

An introduction to the Health Data Research UK Gateway: facilitating discovery and access to health data

06/07/22 | Susheel Varma, Ex- Chief Technology Officer, HDR UK
Current: Head of AI and Data Science, ICO



Our mission

HDR UK's mission is to unite the UK's health data to enable discoveries that improve people's lives

Our 20-year vision is for large scale data and advanced analytics to benefit every patient interaction, clinical trial, biomedical discovery and enhance public health.

Core funders



Programme funders



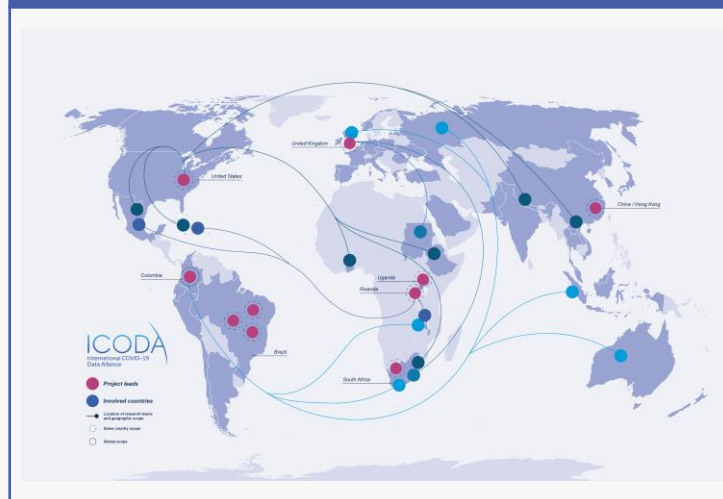
One Institute Partnerships

Three programmes working with local, regional, national and international partners across NHS, academic and industrial ecosystems:

UK Regional Networks



HDR Global



Institute Office



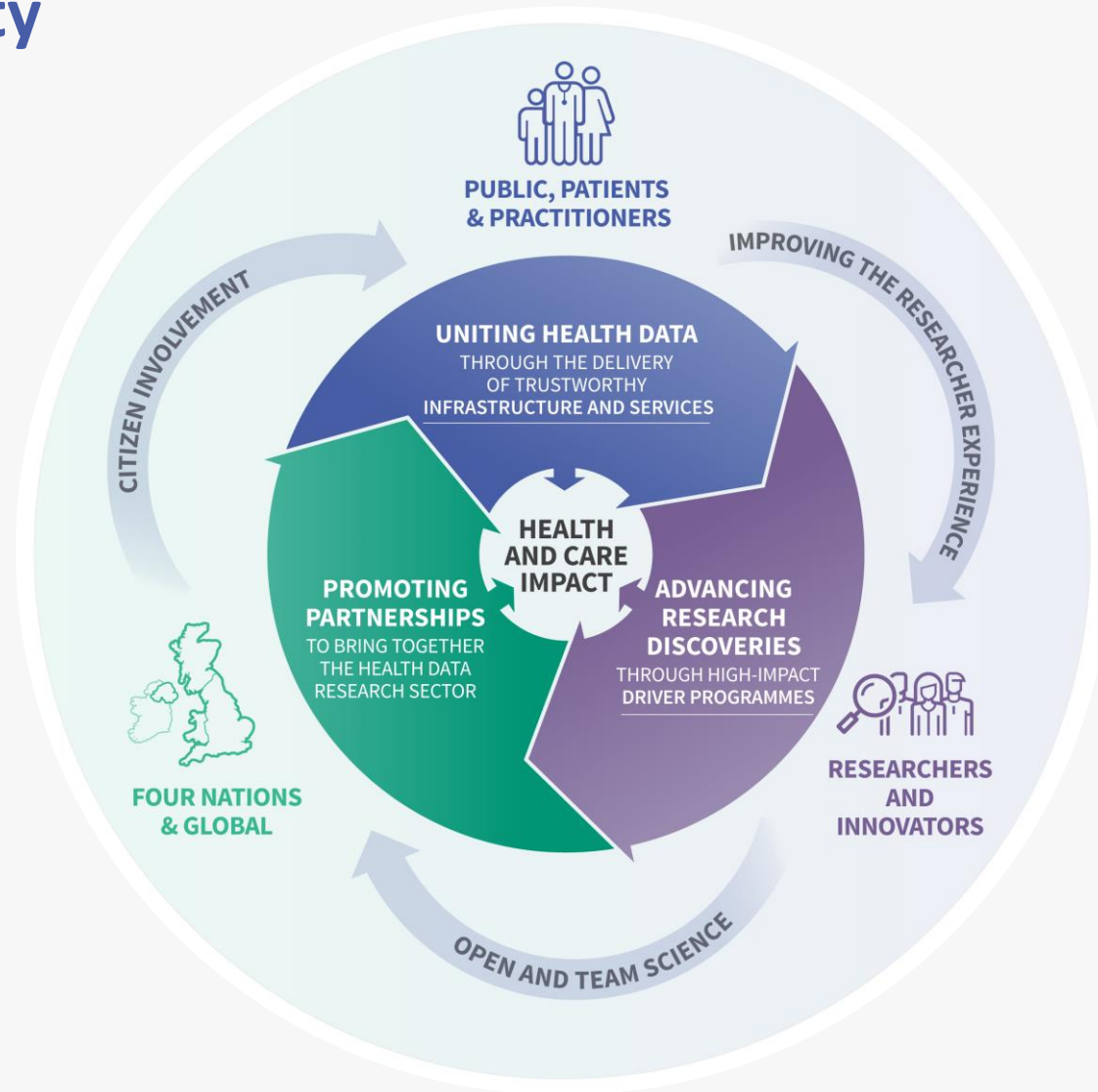
Drawing on health data science expertise and infrastructure across the UK and globally.
Delivering the transformative potential of health data research to impact everyone, everywhere

