

Lightning Talks



SDC-REBOOT
COMMUNITY



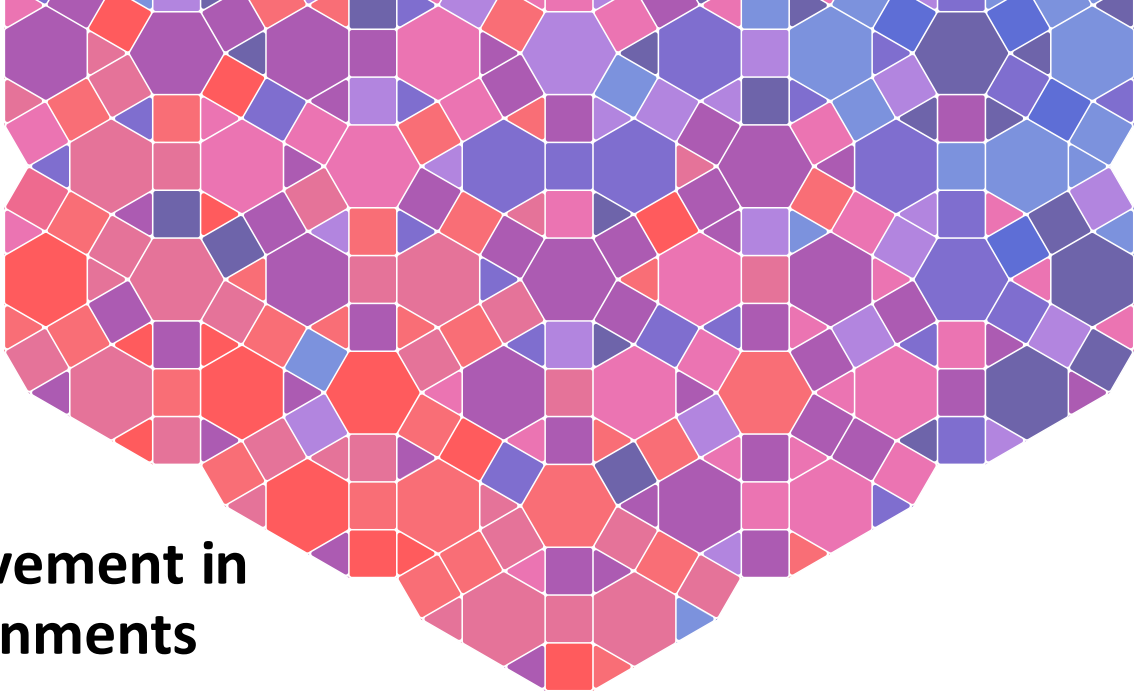
Safe Data Access Professionals



DARE UK

Ifeanyi Chukwu	How we Support AI projects on LASER TRE
Chris Appleton	MOSAIC: Accelerating health and academic research by making data integration secure, compliant, and effortless
Samaira Khan	Embedding Public Involvement in Trusted Research Environments
Dave Buckley	SyftBox – an open-source decentralised network for federated, privacy-preserving data science
Danny Silk	From concept to scaled delivery: adopting a ecosystem approach to drive health R&D in the UK
Eleajo Abubakar	Intersectional Inequality of Multimorbidity Prevalence in Older Adults due to Personal and Residential Circumstances
George Svarovsky	One year of UCL Advanced Research Computing's new TRE
Jamie Kidd	Assessing ML Models in the Scottish National Safe Haven using Medical Images
Jenny Johnston	SATRE Control Alignment - Mapping Exercise
Kostas Kavoussanakis	How to pursue federated access to sensitive data
Chris Wood	Metadata Enhancement
Lewis Hotchkiss	The UK Synthetic Data Community Group
Martin O'Reilly	Federated Research Infrastructure by Data Governance Extension
Tim Machin	SATRE 2.0 Update

Samaira Khan



Embedding Public Involvement in Trusted Research Environments

Introducing PEDRI Good Practice Standards, Planning,
Reflection and Evaluation Toolkit

Samaira Khan (*Senior Partnerships Manager and Operational Lead*)





Equity, diversity and inclusion

Ensure representation of people from different backgrounds.



