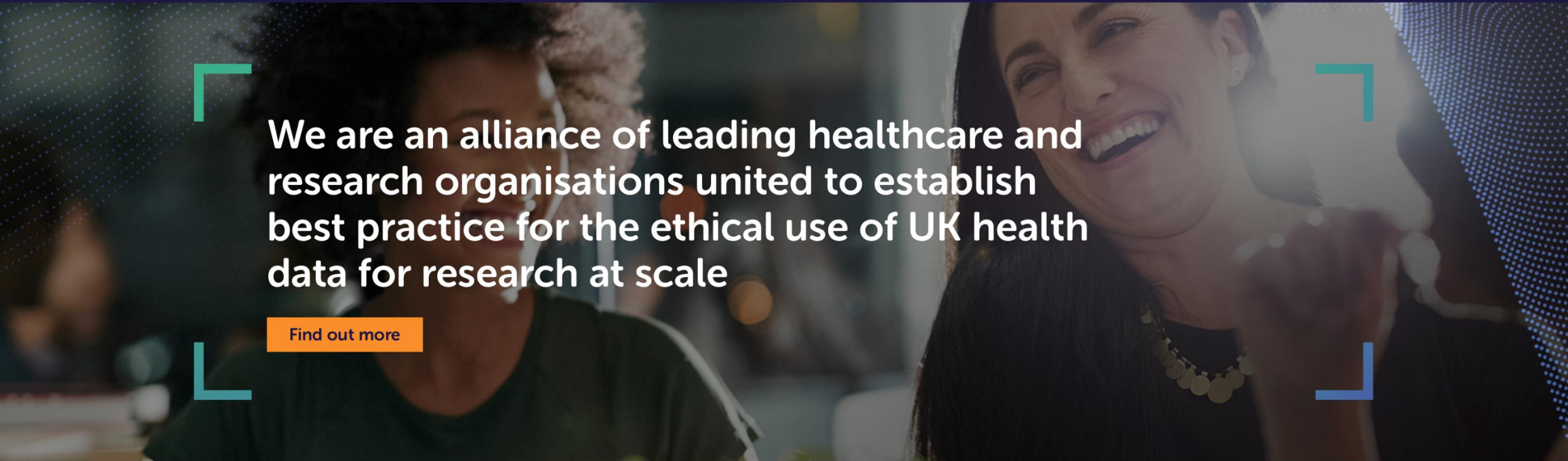
Six areas identified:

- Consistent and proportionate accreditation of safe people.
- Consistent accreditation of safe settings, with a focus on use of cloud computing.
- Involvement of public and patient representatives in the data access decision making process.
- Improved lay explanations of the design and functioning of TREs.
- Enhancing the researcher experience whilst minimising risks to privacy.
- Addressing the technical, governance and process challenges of creating a network of TREs through federation.



The screenshot shows a news article on the UK Health Data Research Alliance website. The page title is "Next steps on the journey to Trusted Research Environments". The article is dated 31 July 2020 and includes social media sharing options for LinkedIn and Twitter. The text of the article states that the UK Health Data Research Alliance Board agreed to move beyond the 'Green Paper' phase of the Trusted Research Environment (TRE) programme at its meeting on 23 July 2020. It mentions the publication of an updated paper, which is the result of a collaborative community effort, and notes that there was agreement on the concept and an aspiration to work towards TREs in the future, while also addressing understandable concerns regarding how quickly TREs can meet different researchers' needs. The article concludes by stating that the organization is now entering the more detailed design and delivery work set out in the next steps of the paper, with six areas identified for further work.

<https://ukhealthdata.org/news/next-steps-on-the-journey-to-trusted-research-environments/>



We are an alliance of leading healthcare and research organisations united to establish best practice for the ethical use of UK health data for research at scale

[Find out more](#)

## What is the UK Health Data Research Alliance?

By combining expertise and a shared commitment to work collaboratively, the Alliance helps researchers to answer some of the most difficult questions and address the most important health challenges faced in the UK. The Alliance is convened by [Health Data Research UK](#).

<https://ukhealthdata.org/>



## Members



Barts Health NHS Trust



Nottingham University  
Hospitals NHS Trust



University Hospitals  
Birmingham NHS  
Foundation Trust



Clinical Practice  
Research Datalink



Cystic Fibrosis Trust



The Brain Tumour  
Charity



NHS Digital



NHS England



NHS Scotland



NHS Wales



Health and Social Care  
Northern Ireland



NHSX



Genomics England



UK Biobank



National Institute for  
Health Research (NIHR)  
Bioresource



Public Health England



HQIP



Royal College of  
General Practitioners  
(RCGP)

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

← TRUSTWORTHY USE OF DATA →



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

**UK Health Data Research Alliance**

- NHS DATA
- COHORT DATA
- REGISTRY DATA
- TRIALS DATA

Convened by

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

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

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

Operated by



↑ PATIENT AND PUBLIC BENEFIT ←

# UK HDR Alliance Gateway: Standardised, Searchable Metadata Catalog of resources

**COVID-19** Health Data Research UK is mobilising teams across the UK and championing the use of data to respond to COVID-19

[Read more](#)

Health Data Research  
Innovation Gateway

[About](#) [Community](#) [Dashboard](#) [COVID-19](#) [Collections](#) [Sign in](#) | [Sign up](#)

Explore datasets, tools and resources used in health research across the UK

Search...