Our future strategy focuses on three integrated areas of activity

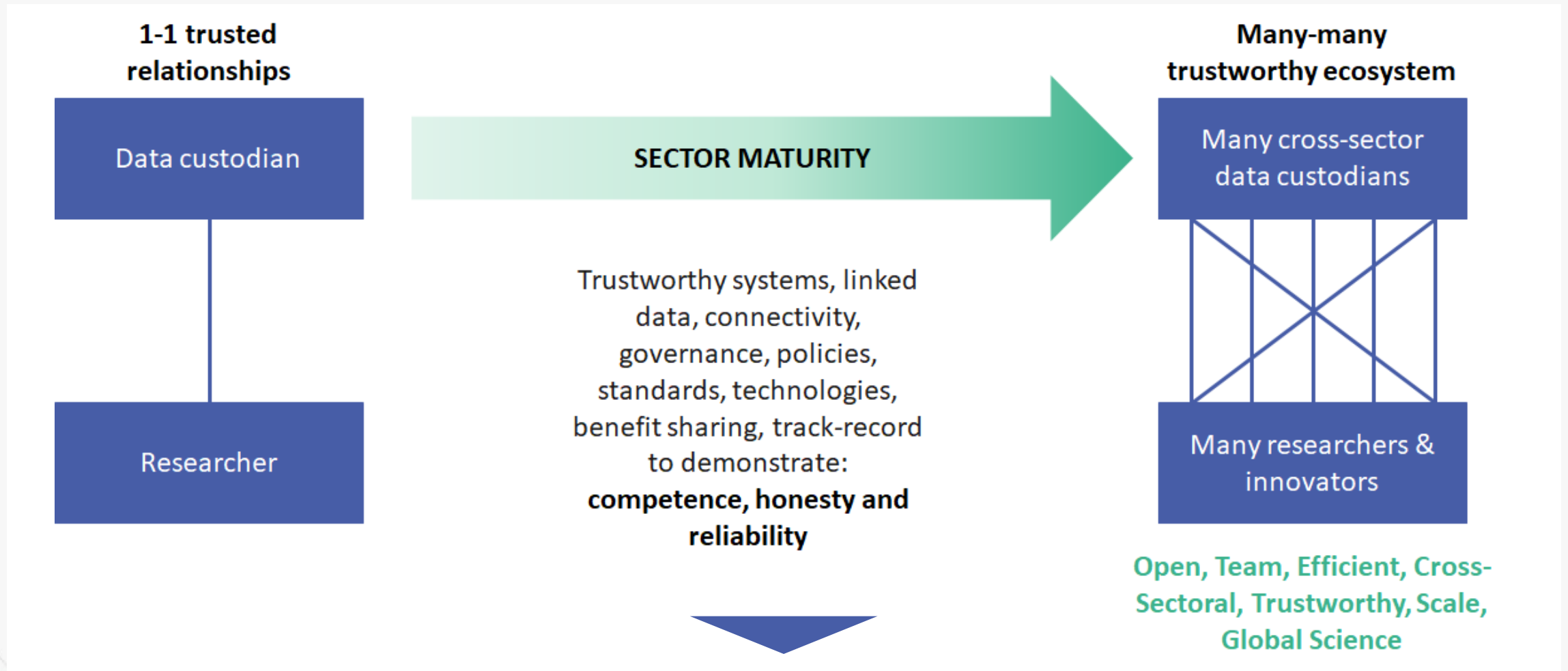
Uniting health data

Advancing research discoveries

Promoting partnerships

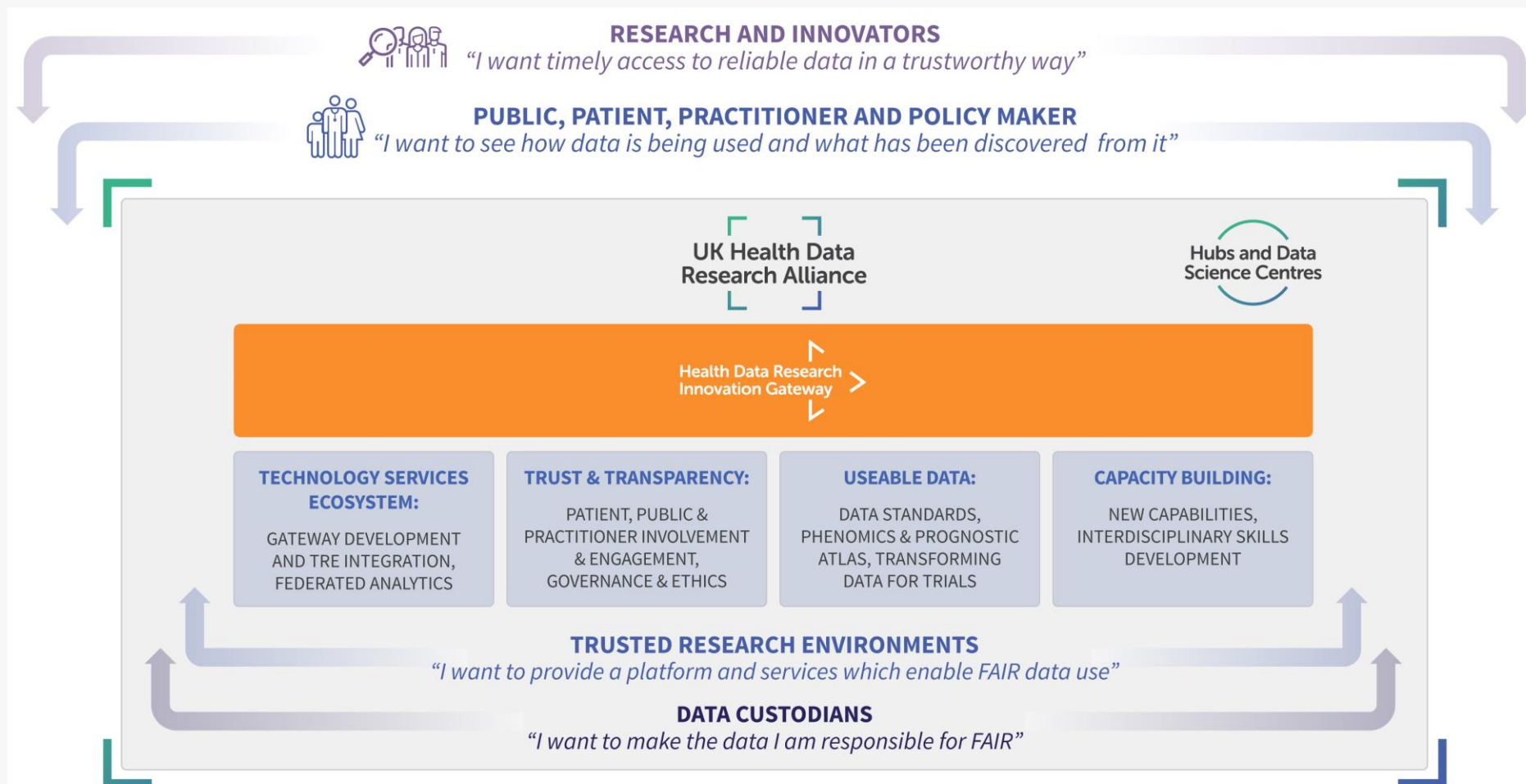


A trustworthy health data research ecosystem – differentiated talent, emerging technology, robust datasets, all connected through a platform



To transform science practice and health globally

1. Research Data Infrastructure and Services



Uniting health data through delivery of FAIR Research Data Infrastructure and Services



Leading healthcare and research organisations united to establish best practice, standards and tools



Our aim is for the Alliance to be **internationally recognised** as a **powerful collaboration** that is **enabling trustworthy use** of health-relevant data at scale and **networked** to deliver **public benefit**.

The UK Health Data Research Alliance (70 members and growing...)



Research Data Infrastructure and Services

We have built a UK-wide FAIR data infrastructure to accelerate and streamline health data research at scale



A collage of logos for various UK health data research organizations, including REATHE, Public Health Scotland, NIHR, Discover-NOW, NHS, CPRD, Imperial College London, biobank, generation scotland, NCIMI, NIHR BioResource, UKRI, Medical Research Council, NHS Digital, Gut Reaction, NHS, University Hospitals Birmingham, GIG NHS, HSC, Health and Social Care, CHIEF SCIENTIST OFFICE, TRIPLE DIRECTORY COORDINATION CENTRE, INSIGHT, RC GP, South London and Maudsley NHS Foundation Trust, DATA-CAN, NHS, University College London Hospitals NHS Foundation Trust, PIONEER, Health Informatics Centre, Human Fertilisation & Embryology Authority, Guy's and St Thomas' NHS Foundation Trust, NHS, Nottinghamshire Healthcare NHS Foundation Trust, and DIGITRIALS.

UK Health Data Research Alliance

Gateway to health data and tools for research

Search, discover and request access to hundreds of datasets, tools and resources for your research. Join the thousands of researchers and scientists worldwide who are already using the Gateway for research and scientific discovery.

Advanced Search

datasets	tools	projects	courses	papers	dataset requests	registered users	searches last month
677	155	257	195	1,293	299	1,560	9,633

View all usage data

Health Data Research Innovation Gateway

Data Utility Evaluation

We have developed a sophisticated tool to help researchers and innovators quickly and accurately access the datasets they need

Transforming the usefulness of data

Watch Later Share

Watch on YouTube

Data quality and standards

Patient and Public Involvement Impacts and Outputs

We work with and learn from the needs and experiences of members of the public and people using the health and care system. These case studies and opinion pieces give some examples of how patients and the public have been involved in our work, and the positive impacts they have had.

Case study
The role of PPIE in enabling rapid and trustworthy access to regional health data to support COVID-19 vaccine research
10 February 2022

Case study
Population Research UK – how public contributors helped to shape research at its earliest stages
19 January 2022
HDR UK sought to embed public and patient

Opinion
CO-CONNECT: A layperson's perspective on using health data, building trust and the benefits of sharing
18 January 2022

Leadership in public involvement

Building Trusted Research Environments - Principles and Best Practices; Towards TRE ecosystems

UK Health Data Research Alliance, NHS

Project leader(s)
Susheel Varma; Tim Hubbard; David Seymour; Nicky Brassington; Simon Madden

The paper acts as a guide for UK data custodians and other organisations involved in data sharing and information governance both within and beyond the health sectors. It identifies and provides principles, approaches and guidelines for data sharing and linkage practice within TREs, as well as examples of best practices drawn from the experiences of colleagues working in the Alliance and internationally.

The guidelines and principles in the paper are structured around the "Five Safes" framework for the access of health data – safe people, safe projects, safe settings, safe data, safe outputs. The White Paper also sets out principles for the federation of Trusted Research Environments - how they can be linked, to substantially enhance the scale of secure and trustworthy data linkage and research.

Governance and ethics

Advance your health data science skills, advance healthcare.

Discover the latest health data science breakthroughs on your coffee break

When it comes to health data science, be expertly informed.

Instead of refreshing your feed, why not refresh your health data science skills?

972 health data science training programmes

10,700 people trained

66 videos on HDR UK Futures

1,212 external users on HDR UK Futures

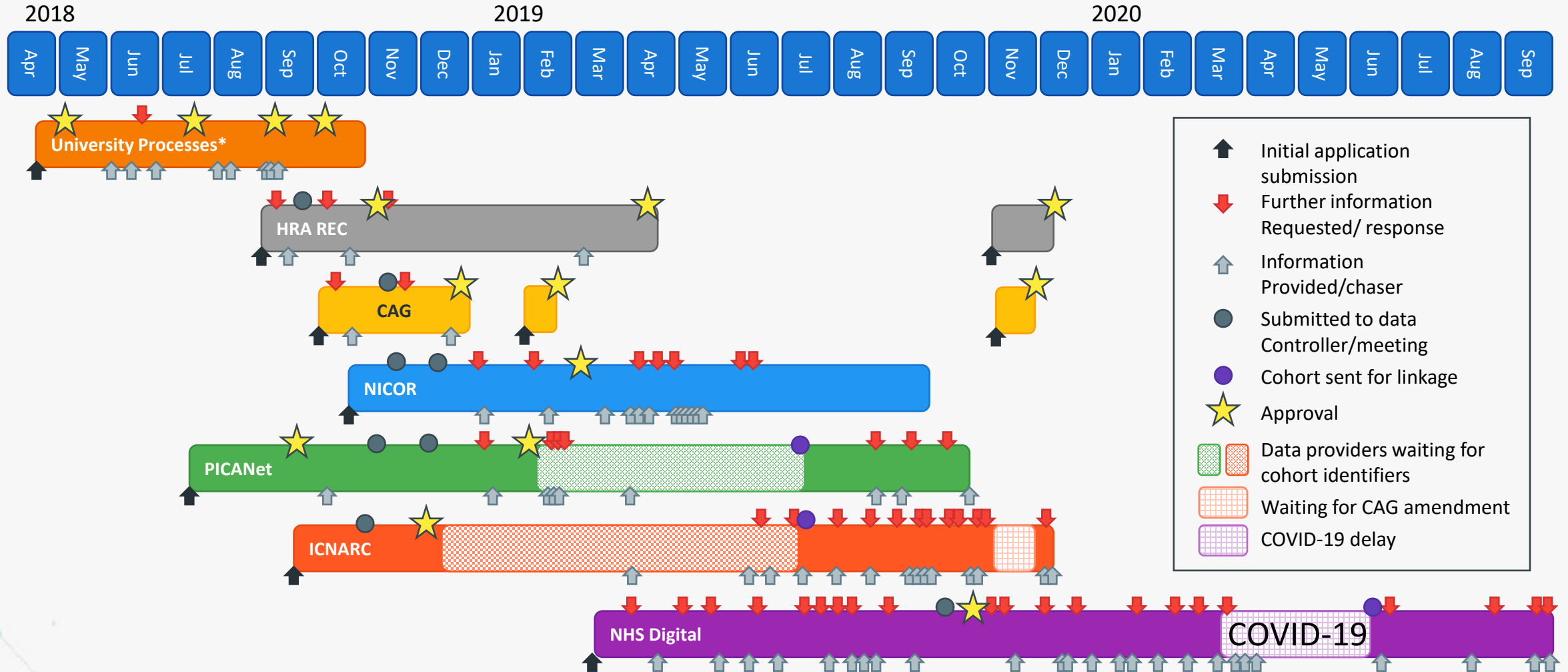
Capacity building



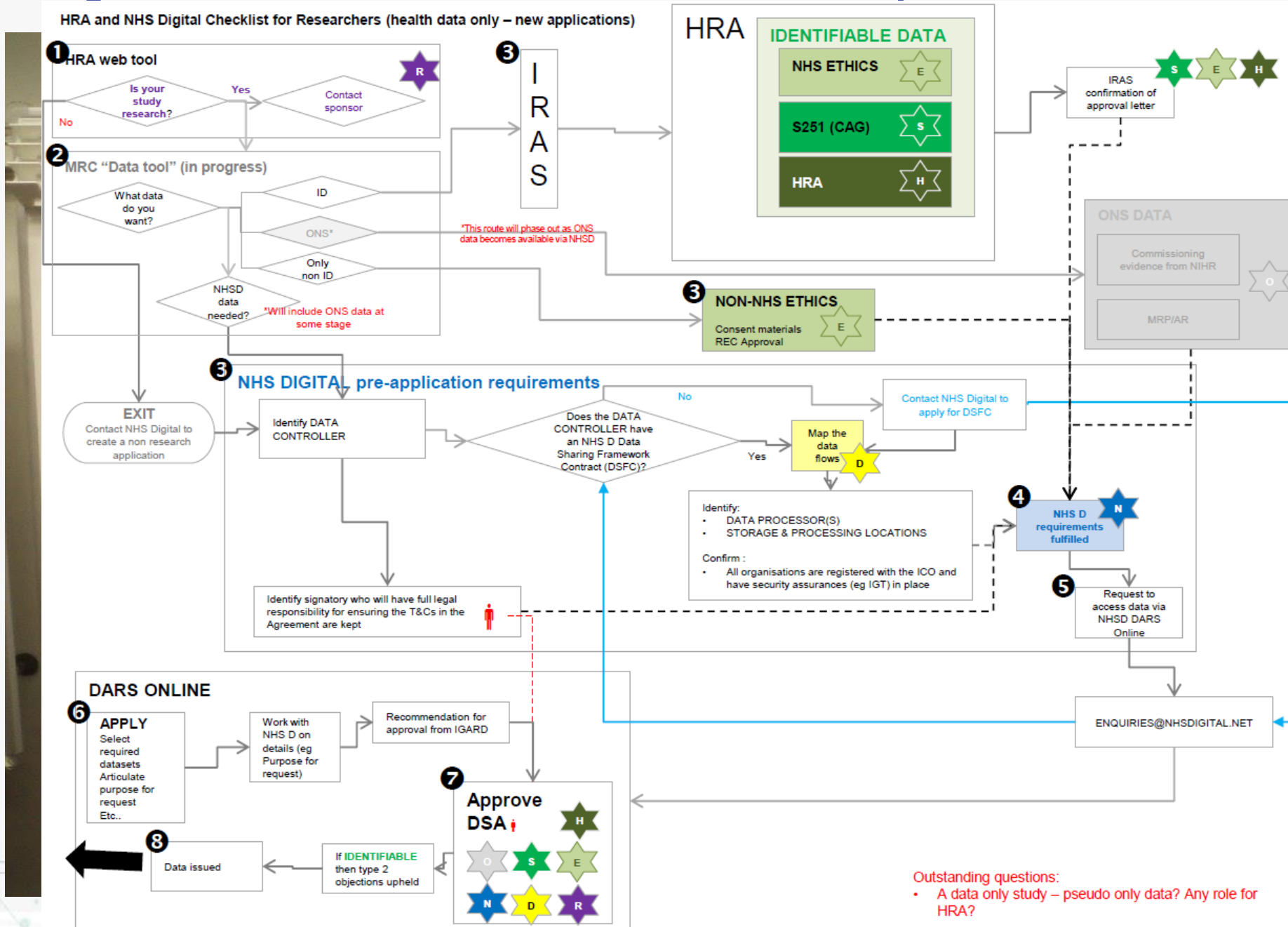
Our role in the landscape of UK-wide health data and research