Data literacy and training

Empower the public to understand and contribute.



Two-way communication

Enable everyone to have open, honest, and clear conversations.



Transparency

Project information is freely accessible for discussions with the public.



Mutual benefit

There is benefit to everyone involved.



Effective involvement and engagement

There are clear tasks, purpose and impacts that are not tokenistic.



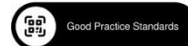
Creating a culture of involvement and engagement

Public involvement and engagement is embedded into ways of working.



“It’s not about getting people to trust you; it’s about demonstrating trustworthiness”

Public Contributor



Our journey so far



Practical
examples



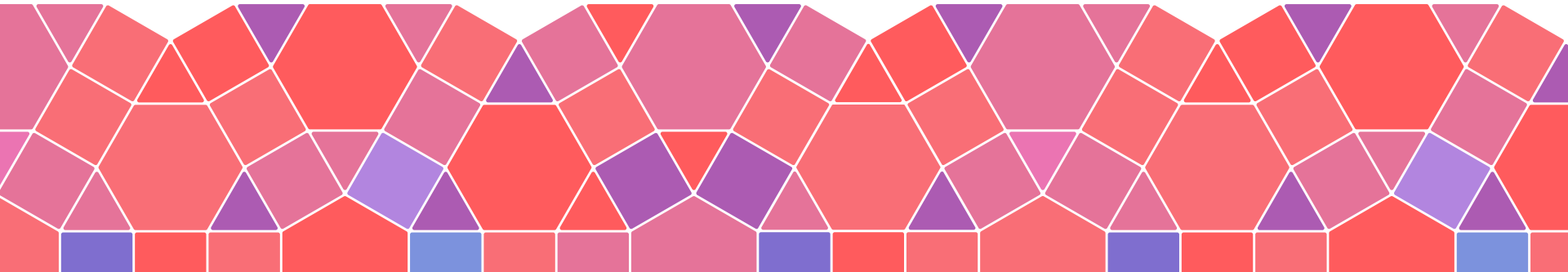
Journey
snapshot




Full journey
log



Guidance



PEDRI Toolkit - Guidance

Item	Guidance
Standards	Which standard(s) do your objective relate to?
Objective	What do you want to achieve?
Key results	How will you know you are making progress? Can you name specific, measurable things like numbers, percentages, or milestones?
Activities	What practical steps or tasks will you do to hit those key results?
Timeline	Can you specify dates or timeframes (e.g. April–June 2025, or Q1 2025)?
Context	What needs to be in place for this to work? Does something else needs to happen first or at the same time? Who else is involved?
Impact	What did you achieve
Diary log (including public partners)	Did anything change? Why? What went well? What was difficult?
 Footer:	What would you do differently?

Working version — format and content not yet final

Example: Public Involvement in Early-Onset CVD Research (Aged 30-50)



Researcher: Dr. Alex, University of Manchester

Project: Shaping cardiovascular disease research using linked health data in a Trusted Research Environment (TRE)

Datasets: BHF Data Science Centre (BHF DSC) linked datasets

Supported by: PEDRI, BHF DSC, community health networks

Showcasing the Good Practice Standards Toolkit in Action



Section	Details
Objective	<ul style="list-style-type: none">Engage adults 30-50 to shape & understand CVD research in a TRE
Standards applied	<ul style="list-style-type: none">Equity, Diversity & Inclusion – recruit diverse contributorsData Literacy & Training – tailored session on TRE & data structureEffective Engagement – clear, non-tokenistic roles (e.g. research questions, reviewing outputs)
Activities & Key Results	<ul style="list-style-type: none">4-6 contributors recruited1.5 hr data literacy session2 engagement sessions: design input + findings review1 plain-language summary co-reviewed1-2 design changes made
Timeline	Oct 2025 – recruit Nov 2025 – training Dec 2025 – session 1 Feb 2026 – session 2 Mar 2026 – publish outputs
Expected Impact	<ul style="list-style-type: none">Research questions refined to include social & occupational stressorsBetter public understanding of TREs leads to increased trustContributors felt heard; feedback led to changes in study design & outputsProcess documented to guide future TRE-based CVD research