[All Datasets](#) [All tools](#) [All Courses](#) [COVID-19](#) [Cancer](#)

**BETA** This is a new service. Your feedback will help us improve it.

**354**  
registered users

**504**  
datasets

**107**  
access requests

**48**  
tools

**6,936**  
searches in last month

# The Gateway is enabling a federated, collaborative and trustworthy approach to health data research at scale: 30+ data custodians, 4 nations, 500 datasets, ~60m population, used by 100s of projects



## Innovation Gateway

Research Questions

Dataset Discovery

Access Requests

Accessed in TREs

Impact

Enables researchers to coalesce on key questions

~500 datasets, four nations, covering multiple diseases

Streamlined and transparent access management

Linked to a federated network of TREs

Discoveries that are improving people's lives

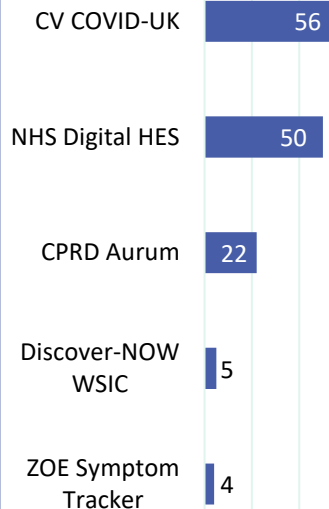
Examples

**111** research questions reviewed as part of COVID-19 prioritisation pipeline

Industry supported by Hubs include:

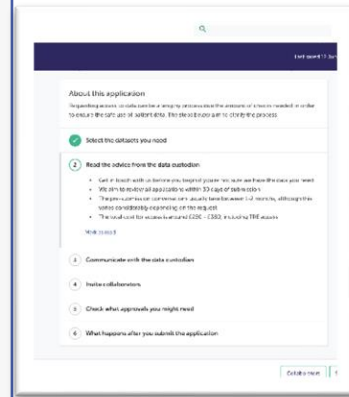


Example Gateway datasets (population, millions)

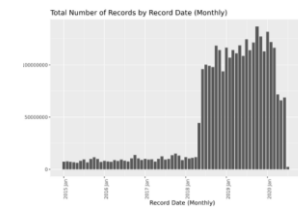


Harmonised access request:

- developed with input from **+20** orgs
- trial underway with **7** custodians



**158** active projects on COVID within TREs  
 BHF Data Science Centre leading the COVID-CVD consortium with **17** researchers across **6** institutions undertaking research on **4b** records (**56.5m** patients)



Over **140** projects underway across Hubs – with large industry, SMEs, charities, the NHS and research organisations



Underpinned and strengthened through continuous patient and public involvement and engagement



# Acknowledgements

- 100,000 genomes project patients and their families, Genomics England Team, NHS England, Health Education England, Public Health England
- Health Data Research UK
- HDR London site partners: Imperial College London, King's College London, London School of Hygiene and Tropical Medicine, Queen Mary's University London & University College London and
- King's College London, Wellcome Sanger Institute, European Bioinformatics Institute

# Secure Anonymised Information Linkage Databank

**Kerina Jones**

Use MY Data Webinar on Trusted Research Environments  
25<sup>th</sup> November 2020

# SAIL Databank

## Established in 2007

### Why was SAIL established?

- Vast amounts of routinely-collected data – untapped potential
- Making the data accessible safely for research
- Linking datasets from different sources – otherwise in silos
- Open up ways to answer important questions that could not otherwise be addressed or without prohibitive effort and cost

### Who funds SAIL?

- Health & Care Research Wales
- Economic & Social Research Council

### How does SAIL provide access to data?

- Within a secure virtual environment subject to procedural, technical, and physical controls
- In combination with disclosure controls applied to the data
- Privacy by design

### What sort of data does SAIL have?

# What data does SAIL hold?

Core Datasets

Core Restricted Datasets

Core restricted datasets **do** require permission from the data providers to use them in addition to the standard application process of IGRP approval and completion of the safe researcher training.

Active Adult Survey

Bowel Screening Wales

Breast Test Wales

Cervical Screening Wales

Congenital Anomaly Register and  
Information Service for Wales (CARIS)