HDR UK is one part
of the ecosystem
& to succeed in our
mission our role is
to enable and
convene other
organisations

The road to hell is paved with good intentions...



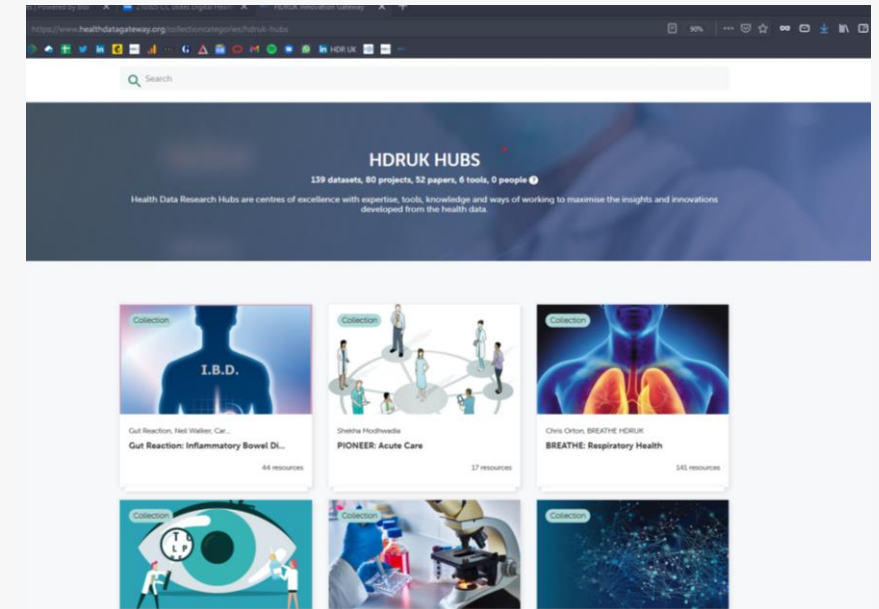
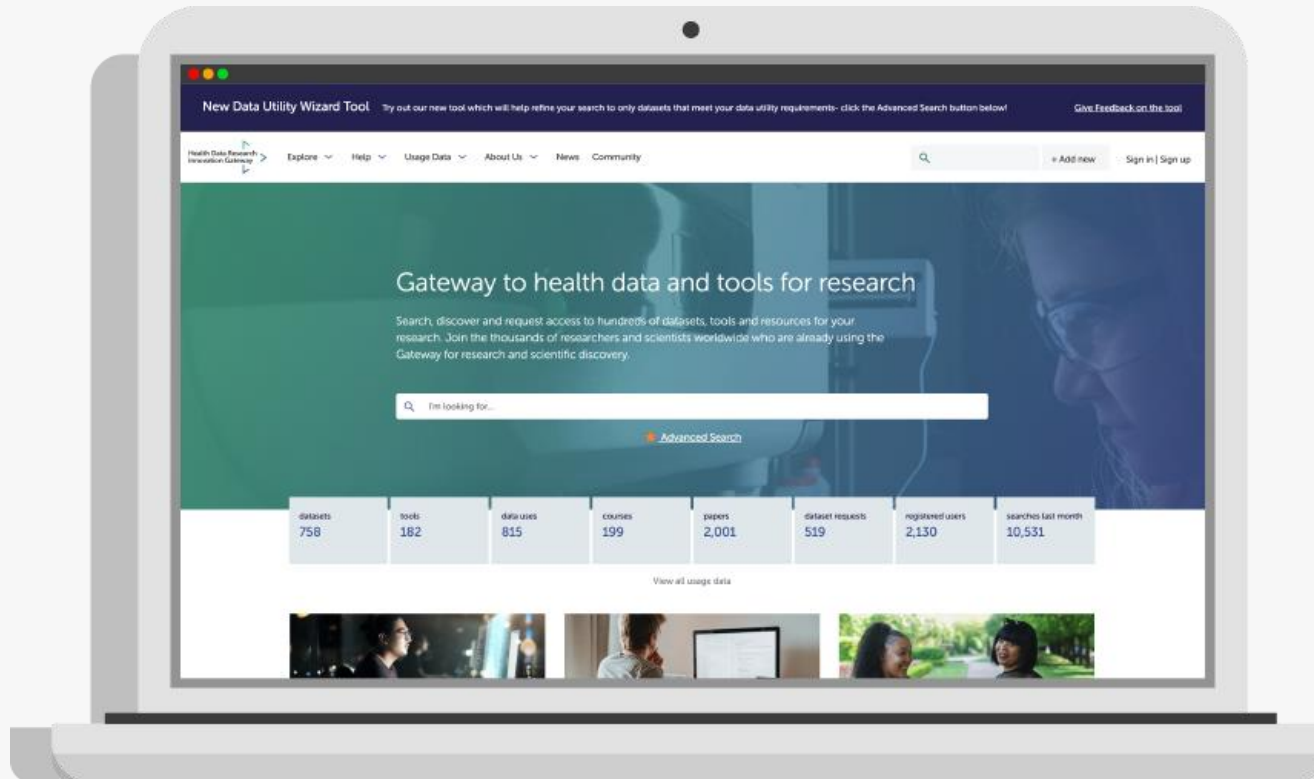
Streamlining access to health data → Accelerate research cycle times



Outstanding questions:
 • A data only study – pseudo only data? Any role for HRA?

Building the Gateway for researchers and innovators to access health data

www.healthdatagateway.org



- **Access to request over 774 datasets**, 188 tools, 199 educational courses and 2000+ publications
- A **dedicated collection of datasets for each Hub – 140 in total**
- **Facilitated 584 requests to access health datasets**, in particular actively supporting the government’s National Core Studies into COVID-19 (“Data and Connectivity”)

“Really impressed with this resource. I think as a gateway to search by data type and indication, it’s a really powerful tool.”

David Leather, GSK

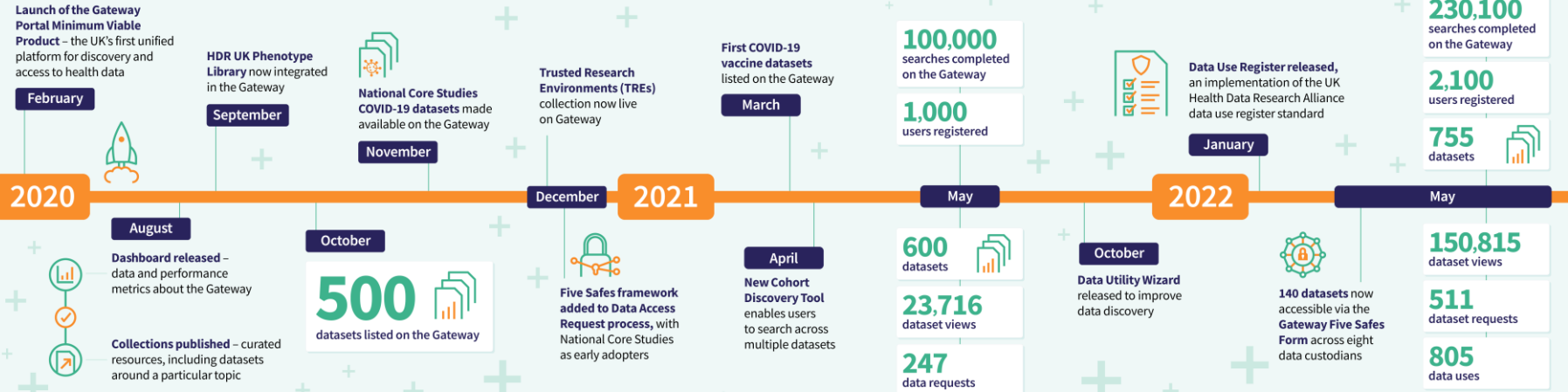
Health Data Research Innovation Gateway

The journey so far and the road ahead



Explore the timeline

Health Data Research Innovation Gateway
HDRUK
Health Data Research UK



● ○ ○ ○ ○

○ ● ○ ○ ○

○ ○ ● ○ ○

→



○ ○ ○ ○ ●

○ ○ ○ ○ ●

Register for a Gateway account
Search, discover and request access to hundreds of datasets available for research and scientific discovery

Sign up to our monthly newsletter
Be the first to know about all the latest Gateway updates by subscribing to our monthly newsletter