Benefits to Alex, UoM and data providers:

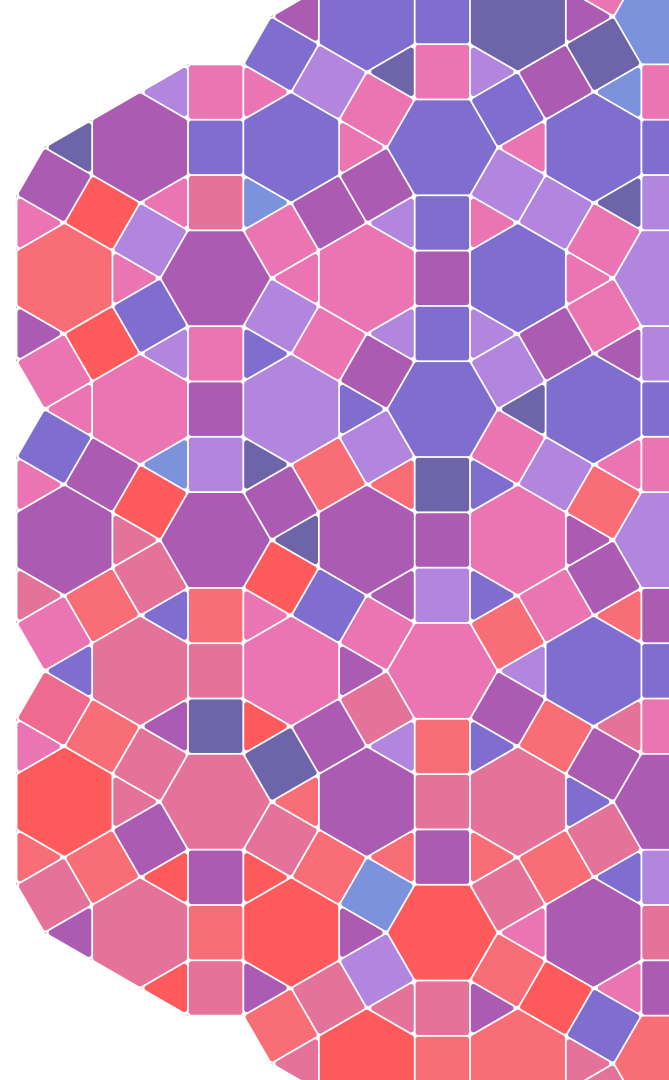
Clear **audit trail** for funders and ethics bodies.

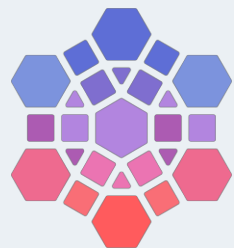
Builds **trust** with data contributors and the public.

Creates a **legacy of learning** for future projects.

Want to find out more?

- Visit our website: <https://www.pedri.org.uk/>
- Explore our membership options:
<https://www.pedri.org.uk/get-involved/memberships/>
- Contact us: contact@pedri.org.uk





PEDRI

Dave Buckley



OpenMined

We are a 501(c)(3) non-profit foundation and mission-driven community that brings together leading minds in AI, security and privacy technology to develop open-source solutions that transform how data can be accessed and shared.