Education Attainment

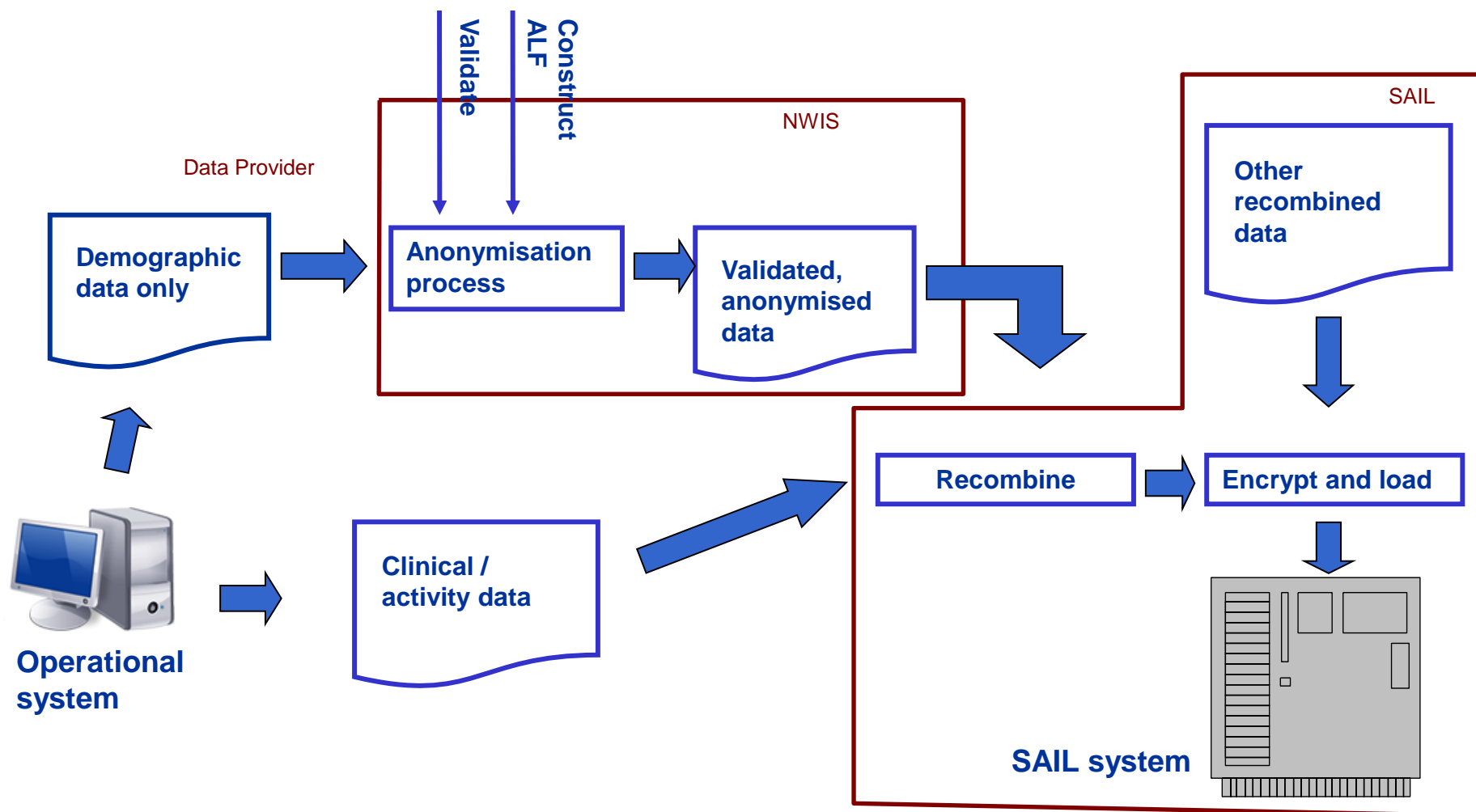
National Survey for Wales

Welsh Cancer Intelligence  
Surveillance Unit (WCISU)