Health Data Research Innovation Gateway
HDRUK
Health Data Research UK

Connect with us

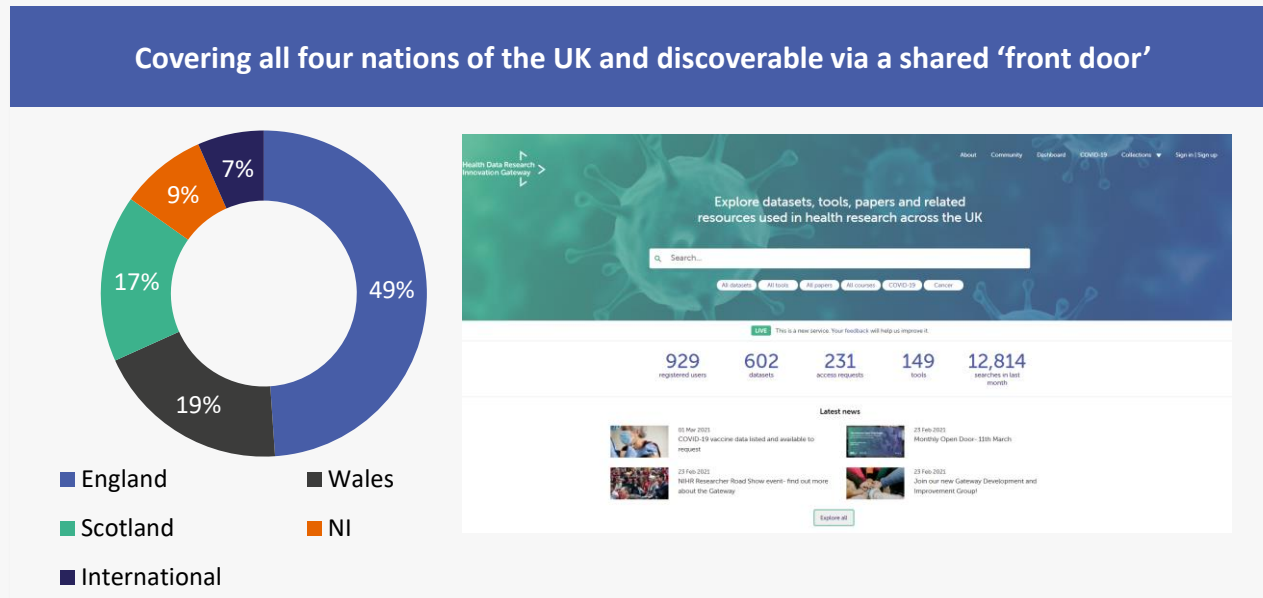


SCAN ME

A platform already used by custodians to make their data available

- ✓ An **independent Alliance** of data custodians dedicated to improving human health
- ✓ **50** member organisations (incl. national bodies, hospital trusts, charities, cohorts)
- ✓ Guided by Principles for Participation based on **transparency and public benefit**
- ✓ **Collaborating** to design **best practice and standards** for the use of health data

UK Health Data Research Alliance



Accelerate FAIR-ness in a federated world – Metrics, Cookbooks, Data Utility Framework, Maturity levels

FAIR cookbook

F Findability

- Exemplar recipes:**
- Unique, persistent identifiers
 - Search engine optimization

→ More about Findability

A Accessibility

- Exemplar recipes:**
- Transferring data with SFTP
 - Downloading data with Aspera

→ More about Accessibility

I Interoperability

- Exemplar recipes:**
- Selecting terminologies and ontologies
 - Creating a metadata profile

→ More about Interoperability

R Reusability

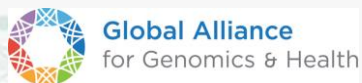
- Exemplar recipes:**
- Data licenses
 - Declaring data's permitted uses

→ More about Reusability

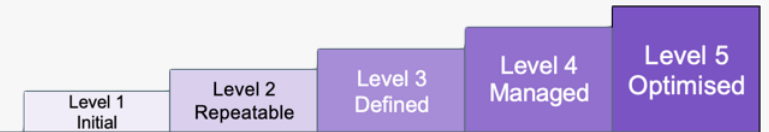
I Infrastructure

A Applied Examples

A Assessment



Category	Dimension	Definition	Bronze	Silver	Gold	Platinum
Data Documentation	Documentation Completeness	Preparation of metadata (as in the current HDR UK metadata specification) which is available in the expected format	65%-70%	71%-80%	81%-90%	91%-100%
	Availability of additional documentation and support	Available dataset documentation in addition to the data dictionary	Past journal articles documenting knowledge of the data exists	Comprehensive ReadMe describing extracting and use of data, Dataset FAQs available, Visual data model provided	As Silver, plus dataset publication was supported with a journal article explaining the dataset in detail, or dataset training materials	As Gold, plus support personnel available to answer questions
	Data Model	Availability of clear, documented data model	Known and accepted data model but some key fields un-coded or free text	Key fields codified using a local standard	Key fields codified using a national or international standard	Data Model conforms to a national standard and key fields codified using a national international standard
	Data Dictionary	Provided documented data dictionary and terminologies	Data definitions available	Definitions compiled into local data dictionary which is available online	Dictionary relates to national definitions	Dictionary is based on international standards and includes mapping
	Provenance	Clear description of source and history of the dataset, providing a "transparent data pipeline"	Source of the dataset is documented	Source of the dataset and any transformations, rules and exclusions documented	All original data items listed, all transformations, rules and exclusions listed and impact of these	Ability to view earlier versions, including versions before any transformations have been applied data (in line with deidentification and IR approval) and review the impact of each stage of data cleaning
Technical Quality	Data Quality Management Process	The level of maturity of the data quality management processes	A documented data management plan covering collection, auditing, and management is available for the dataset	Evidence that the data management plan has been implemented is available	Externally verified compliance with the data management plan, e.g. by ISO, CAC, ISO or other body	
	Data Management Association (DAMA) Quality Dimensions	Technical data quality dimensions: Completeness, Uniqueness, Accuracy, Validity, Timeliness and Consistency	These elements will be calculated with data profiling tools, and the category breakdown evaluated following further data collection			
Coverage	Pathway coverage	Representation of multi-disciplinary healthcare data	Contains data from a single speciality or area	Contains data from multiple specialities or services, within a single tier of care	Contains multimodal data or data that is linked across two tiers (e.g. primary and secondary care)	Contains data across more than two tiers
	Length of follow up	Average timeframe in which a patient appears in a dataset (follow up period)	Between 1 - 6 months	Between 6 - 12 months	Between 1 - 30 years	More than 30 years
Access & Provision	Allowable users	Allowable dataset usages as per the licensing agreement, following ethical and IR approval	Available for specific academic research uses only	Available for academic and non-profit (e.g. charity, public sector) uses only	Available for limited commercial uses (e.g. relating to a specific domain), in addition to academic and other non-commercial uses	Available for wider commercial uses (in line with ethical and IR approval), and addition to academic and other non-commercial uses
	Time Lag	Lag between the data being collected and added to the dataset	Approximately 1 year	Approximately 1 month	Approximately 1 week	Effectively real-time data
	Timeliness	Average data access request timeframe	Less than 6 months	Less than 3 months	Less than 1 month	Less than 2 weeks
Value & Interest	Linkages	Ability to link with other datasets	Identifiers to demonstrate ability to link to other datasets	Available linkages outlined and/or List of previously successfully linked datasets	List of restrictions on the type of linkages detailed. List of previously successful dataset linkages performed, with navigable links to linked datasets via a DOI/URL	Existing linkage with reusable or downstream approvals
	Data Enrichments	Data sources enriched with annotations, image labels, phenomes, derivations, NLP derived data labels	The data include additional derived fields, or enriched data	The data include additional derived fields, or enriched data used by other available data sources.	The derived fields or enriched data were generated from, or used by, a peer-reviewed algorithm.	The data includes derived fields or enriched data from a national report.



	Level 1 Initial	Level 2 Repeatable	Level 3 Defined	Level 4 Managed	Level 5 Optimised
Findable					
F1. (meta)data are assigned a globally unique and eternally persistent identifier.	No URI or PID and no documentation	PID without metadata or documentation	PID with limited metadata, just enough to understand the data	PID with standardised metadata registered or indexed in a trusted data repository	Extensive metadata and rich additional documentation available and searchable in a trusted data repository
F2. data are described with rich metadata.					
F3. (meta)data are registered or indexed in a searchable resource.					
F4. metadata specify the data identifier.					
Accessible					
A1 (meta)data are retrievable by their identifier using a standardized communications protocol.	No user licence / unclear conditions of reuse / metadata nor data are accessible	No metadata and user Access restrictions apply with only bespoke access	Appropriately licensed and limited (meta)data retrievable using standardised protocols	Public access (after registration) With (meta)data accessible (even when data is no longer available)	Open Access (unrestricted)
A1.1 the protocol is open, free, and universally implementable.					
A1.2 the protocol allows for an authentication and authorization procedure, where necessary.					
A2 metadata are accessible, even when the data are no longer available.					
Interoperable					
I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.	Proprietary, non-open format data	Proprietary format accepted by certified and trusted data repository	Non-proprietary, open format (archival format)	Data additionally harmonised/standardised using a standard vocabulary	Data is additionally linked to other data to provide context
I2. (meta)data use vocabularies that follow FAIR principles.					
I3. (meta)data include qualified references to other (meta)data.					
Re-usable					
R1. meta(data) have a plurality of accurate and relevant attributes.	No clear provenance of data (to facilitate replication and reuse)	Explication of how data was or can be used is available with user access restrictions	Data automatically usable by machines and (meta)data meet domain-relevant community standards	Data stored in a trusted data repository	Data is reliable and tested against gold standard (reference data)
R1.1. (meta)data are released with a clear and accessible data usage license.					
R1.2. (meta)data are associated with their provenance.					
R1.3. (meta)data meet domain-relevant community standards.					

The Health Data Gateway – Cohort Discovery

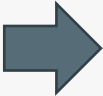
Solution: Cohort Discovery enables researchers to **discover, assess and request access to potential datasets** that exactly match the research project cohort definition using standardised **inclusion & exclusion criteria** and **co-variables**



Researchers

Datasets with **female** patients between **18-35** who have **asthma** and **diabetes** and who are **not smokers** and **not pregnant**

Cohort Definition



Cohort Query

Health Data Gateway

Total Patients: 23K

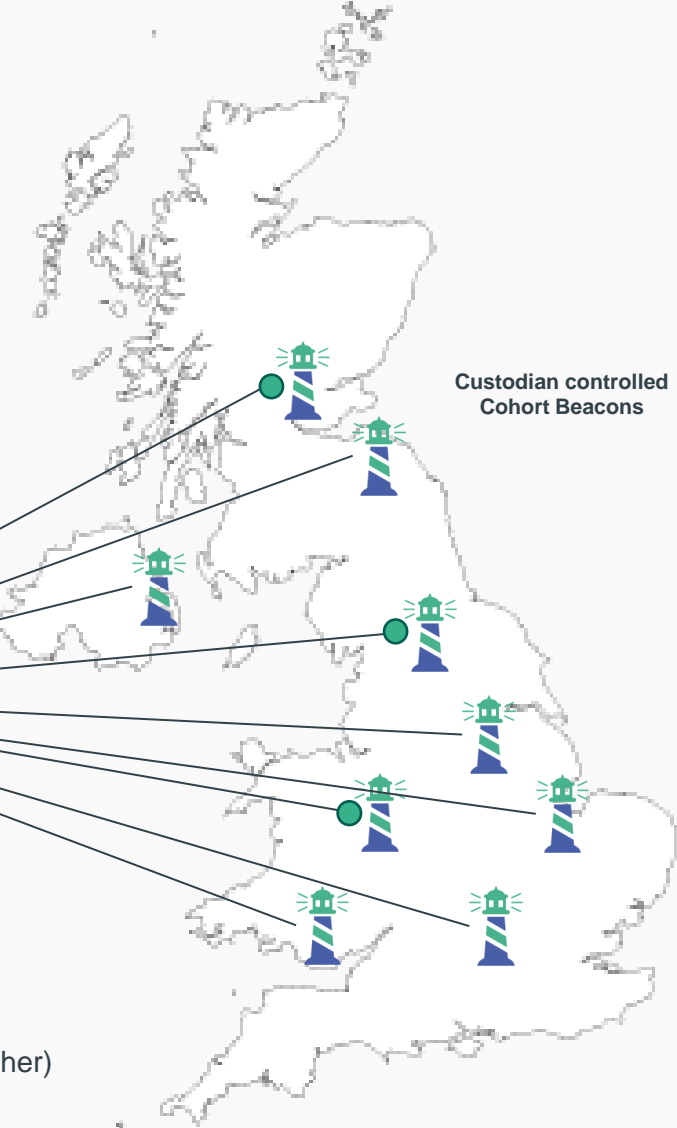
- Dataset 1: 20K
- Dataset 2: 3K
- Dataset 3: 1K
- ...

