

UK Health Data Research Alliance

- Recovering pure mathematician with a 17 year meandering career in:
 - Computational Physiology, Health Informatics, Bioinformatics
 - Cardiac Electrophysiology, Orthopedics, Virology, Medical Devices, Dementia, Colorectal Cancer, Rare Disease
 - Technologist, Global Standards and expertise in setting up and operating national and international Research Infrastructures (4)
- Education/Awards:
 - PhD in Computational Systems Biology Cardiac electrophysiology
 - Exec. MBA in Management of Research Infrastructures
 - Fellow of the British Computer Society
- Hobbies
 - Not many people know that I'm a competitive lockpicker/safecracker
 - You're safe I don't do it professionally of course; I would say that wouldn't I;)
 - You had me at rescue/retired greyhounds
 - Clearly enjoy food:) Although I'm not sure what I'm doing in front of a salad!!



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

UK Health Data Research Alliance

National TREs

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

Commercial TREs

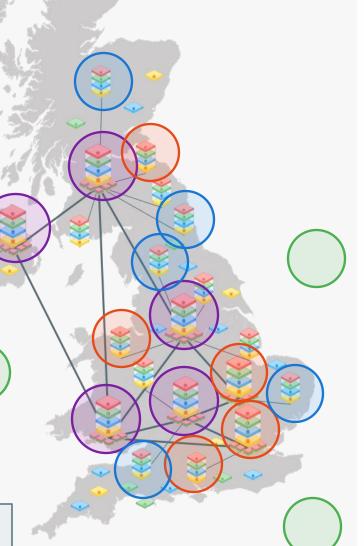
- AIMES
- AridhiaDRE
- AzureTRE
- AWS ServiceBench
- Lifebit
- ...

International TREs

- Terra.bio
- TEHDAS
- MedCo
- ICODA
- •

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
- ...



Layout for illustration only

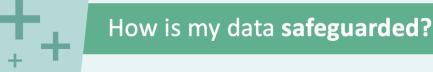
The future of TREs is already here, just unevenly distributed

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.







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



Safe People

Only trained and specifically accredited researchers can access the data



Safe Projects

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



Safe Settings

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



Safe Data

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



Safe Outputs

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

Why are they important?



TREs make research safer.

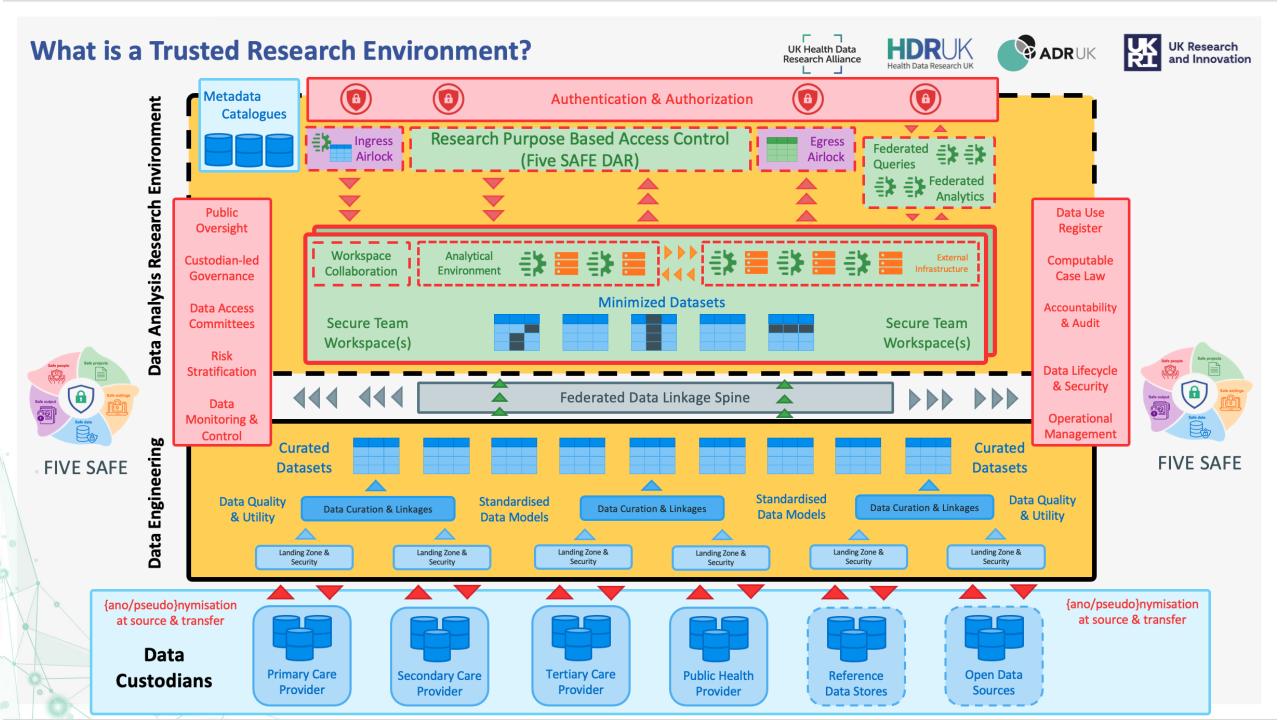
Making data available through a TRE means that people can be confident that their personal health data is accessed securely and their privacy protected. TREs help make research efficient, collaborative and cost effective, providing rich data that enables deep insights which will go on to improve healthcare 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.