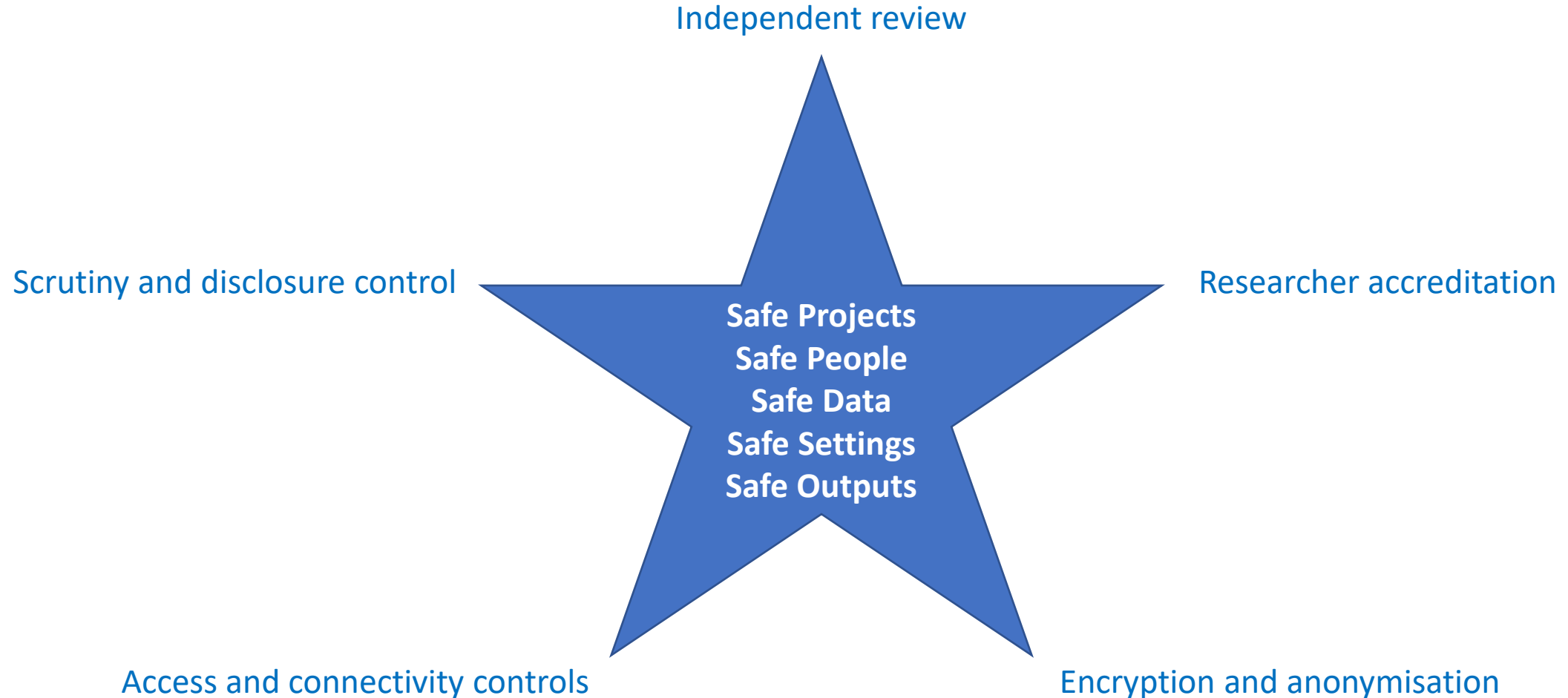
Welsh Health Survey



# How does SAIL link without identifiable data?



# How does SAIL enact privacy by design?



**Combination of physical, technical and procedural controls applied to the data and the environment**

# Public involvement & engagement

## Mission statement

- To act upon our commitment to Public Involvement & Engagement by implementing a strategy that promotes public inclusivity, that is proportionate and sustainable, and that demonstrates respect and trustworthiness towards the public, enabling their voices to be heard and included.

## Consumer panel

- Public perspectives on data intensive research and on new developments
- Views on data protection issues
- Discuss proposals and research findings
- Membership of Information Governance Review Panel
- Review information designed for a lay audience
- How best to engage with the public/group



*SAIL Consumer Panel*

## The wider public and particular groups

- Events, surveys, workshops, etc.
- Research projects

<https://saildatabank.com/wp-content/uploads/200416-Public-Involvement-Engagement-Policy.pdf>

# Challenges and developments

- **Free-text data**
  - Public views on safeguards in the use of free-text data such as from clinical letters and notes
- **Genetic data**
  - Public preferences on access models for genetic data linked to health and administrative data
- **Connected devices**
  - Public views, knowledge, benefits and concerns on the use of device data for research
- **Adoptions data**
  - Working with the adoptees, adopters, birth parents and the general public on the use of data about adoptions with health and administrative data



Thank you for listening

Any questions?



# NHSD Trusted Research Environment (TRE) Service for England

## Overview

# What is the NHS Digital TRE Service for England and why are we delivering it?

## Our Vision:

- Building on NHSD's place in law as England's Safe Haven, we want to be the place to go for safe and timely access to the most comprehensive store of nationally collected, linkable, quality assured health and care data for secondary use... a one-stop, trusted health research service where researchers live and do great work, day-in, day out

## Our mission:

- ✓ *to provide approved researchers with rapid, safe and trustworthy access to essential health and care data*
- ✓ *Enabling timely research at scale... on outcomes associated with COVID-19\* (and other research) to guide national and international decision making, and recommend potential interventions to reduce the severity of COVID-19 outcomes*

The **TRE Service is strategically important** both in the immediate and longer term:

- Immediately to respond to COVID-19, as a service directly commissioned by SAGE\*
- In the longer term as an enduring service (post COVID-19) as NHSD **channel shifts** a number of our users of data, moving from a data dissemination model (only) ...to a safe access model – starting with research clients and research teams

\*One of 4 COVID-19 TRE's across the home nations

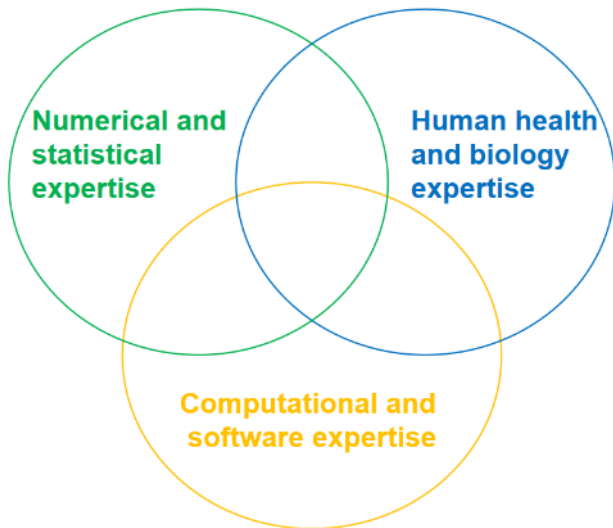


# Who receives the service? Some profiles

## Typical research users:

Clinical researchers, data researchers, data analysts, data scientists

coming from Universities [HE], clinical, research organisations and public bodies