Cohort Query Engine

Query Resulting

- Researchers are able to **reuse the cohort query to define their research protocol** when submitting their data access request
- Researchers will be able to **reuse and compare cohort definitions** between similar protocols
- Cohort definitions will be able to **reuse phenotype definitions** (asthma, diabetes) without resorting to using ICD-10, Read, SNOMED-CT codes

- Statistical Disclosure Control Policies:**
- User validation (e.g. Bona-fide Researcher)
 - Low number suppression (e.g. >50)
 - Query Count binning (e.g. 50, 60, 70, -)
 - Query Rate limiting



Custodian controlled Cohort Beacons

Layout for illustration purpose only

Improving data access – our work so far

- ✓ Worked with Alliance members to develop a **common data access form** based on the **Five Safes*** for datasets listed in the Gateway
- ✓ We went through an **harmonisation process** to identify a set of **essential questions** common to all custodians
- ✓ Has launched on the Gateway for **283 datasets** and **12 custodians** (2 more soon). Look to continue to encourage datasets listed in the Gateway to use the new form
- ✓ Had **584 data access requests** using the new form (requesting 90 datasets)



Health Data Research Innovation Gateway > About Community

Data Access Request Breathe Last saved 12 Jun 2020 13:45 Save now

About this application

Safe people
Safe project
Safe data
Safe settings
Safe outputs

About this application

Requesting access to data can be a lengthy process due to the amount of checks needed in order to ensure the safe use of patient data. The steps below aim to clarify the process.

1 Select the datasets you need

2 Read the advice from the data custodian

- Get in touch with us before you begin if you're not sure we have the data you need
- We aim to review all applications within 30 days of submission
- The pre-submission conversations usually take between 1-2 months, although this varies considerably depending on the request
- The total cost for access is around £250 - £380, including TRE access

[Mark as read](#)

3 Communicate with the data custodian

4 Invite collaborators

5 Check what approvals you might need

[Collaborators](#) [Submit application](#) [Next](#)

Data Access Request Breathe Last saved 12 Jun 2020 13:45 Save now

About this application

Safe people
Individuals
Organisations

Safe project
Safe data
Safe settings
Safe outputs
Post-submission

Safe people

Please identify any persons or organisations who will have access to the data

Person 1

Role
Chief investigator

Title
Dr

Full name

Job Title

Affiliation / organisation

ORCID

Email

Guidance Answers Notes Messages

Role

Each application must have details for least 1 person. Chief investigator relates to the main researcher on a study. If you are an independent applicant and not affiliated with an institution, please select "citizen scientist".

Please make sure to select "add more people" and add details for every person who will have access to the data, even if they are independent partners.

Guidance from Breathe

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Suspendisse interdum dolor eu hendrerit pulvinar. Sed dolor metus, venenatis id quam ut, rutrum pretium acca. In sit amet turpis posuere, rutrum enim pulvinar, hendrerit libero. Duis vitae mauris vel risus condimentum pharetra.

[View all guidance in a new window](#)

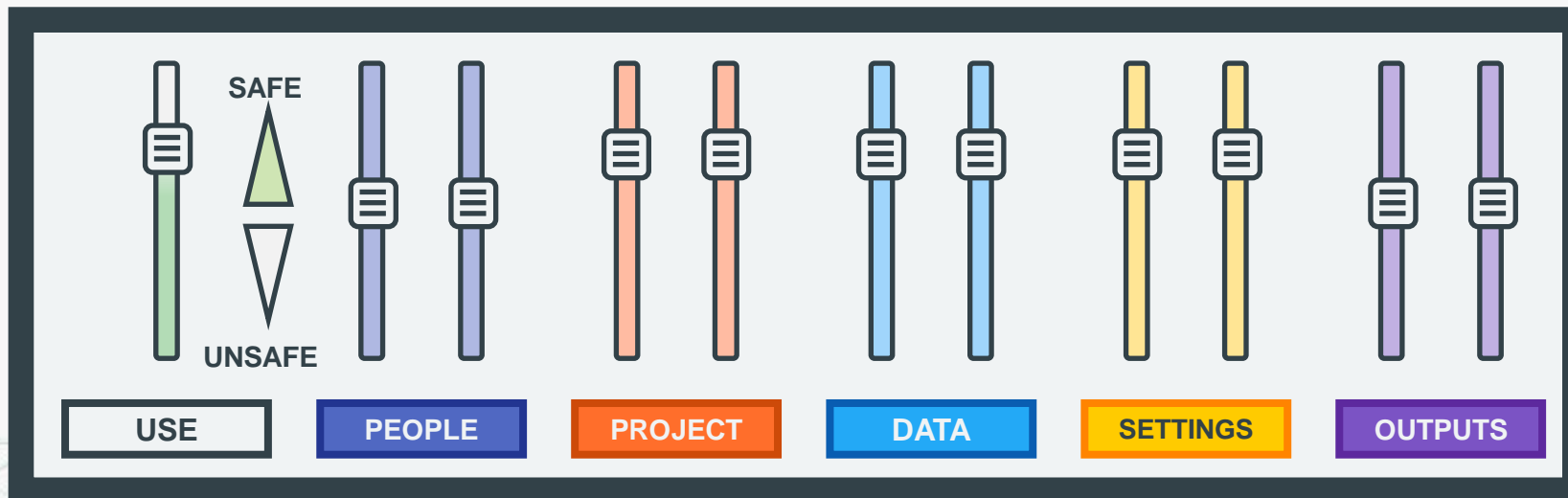
[Collaborators](#) [Submit application](#) [Next](#)

Data Governance – Risk Profiles based on the Five SAFE principles

	SAFE People	SAFE Project	SAFE Data	SAFE Setting	SAFE Output	
Public Data	None	None	Low	None	None	General Public
Aggregated Data	Low	Medium	Medium	Medium	Low	Data Journalists
De-Identified Data	Medium	High	High	High	Medium	Researchers
Sensitive Data	High	Very High	Very High	Very High	High	Hospital Analysts
Identifiable Data	Very High	Very High	Very High	Very High	Very High	Clinical Staff

Data Sensitivity

Example Users



Data Use Registers

Getting the greatest use and benefit from the UK's rich health data is an important goal– but it is essential that those whose data is being used can see what is being done with it.

The Gateway data use register functionality clearly shows how datasets published on Gateway are being used, by whom and most importantly for what purpose.



What benefits does the data use register offer?



Automated process: Data custodians who use the Gateway data access management system are automatically told about use of their datasets. This cuts down on administration time and provides the public with real-time data.



Closing the loop on impact: By providing a clear link between data use and research outputs we can better demonstrate the impact and value of using health data for research.



Alignment to Five Safes Framework: To show that only qualified people in secure settings can access the data, and for purposes that benefit the public, the content and structure of individual data uses follows the Fives Safes framework.

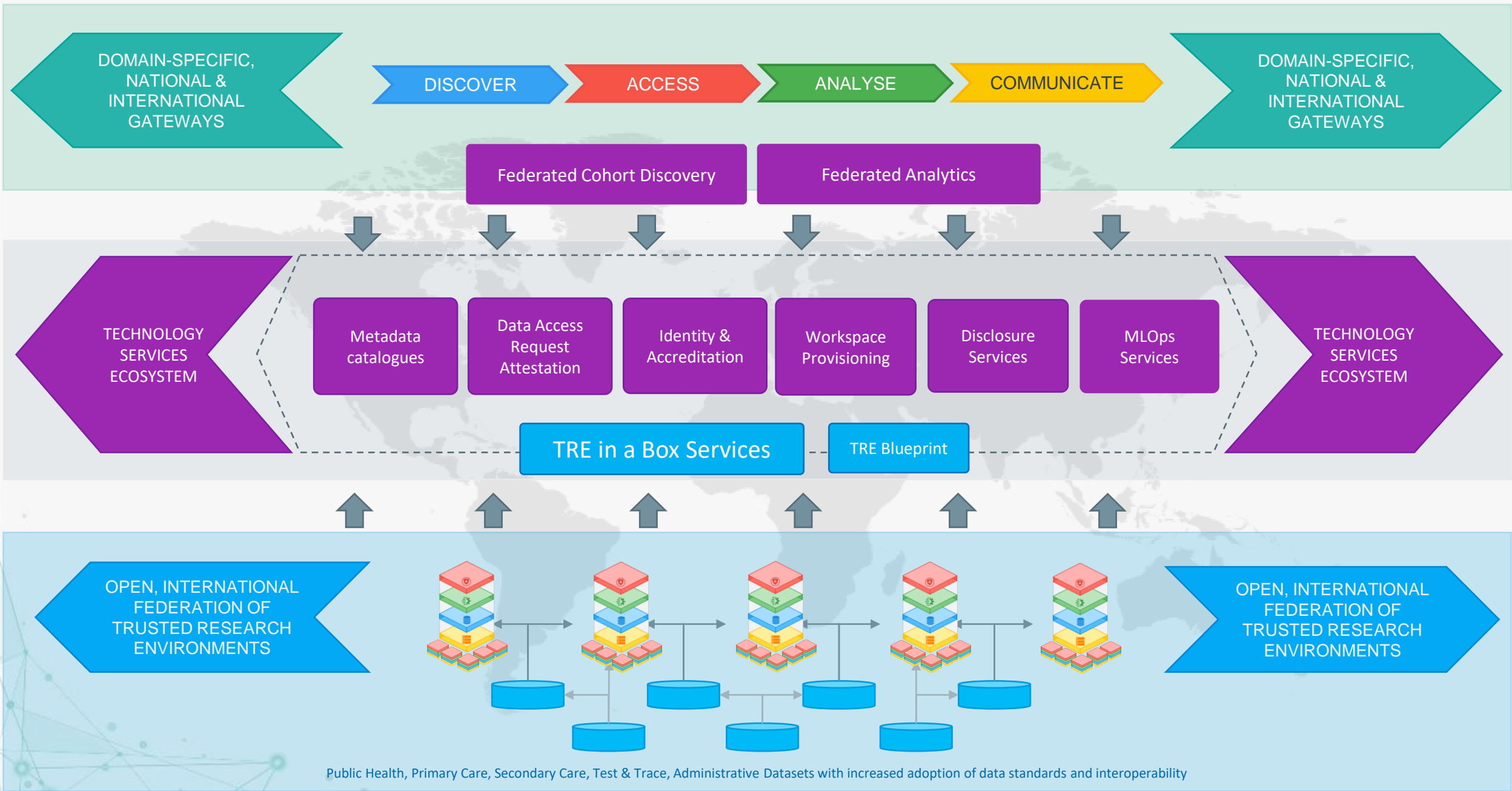


Best practice: The register implements a **national standard for data use registers**, developed after a wide consultation with data custodians and the public.