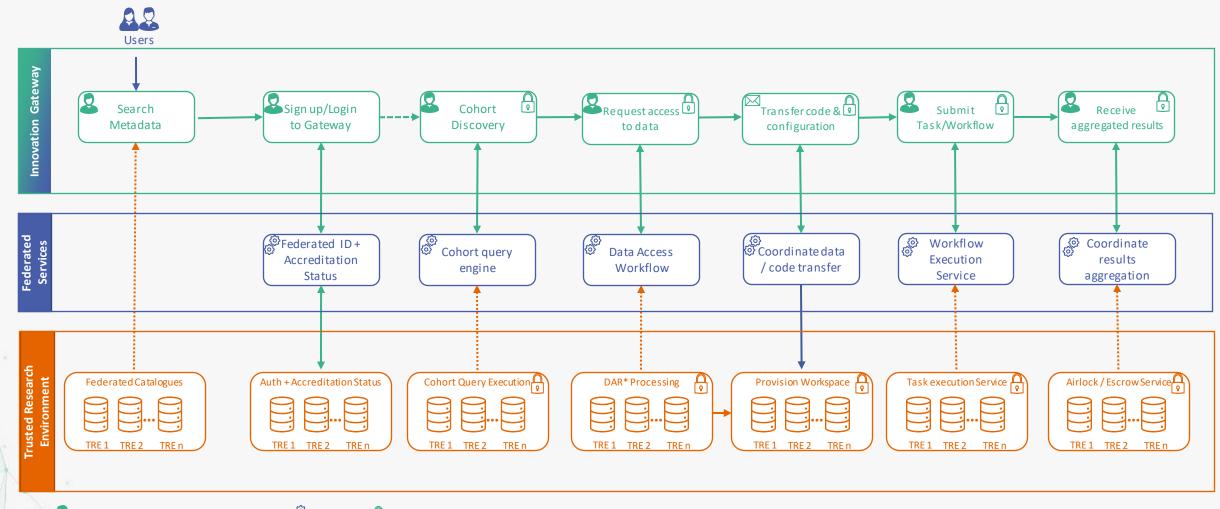
Learn more about TREs





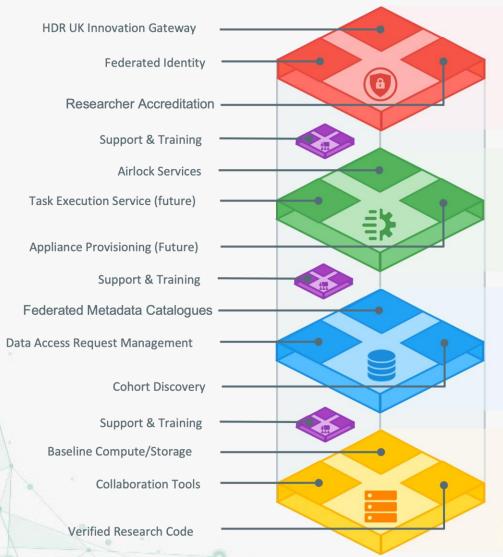
Federated TRE User Journey - COVID-19 National Core Studies - Data & Connectivity





FOREST - Federated, Open & Resilient Ecosystem of Secure TREs





Identity Federation

Provides authenticated, authorized and auditable access to federated resources using standardized single sign-on and identity federation

Analytics Federation

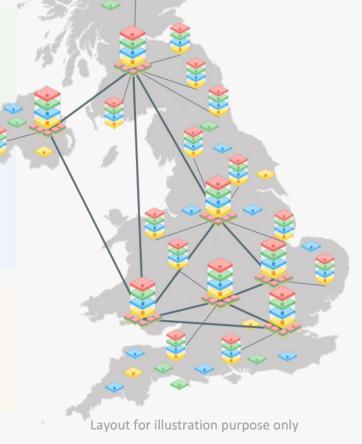
Reuse and combine portable tools and workflows to enhance healthcare delivery with advanced data-driven translational insights.

{Meta}Data Federation

Discover, explore, organize and securely access federated data for accelerating the translation into healthcare.

Infrastructure Federation

Immediate access to advanced & flexible hybrid cloudbased computational resource including access to specialized accelerators and container orchestration services.



FOREST – Federation Capability Maturity Model (Unbundling TREs)

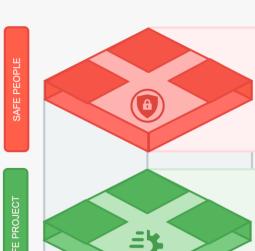
UK Health Data Research Alliance

Level 0

Level 1

Level 2

Level 3



Locally Provisioned Accounts

- Locally verified authorization
 - Local institutional access only
- Locally provisioned accounts with 2FA
- Consortium verified authorization
- Network restrictions and tunnels
- Self registration with 2FA
- Externally validated user claims
- Additional internal verification
- Browser based access

- Federated Identity SSO
- Externally validated standardized user passports & visa claims
- User delegation
- **SCIM**
- Browser based access



Locally provisioned default software

- Remote Desktop Access
- Manual verification of code/lib
- Shared folders for collab.
- Local access to databases

- Locally provisioned software
- Remote desktop access
- Whitelisted code/lib import
- Shared folders for collab.
- Remote access to external databases
- Self-service code/lib import with manual verification
- Managed data analytics
- Shared project workspace -Git, wiki, file store etc.
- Remote access to external databases
- Remote workflow/task submission
- Self-service code/lib import
- External project workspace
- Automated appliance / workspace provisioning





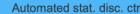
Locally discoverable metadata

- Locally provisioned data
- Local data access request process
- No data ingress / egress
- Manual statistical disclosure control
- Data release to user

- High-level Metadata registered externally
- Local technical metadata
- External data access request
- Local data provisioning
- Data ingress & record linkage
- Manual statistical disclosure control
- Semi-automated metadata publishing
- Local/External technical metadata
- External cohort discovery
- External data access request
- Data ingress / egress
- · Semi-auto stat. disc. ctrl.