Based on their professional backgrounds and roles, different users will come with **different blends of stats, computational and health and care domain expertise**

Their expertise and current roles in the Health, Care and Life Science sectors also underpins their respective **TRE service needs, data needs and tooling choices**

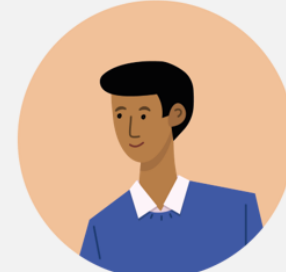
### THE MACHINE LEARNING SPECIALIST



**Beatriz**

Bioinformatician and cancer data scientist

### THE “DATA WRANGLING” EXPERT



**Lucien**

Data analyst in a public body insights team

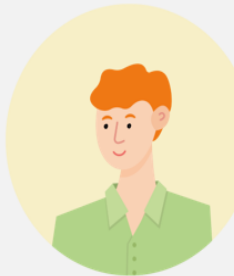
### THE “NUMBERS’ CRUNCHER”



**Tessa**

Statistical epidemiologist in a large university

### THE “TIME CRITICAL” INVESTIGATOR

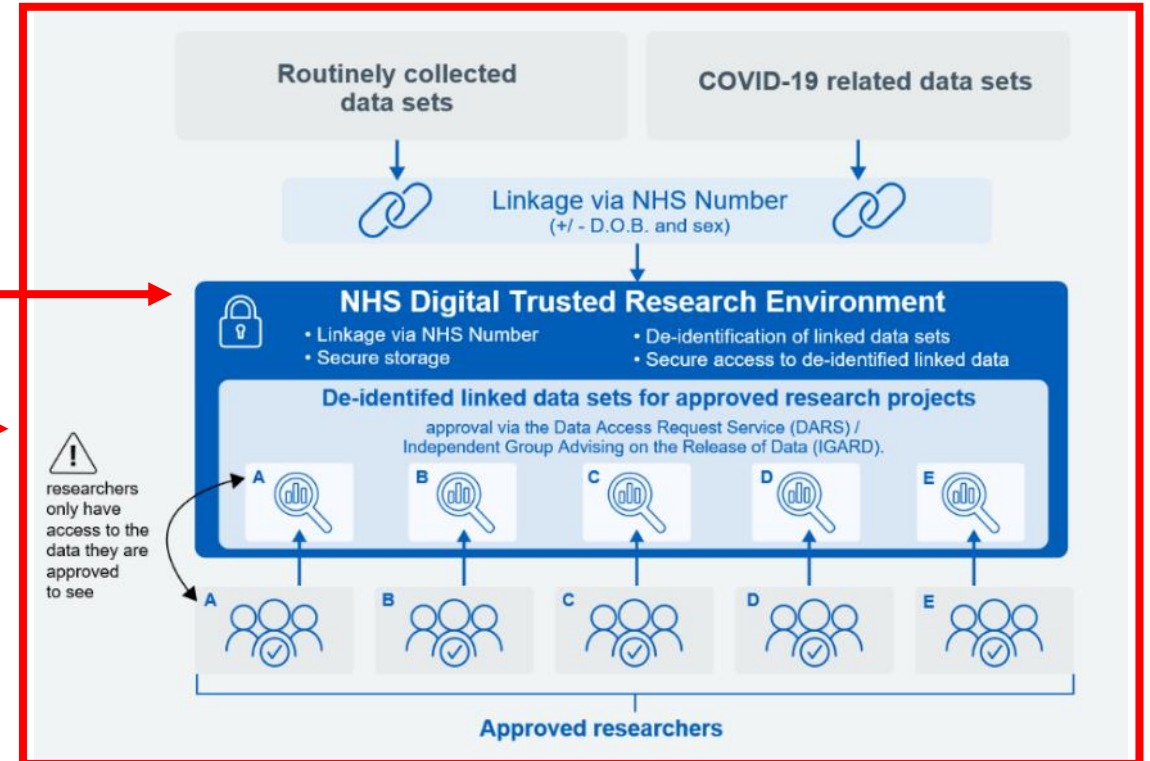


**Alexei**

Cardiologist and clinical researcher at an NHS trust

# The TRE Service and its Safe Setting... what's the difference?

- The TRE service we offer uses NHS Digital's new Data Platform ['DPS'] as its 'safe setting' and it's at the heart of the service offer
- But, there are other key common characteristics of TRE Services aside from a technical platform – Known as the '5 Safes framework' – more on those in a minute
- For now,
  - whenever you see the term 'NHSD TRE' - think: a *service* that uses DPS as its safe *setting*
  - And also note – in technology terms – no matter how we gear the service to meet different research needs of, say cancer research, cardio-vascular research or cross-cutting research across a number of conditions...
  - ... they are all **using a single system** – the DPS - that puts walls round particular service offerings and access rights