View the Data Use Register

For more information, visit the webpage

<https://www.hdruk.ac.uk/access-to-health-data/data-use-registers/>



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**.

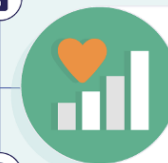
How is my data safeguarded?

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.



Safe People

Only trained and specifically accredited researchers can access the data



Safe Projects

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



Safe Settings

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



Safe Data

Researchers only use data that have been de-identified to protect privacy



Safe Outputs

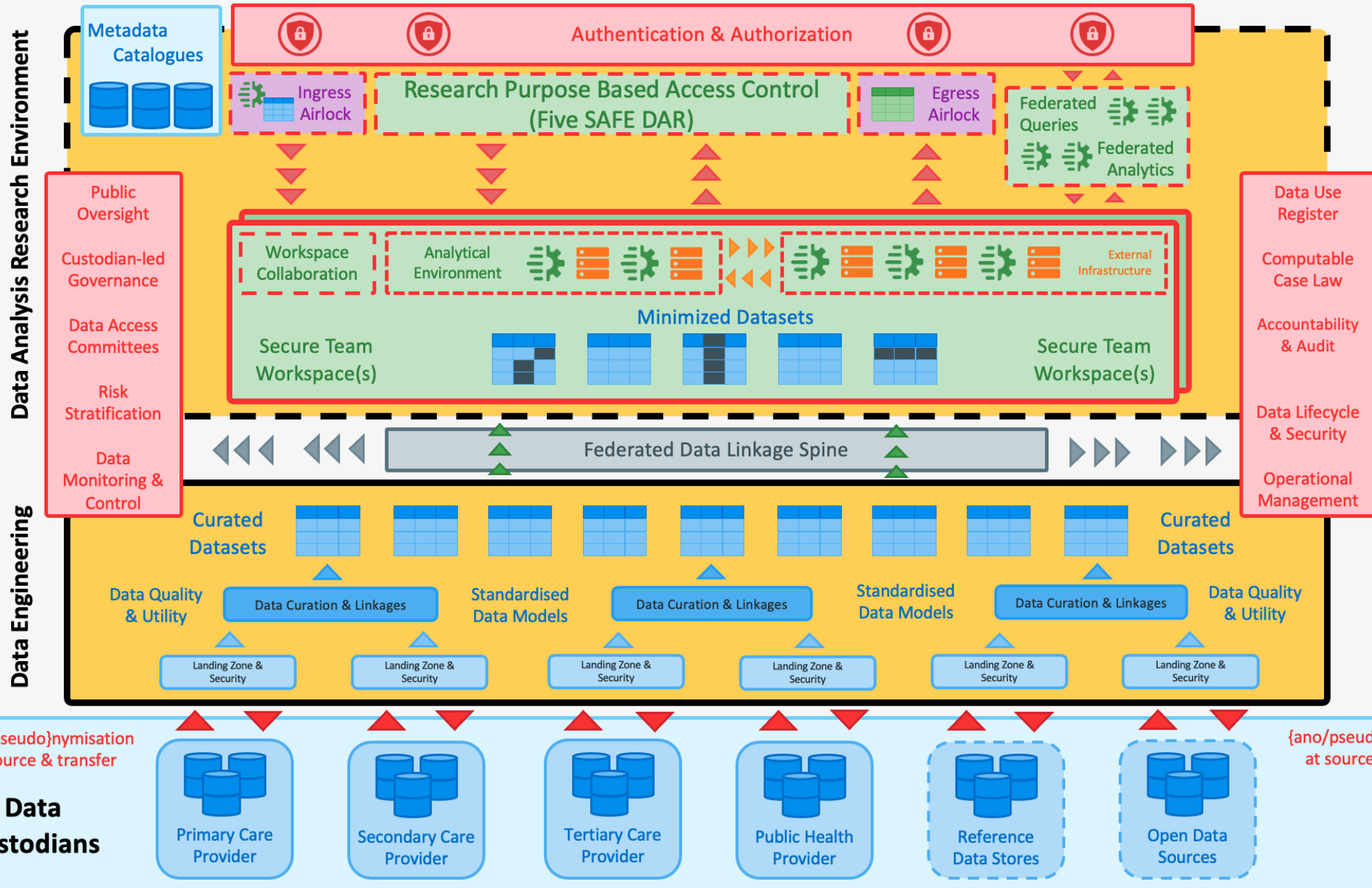
All research outputs are checked to ensure they cannot be used to identify subjects

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

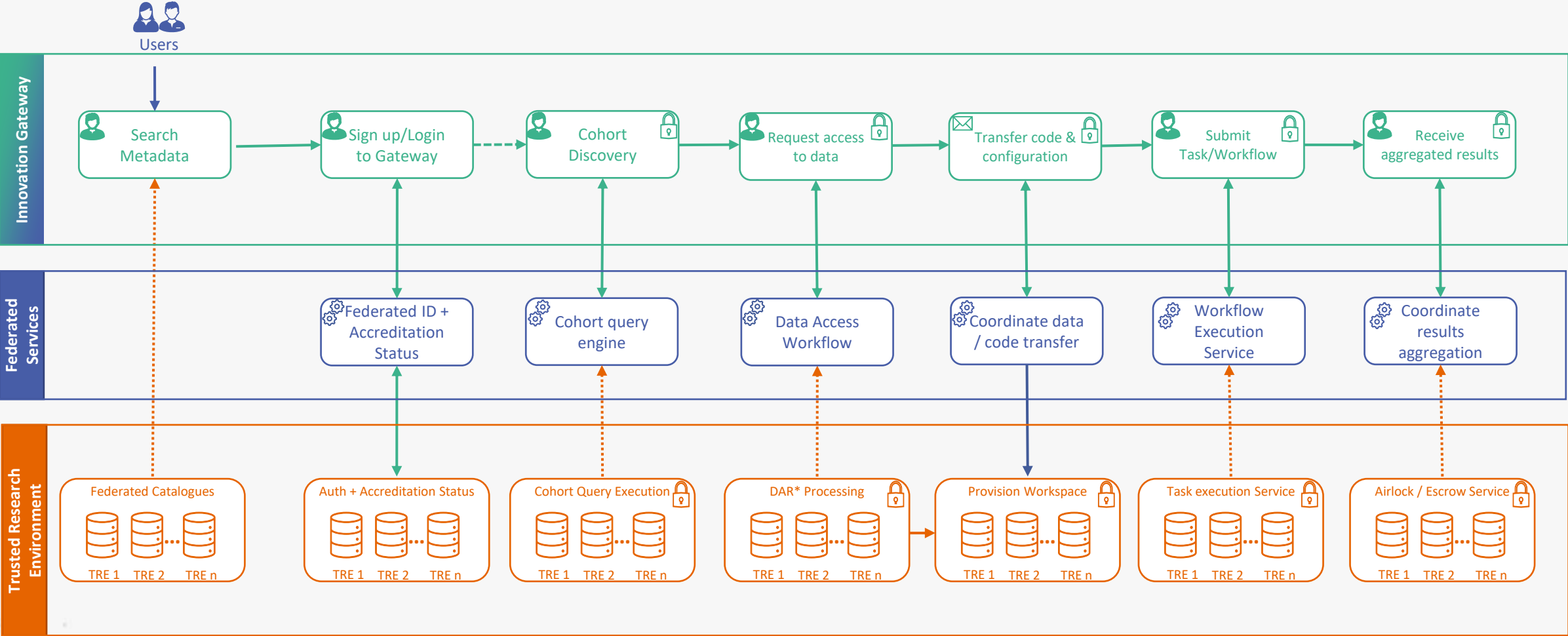
Learn more about TREs



What is a Trusted Research Environment?



Federated TRE User Journey – COVID-19 National Core Studies – Data & Connectivity



User action
 Automated action
 Process
 Authenticated user action

Trusted Research Environments Landscape View (research only – non-exhaustive)

National TREs

- HSCNI - Honest Broker Service
- PHS - Scottish National Safe Haven
- NHS Digital
- UKSeRP
- ONS - SRS
- OpenSAFELY

Commercial TREs

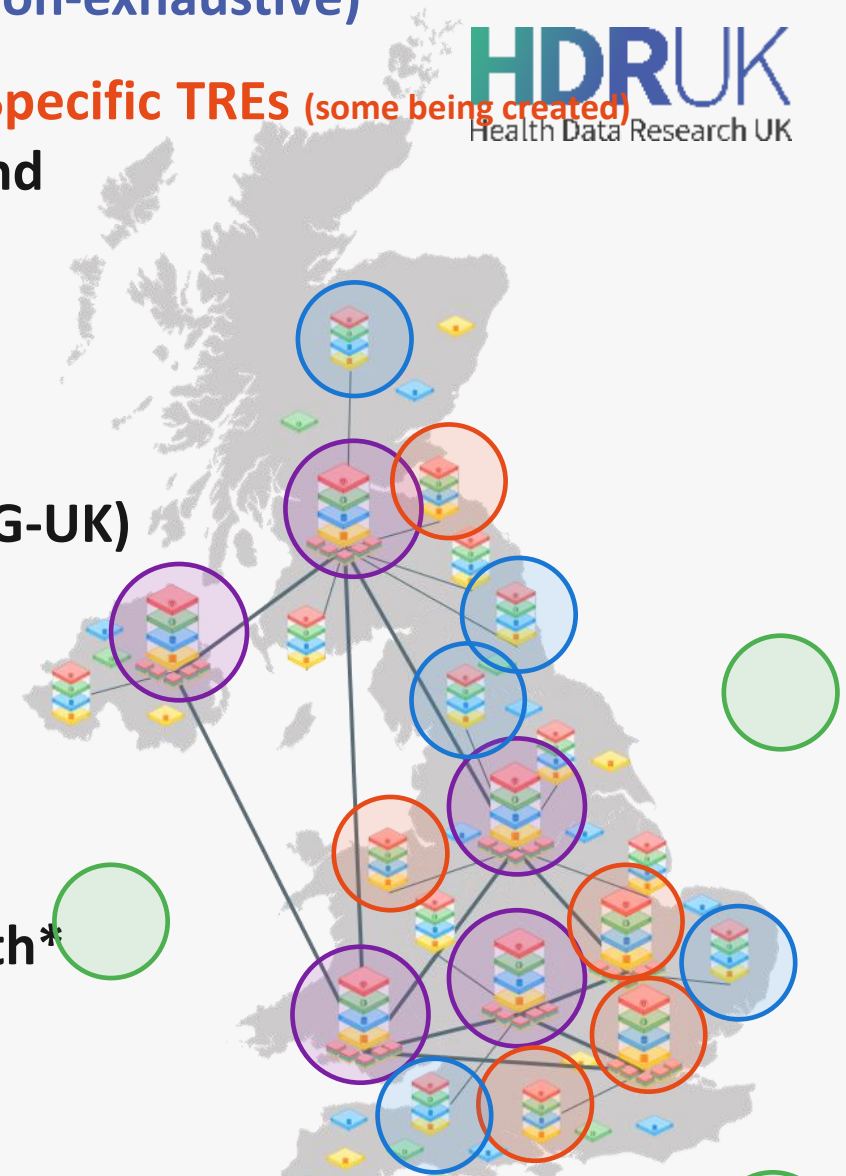
- AIMES
- AridhiaDRE
- AzureTRE
- AWS ServiceBench
- Lifebit
- ...