- Federated discovery of metadata
- Metadata linked externally
- Deep cohort phenotype query extraction
- Remote data access with access control
- Automated stat. disc. ctrl









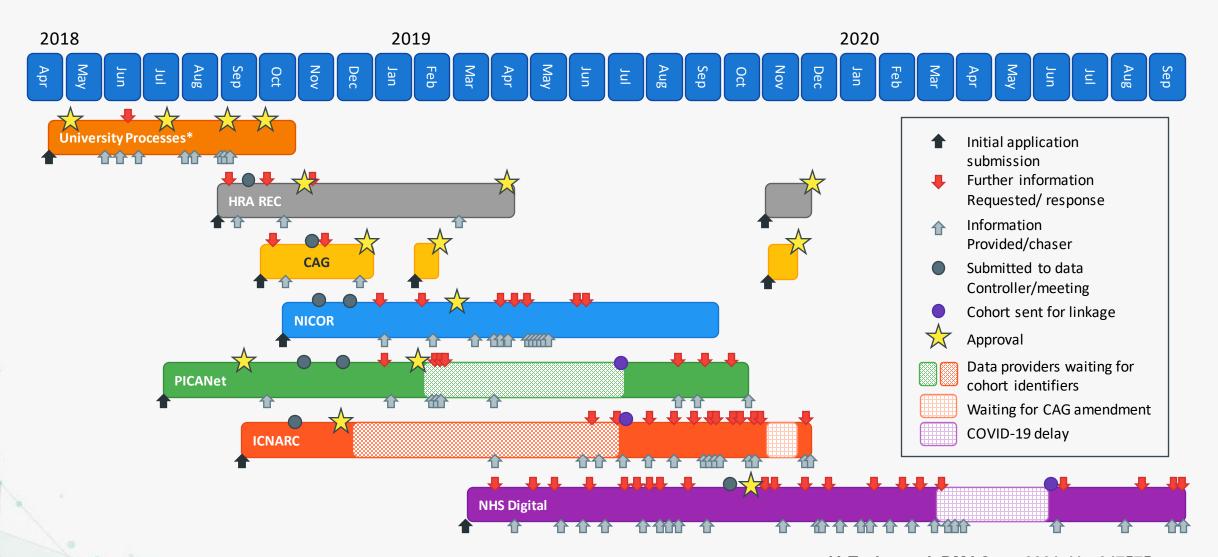
- Manual machine provisioning
- Data hosted on-prem
- Tenancy isolation
- No USB, copy/paste, internet access
- Multi-tenant Private Cloud
- Mixed OS environments
- Remote desktop and/or SSH access
- SEIM, Monitoring, Audit trails
- Multi-tenant Hybrid cloud
- Mixed OS environments
- Remote desktop and/or SSH access
- Burstable to access external services
- SEIM, Monitoring, Audit

- Fully Public cloud
- Burstable access to external services
- Software-defined workspace with compute / data / networking and security perimeter def.



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





Data Governance Structures / Design Patterns



	Governance Structures	Relationship	Data Availability	Permissions	Governance Challenges	IP, License & Contracts
• - X	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

Data Governance Design Pattern - Maturity Models (WIP)



Governance Patter	yns Janeway	/ Sisko	K	Kirk		Picard		
IP, License & Contra	Closed Restricte		Clique Fed. Analytic	TREs TREs TREs	Open Access	Citizen Science		
Data Sensitivity	Closed Restricte	One-to-One CI	lique Fed. Analytic	cs Fed Query	Open Access	Citizen Science		
User Qualificatio	Closed Restricte		Query	TRES	Open Access	Citizen Science		
Algorithmic Governa	Closed Restricte		Clique TRES	Fed. Query	Open Access	Citizen Science		

Not all governance structures and patterns are appropriate for all types of research collaborations



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













The future - building on this excellent foundation 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 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
- Patient Involvement and Engagement at all stages of the research lifecycle
- Training the current and next generation of scientists to responsible, ethical and equitable use of Al



Thank you for listening

Here's some obligatory holiday snaps from the peak district

bit.ly/HDRUK-TRE



hdruk.ac.uk



healthdatagateway.org



Email: susheel.varma@hdruk.ac.uk

Twitter: @susheelvarma



Do we want a "FORESTRE Commission" responsible for increasing the value of TREs – expanding, promoting, protecting the sustainable management of TREs across all four nations?