# The '5 Safes framework' and how we comply

**! All TREs MUST follow 5 safes !**

**5 Safes:** Safe Projects + Safe People + Safe Data + Safe Setting + Safe Outputs = **Safe use**



[Click here for more on 5 Safes](#)



- NHS Digital's 'Data Access Request Service' (DARS) ensures Data Sharing Agreements (DSAs) are in place
- These **DSAs** set out the purpose of the data sharing, what happens to the data at each stage, sets standards and helps all the parties to be clear about their respective roles.
- There's also an **Independent Panel** to review research project selection



- NHSD's Data Access Request Service also uses a **Data Security Protection Toolkit [DSPT]** to ensure researchers come from trusted organisations, have the right credentials and have undergone the right IG training



- Part of NHSD's 'safe technology' is a '**De-identification**' solution - **applied on all data that is shared to pseudonymise is**, greatly reducing the risk of patient/person re-identification
- Approved **Data Sharing Agreements** also carry stipulations on required '**data minimisation**' – **placing clear limits so that researchers only get access to the data they need [and no more]** to fulfil their agreed research purpose.



- NHSD's '**Secure Environment**' – **i.e. its safe setting** has multiple security layers, including two-factor authentication for safe user login; end-to-end encryption + is compliant with NHS Digital Cloud Security Guidance for Safe Compute



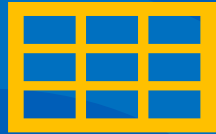
- **The NHSD TRE 'Safe Outputs team'** – 'Output checkers' use a Safe Outputs Policy and dataset statistical disclosure rules to check and approve research findings for publication – thus ensuring nothing leaves the safe setting that shouldn't

# What does NHS Digital's TRE Service offer?

## Day 1 - TRE service

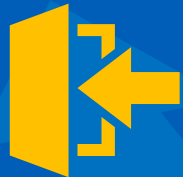


Remote Access &  
Log In



De-identified  
Data products

Analyse  
(e.g R Studio,  
Databricks)



Exit [or  
start  
over]



Safely publish  
(via Escrow)



Collaborate

The screenshot displays the NHS Digital interface, which includes a sidebar with navigation options: Home, Workspace, Recents, and Data. The main content area shows a Databricks workspace with a SQL query: `SELECT * FROM dss_corporate.org_daily`. Below the query, a table lists data rows with columns: ORG\_KEY, ORG\_CODE, ORG\_TYPE\_CODE, NAME, SHORT\_NAME, and ORG\_OPE. The table contains 5 rows of data. Below the table, there are options to 'Download preview' and 'Download full results (max. 1 million rows)'. The bottom part of the screenshot shows a file export list for 'export (1).csv' with columns for Name, Last modified, and Size. The file list includes 'export (1).csv' (59.0 MB), 'export (2).csv' (371.4 KB), 'export (3).csv' (55.7 MB), and 'export (4).csv' (61.8 MB).

ORG_KEY	ORG_CODE	ORG_TYPE_CODE	NAME	SHORT_NAME	ORG_OPE
1	5FEM8	PU	ELMLEIGH HOSPITAL	ELMLEIGH HOSPITAL	2009-10-0
2	5FEM9	PU	HANWAY GROUP PRACTICE	HANWAY GROUP PRACTICE	2009-10-0
3	5FEM0	PU	THE HEALTH HOUSE	THE HEALTH HOUSE	2009-10-0
4	5FEN0	PU	HEWAT HOUSE	HEWAT HOUSE	2009-10-0
5	5FEN1	PU	HEYWARD ROAD SURGERY	HEYWARD ROAD SURGERY	2009-10-0

Name	Last modified	Size
export (1).csv	May 22, 2020 8:26:05 PM GMT+0100	59.0 MB
export (2).csv	May 22, 2020 9:39:30 PM GMT+0100	371.4 KB
export (3).csv	May 22, 2020 9:40:01 PM GMT+0100	55.7 MB
export (4).csv	May 22, 2020 9:40:02 PM GMT+0100	61.8 MB

## Since April 2020

- **LIVE:** NHSD's initial 'TRE Service', rapidly created to address immediate CV-19 research needs

## Since 15 July 2020

- **LIVE:** First research customer up to access our more enduring, strategic service – a partnership with the BHF for research on Cardio-vascular disease
- With access to the **first datasets** in line with the approved of the Data Sharing Agreement



## Since 31 August 2020

- TRE service ready for business for Cancer researchers, with two projects coming down the track now\*
- **Nov '20:** Cancer data x 4 being loaded into the TRE ready for first projects