International TREs

- Terra.bio
- TEHDAS
- MedCo
- ICODA
- DNASStack
- ...

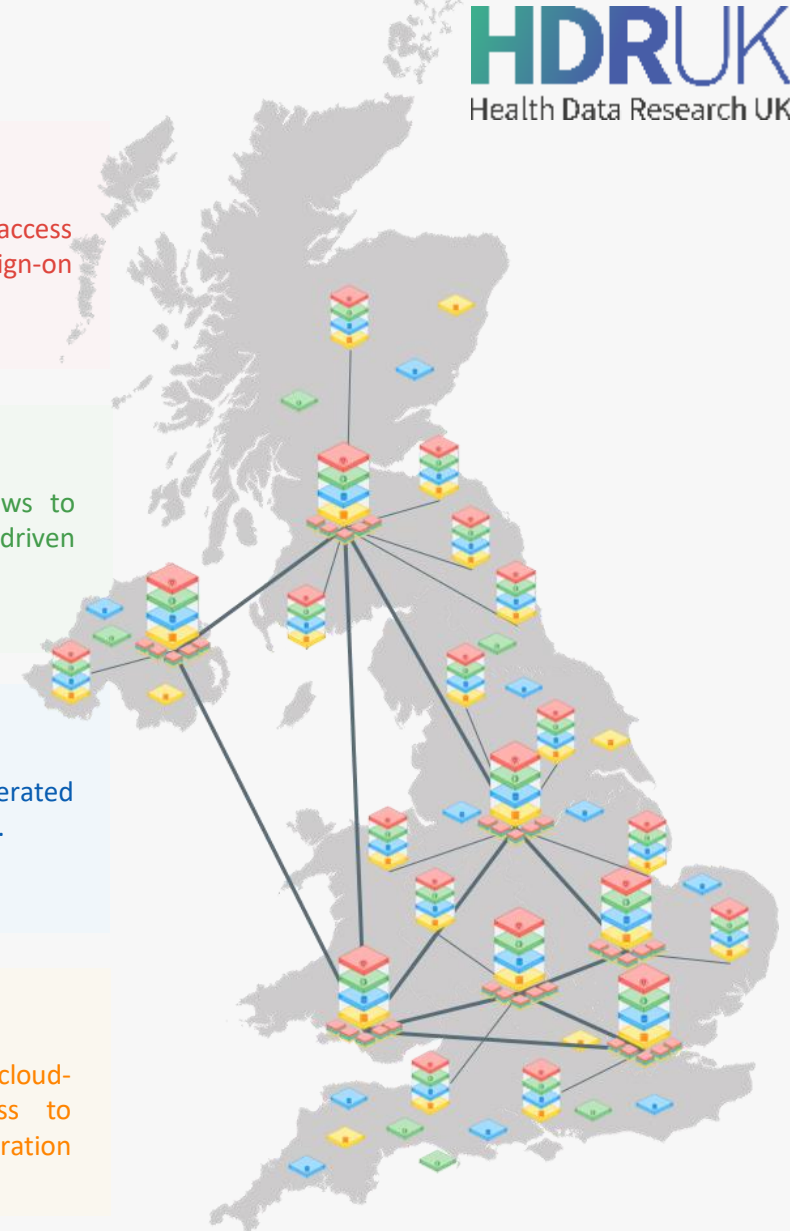
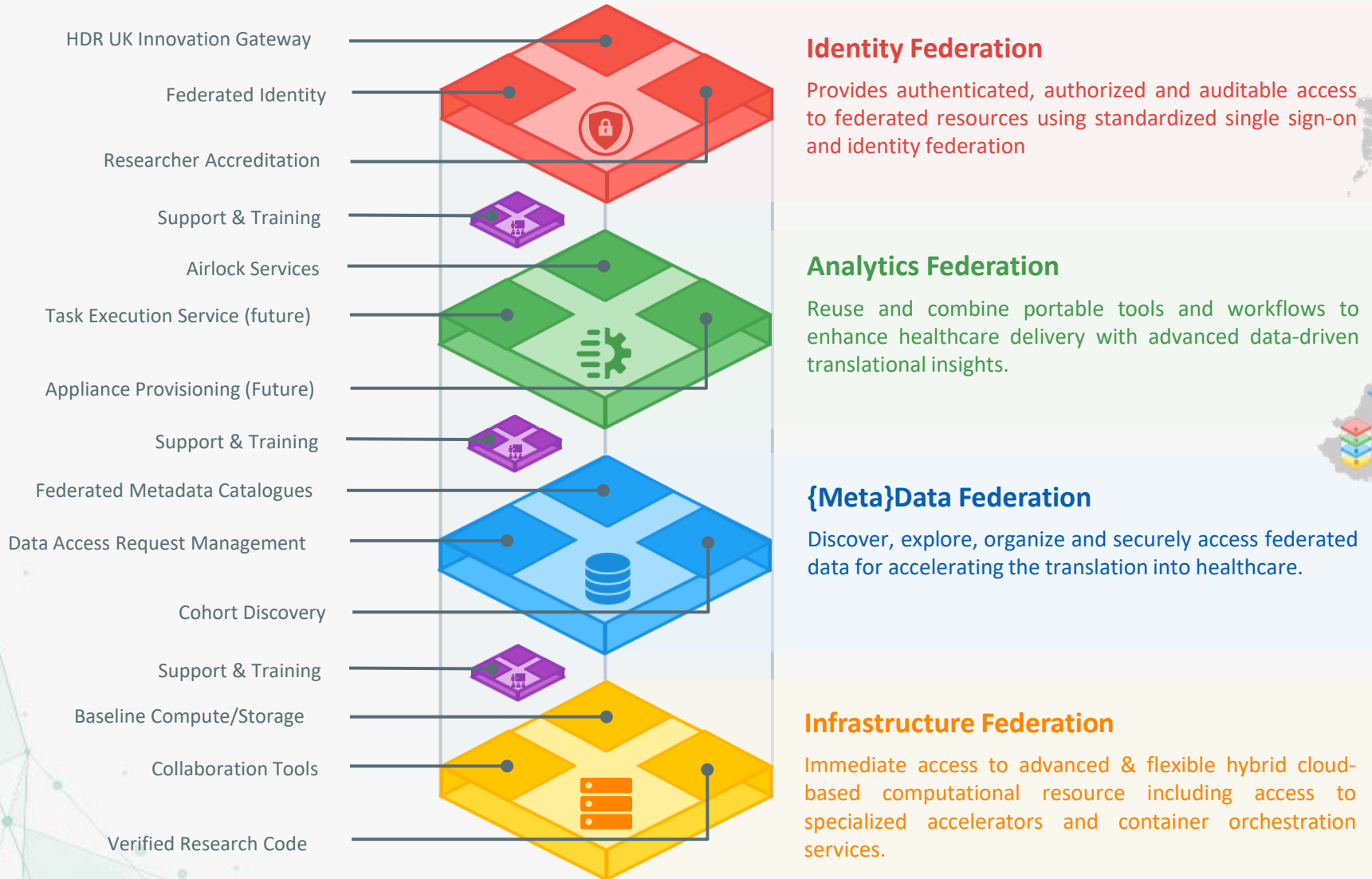
Domain/Location-Specific TREs (some being created)

- Genomics England
- QResearch
- EMBL-EBI*
- CPRD
- Turing Institute
- MRC-CLIMB (COG-UK)
- PIONEER
- eMedLab
- DISCOVER-NOW
- UKHSA
- CIPHA
- Our Future Health*
- ...



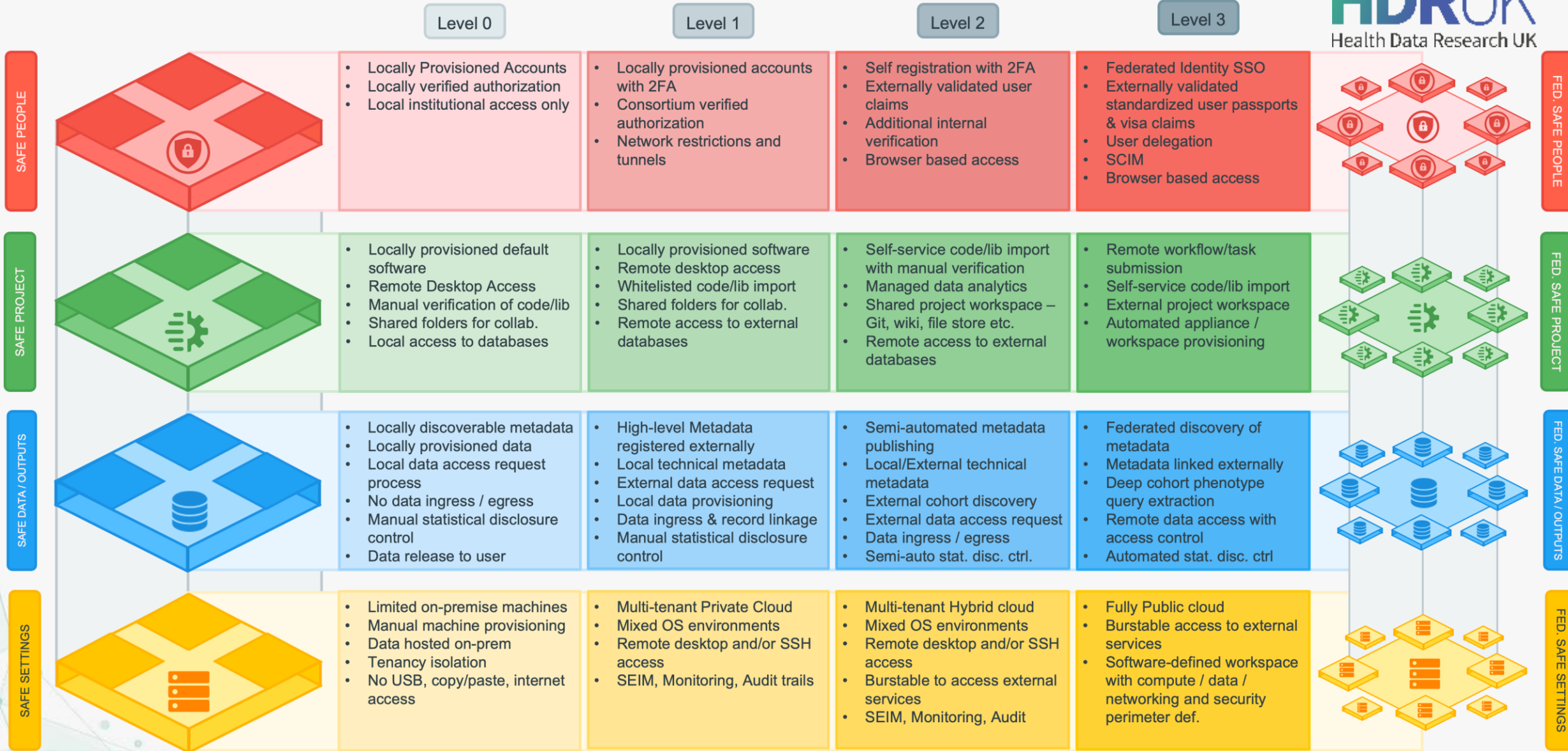
The future of TREs is already here, just unevenly distributed & fragmented