## When do things happen?

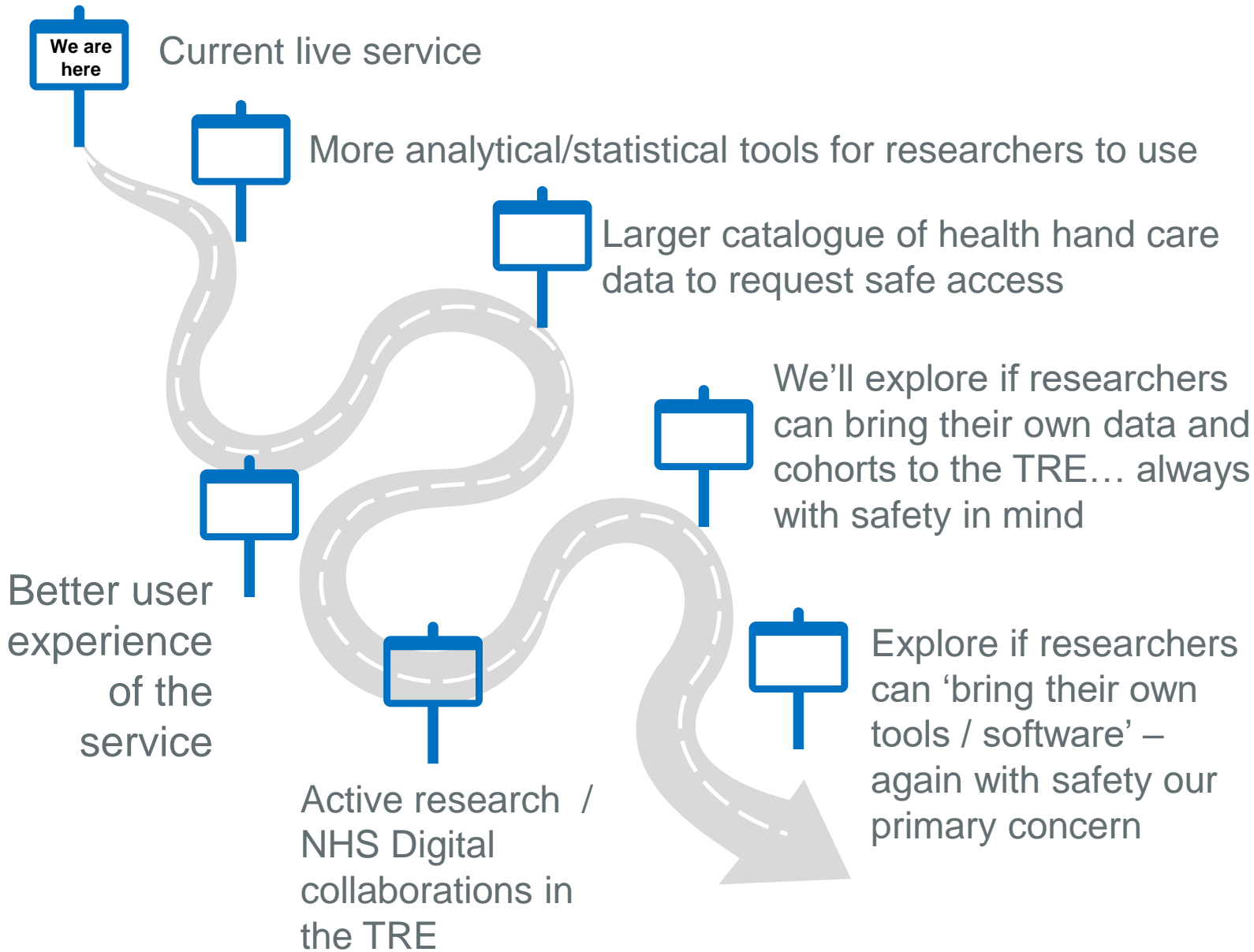


## Since September 2020, onwards

- More prospective research projects coming forward inc. research projects on COVID-19 National Core Studies [x 4], Infection Control, Biosecurity & Early Disease Detection



# Developing and enhancing the service



## Channel Shift



# How do researchers, patients and the public alike find out more, find answers to questions they have, or even get involved?

## General service information

- Service Information: <https://digital.nhs.uk/coronavirus/coronavirus-data-services-updates/trusted-research-environment-service-for-england>
- Access to service: researchers contact NHSD via 'DARS Online'
- Active user queries: via NHS Digital National Service Desk

## Wider engagement / getting involved

- How can people help? Great public and patient engagement and involvement is a watchword of a great service
- Though it's a challenge to speak to everyone individually – we actively encourage people to go through groups like **use MY data** to have their say on things like
  - How we design and build the service
  - How we run it day-to-day
- We're actively working with Chris, John and the use MY Data team - so please come join in

# How the TRE service matters...

“Rapid, safe and trustworthy access to essential health and care data for timely research at scale”

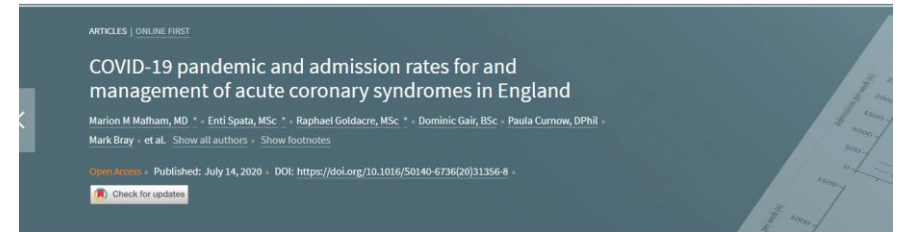
## Early NHSD TRE Service work

- First publication in *The Lancet* out now, with national press coverage
- **20 more publications** in progress