Open, Federated and Interoperable Technology Stack for Trusted Research Environments



Layout for illustration purpose only

FOREST – Federation Capability Maturity Model (Unbundling TREs)



Data Governance Structures / Design Patterns

	Governance Structures	Relationship	Data Availability	Permissions	Governance Challenges	IP, License & Contracts
	Closed	One → None	Very Low	Very Low	Limits to collaboration	Enshrined in Law
	Restricted	One → Some	Very Low	Low	Limits to collaboration	Enshrined in Law
	One-to-one	One → One	Low	Low	Power asymmetry	DSA, Contract
	Clique	Some → Some	Low/Medium	Medium	Trusteeship	Consortium Agreement
	Trusted Research Environments	Some → Many	High	Low/Medium	Trusteeship / user/project validation	DSA, Contract
	Federated Analytics	Some → Many	High	Low/Medium	Trusteeship / user/project validation	DSA, Contract
	Federated Query	Some → Many	High/Medium	Low/Medium	Trusteeship / user validation	DSA, Contract
	Open Access	Some → Many	High	High	Revocation of rights	License
	Citizen Science	Many → Many	High	High	Uneven capacity for analysis	Contract or License

Operationalizing Data Governance at scale – TREs are not a silver bullet

- Principles – FAIR, CARE?, SAFE?
- Requirements, Roles & Responsibilities, Rules of participation
- Interoperable Standards, Processes, Policies, Frameworks
- Modular Design Patterns, Software, Libraries
- Exemplars & Extensible Use Cases, Training, Awareness, Community Engagement



We need a set of **PRIME Directives** to help operationalize, interoperate and federate across data governance patterns

The future - building on these excellent foundations and partnerships to deliver a step change in benefits for UK science and population

- **Data as Infrastructure** – building blocks to create interoperable global networks focused on collaboration and sharing
- **Clinical/Scientific Driver Programmes** that help guide our development efforts and pilot tools
- **UK wide** and at the centre of an international collaborative network of science
- **Diverse data types beyond just NHS data** (omics, biomedical, wearables, social sciences)
- **Thought leadership and expertise for innovative approaches to trust, governance and standards** for data collection and access
- Exemplar of **team science** and **partnership** working across industry, academia and healthcare
- **Patient Involvement and Engagement** at all stages of the research/innovation lifecycle
- **Training** the current and next generation of scientists to responsible, ethical and equitable use of AI

DARE UK - Key needs and opportunities

dareuk.org.uk



1. **Data and discovery**, including technical standards such as data standards, metadata catalogues and common data models
2. **Access and accreditation**, including governance standards, rules and frameworks for enabling data access e.g., accreditation of TREs and researchers
3. **Digital research infrastructure**, including the physical and software infrastructure
4. **Capability and capacity**, including shortages of data scientists and data engineers, and career pathways for those maintaining infrastructure
5. **Maintaining trust**, namely demonstrating trustworthiness, gaining public and organisational trust, and addressing risk aversion of those holding data
6. **Funding and incentives**, including the current research culture, need for sustained funding of infrastructure and research, and the responsibilities of different groups involved



COVID-19 National Core Studies – Data & Connectivity

bit.ly/HDRUK-NCS

HDRUK
Health Data Research UK

HDRUK
Health Data Research UK

UKRI UK Research and Innovation

Data available for COVID-19 research across the UK

August 2021

This document sets out the data sets available for COVID-19 research across a network of UK Trusted Research Environments (TREs), made available by the Data and Connectivity National Core Study, led by Health Data Research UK in partnership with Office for National Statistics.

Office for National Statistics | NHS Digital | SAIL Databank | Public Health Scotland | HSC | British Heart Foundation Data Science Centre | OpenSAFELY

Data and Connectivity Partners

The programme works with partners to deliver streamlined access to a variety of health and administrative data via secure Trusted Research Environments across the four nations of the UK.

- Office for National Statistics** – The ONS – programme co lead
- NHS Digital** – NHS Digital (England)
- SAIL DATABANK** – SAIL Databank / University of Swansea (Wales)
- Public Health Scotland** – Public Health Scotland / EPCC University of Edinburgh (Scotland)
- HSC** – HSC Honest Broker Service (Northern Ireland)
- British Heart Foundation Data Science Centre** – British Heart Foundation Data Science Centre
- OpenSAFELY** – Open Safety

Key outputs across the NCS programme to date

- 489** Publications in academic journals, with 124 pre-prints
- 93** High-quality datasets made available via the Health Data Innovation Gateway, described in this brochure
- 190** NCS-linked data uses in the Data Use Register, a new standard for transparent public reporting shared in this HDRUK white paper
- 847** Researchers supported through 86 projects

National Core Studies Impact Report

Oct - Dec 2021

The COVID-19 National Core Studies (NCS) are a crucial part of the UK's ongoing pandemic response. They are enabling the UK to use health data and research to inform both our near and long-term responses to COVID-19, as well as accelerating progress to establish a world-leading health data and research infrastructure for the future

The 6 National Core Studies are:

Epidemiology and Surveillance led by Ian Diamond (Office for National Statistics) collects and analyses data to inform restrictions and protection against imminent outbreaks.

Clinical Trials Infrastructure led by Patrick Chinnery (Medical Research Council) accelerates delivery of large scale COVID-19 trials for drugs and vaccines.

Transmission and Environment (also known as PROTECT) led by Andrew Curran (Health and Safety Executive) improves understanding of COVID-19 virus transmission in different settings and environments.

Immunity led by Paul Moss (University of Birmingham) supports research to improve understanding of immunity against COVID-19, to inform back-to-work policies.

Longitudinal Health and Wellbeing led by Nish Chaturvedi (University College London) and Jonathan Sterne (University of Bristol) uses data from longitudinal studies to address the impact of COVID-19 and inform mitigating strategies

Data and Connectivity led by Andrew Morris (Health Data Research UK) in partnership with Office for National Statistics makes UK-wide health and administrative data available to catalyse COVID-19 research.

This new quarterly report aims to:

- Communicate the impact the National Core Studies are having on COVID-19 response
- Promote NCS Open Science data, tools and resources to ensure they are taken up
- Highlight where advances & learning gained during NCS translates into a legacy of stronger health threat preparedness

Managed by:



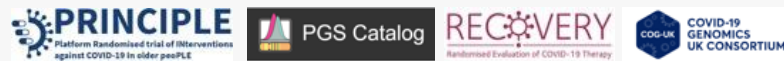
Funded by:



Are there examples of research being conducted in TREs?

RE-USABLE

1,168 COVID-19 pre-prints and 98 published papers. Better Care, Understanding Causes of Disease, Clinical Trials & Public Health



Public & Patient Engagement
National Core Studies
SAGE Reporting



FINDABLE

293 COVID-19 research projects using national data custodian data
300 Health Data Research Hub contracts with industry & academia



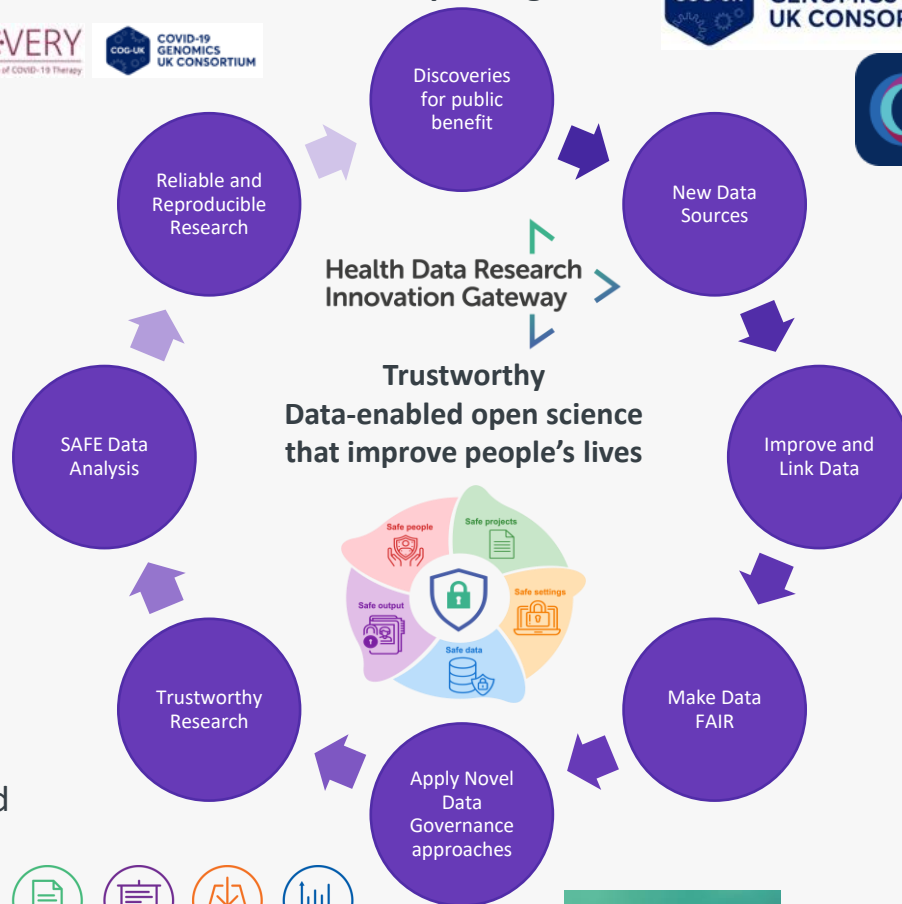
>22,000 patient and public participants in health data research
HDR Standards: Trusted Research Environments, Data Utility, Federation

Streamlined data access request process, harmonised across TREs. Time from application to active research: **3-14 days**¹



¹ Wales and Scotland data only

INTEROPERABLE



Example: Zoe COVID-19 symptom tracker dataset

Example: Hubs + CVD-COVID-19 For the first time, linked health data resource covering **54.4 million people**



112 datasets set up in 5 national trusted research environments – by National Core Studies Data & Connectivity



701 discoverable datasets
>16,000 monthly searches

ACCESSIBLE

HDRUK
Health Data Research UK

UK Health Data
Research Alliance

NHS^X

Thank you for listening

Here's some obligatory holiday snaps from the peak district

bit.ly/HDRUK-TRE



hdruk.ac.uk



healthdatagateway.org



Email: enquires@hdruk.ac.uk

Twitter: [@HDR_UK](https://twitter.com/HDR_UK)