## This is just the start... The enduring TRE service enables safe and timely research at scale for

- Evidence-based decision-making
- Benchmarking [e.g. a clinical intervention, a service etc.]
- Health and care surveillance/monitoring over time
- Evaluations of clinical & financial impact
- Improved care commissioning, assessment and planning

## THE LANCET



Summary

### Summary

**Background**

Several countries affected by the COVID-19 pandemic have reported a substantial drop in the number of patients attending the emergency department with acute coronary syndromes and a reduced number of cardiac procedures. We aimed to understand the scale, nature, and duration of changes to admissions for different types of acute coronary syndrome in England and to evaluate whether in-hospital management of patients has been affected as a result of the COVID-19 pandemic.

**Methods**

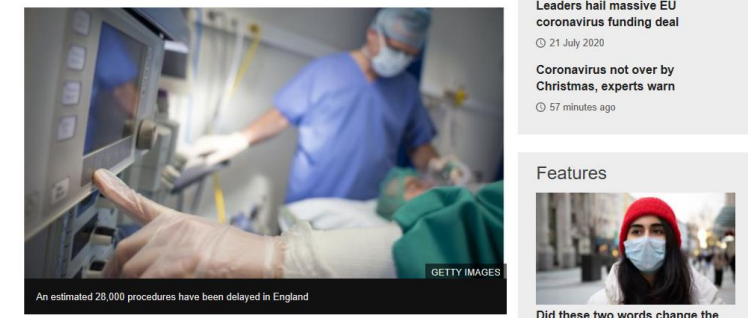
We analysed data on hospital admissions in England for types of acute coronary syndrome from Jan 1, 2019, to May 24, 2020, that were recorded in the Secondary Uses Service Admitted Patient Care database. Admissions were classified as ST-elevation myocardial infarction (STEMI), non-STEMI (NSTEMI), myocardial infarction of unknown type, or other acute coronary syndromes (including unstable angina). We identified revascularisation procedures undertaken during these admissions (ie, coronary angiography without percutaneous coronary intervention [PCI], PCI, and coronary artery bypass graft surgery). We calculated the



### Fewer heart attacks seen by NHS amid coronavirus

15 July 2020

Coronavirus pandemic



Hospital admissions for heart attacks dropped by a third across England when coronavirus took off in the UK and the nation went into lockdown, say researchers in *The Lancet* journal.

### Top Stories

- UK 'actively avoided' investigating Russian threat**  
Government failed to look into possible interference in the EU referendum, says parliamentary committee.  
10 minutes ago
- Leaders hail massive EU coronavirus funding deal**  
21 July 2020
- Coronavirus not over by Christmas, experts warn**  
57 minutes ago

### Features

Did these two words change the course of the outbreak?

Thank you

Q&A

**Connect with us**

@nhsdigital

company/nhs-digital

www.digital.nhs.uk

# TRUSTED RESEARCH ENVIRONMENTS

*A patient's perspective*

John Marsh

25 November 2020



# TRE's POTENTIALLY OFFER MAJOR BENEFITS

- Bring together data from diverse sources to enable new data-driven research and product development driving forward health outcomes and reduced healthcare costs
- Research-ready databases
- Centralised control to manage security, access and usage

# TRE's MIGHT INTRODUCE ADDED RISKS

- Bring together data from diverse sources
- Research-ready databases
- Centralised control to manage security, access and usage

# THEREFORE PATIENTS NEED TO KNOW

- What's the purpose of the TRE and what is it intended to deliver?
- What data is included, what is its source and has it been legitimately collected and transferred to the database?
- What controls are there over access to the data and over who can use it and for what purposes?
- Have the technical and operational environment been independently verified?
- Who has been authorised to access the data and for what purpose?
- What benefits have been delivered and who gets the benefits?
- Have there been any screw-ups, what's been learnt from them and what's been done to prevent repetitions?



# THIS PATIENT'S RECOMMENDATIONS

- Develop a comprehensive health TRE within the legal and ethical boundaries
- Clear, understandable, up to date patient information is as important as technical development for long term benefit – must be available at launch
- Get on with it and deliver asap so we get the benefits asap

# TWO FINAL WISHES

1. No screw-ups!
2. Best of luck!



# Secret Garden or Open Landscape - Views on Trusted Research Environments

Panel discussion and Q&A

Please tweet with:

@useMYdata  
#datasaveslives

[www.useMYdata.org.uk](http://www.useMYdata.org.uk)  
[join@useMYdata.org.uk](mailto:join@useMYdata.org.uk)  
[@useMYdata](https://twitter.com/useMYdata)

[#datasaveslives](https://twitter.com/hashtag/datasaveslives)