AISI AI SAFETY
INSTITUTE

reddit

Microsoft

DAILYMOTION

DeepMind



Meta

Google

UNIVERSITY OF
OXFORD

GOV.UK

UN PET Lab
Artificial Intelligence
Privacy Enhancing Technologies Lab

COMMON
CRAWL

<https://openmined.org>

Attribution-based control

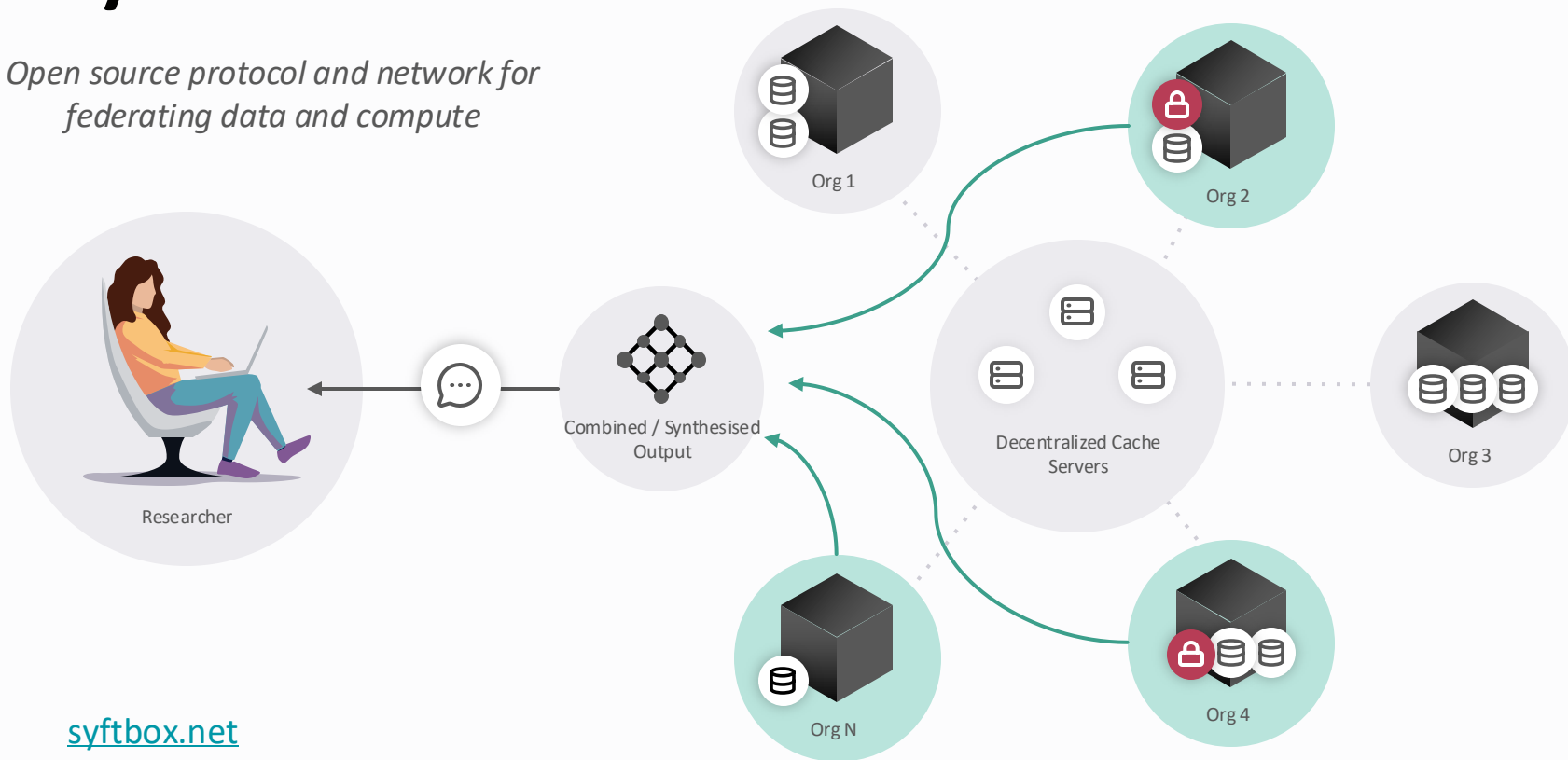
Data owners control what analyses and computations their data supports

Researchers search, discover and choose which sources they use to inform their research outputs

openmined.org/attribution-based-control

Syftbox

*Open source protocol and network for
federating data and compute*




syftbox.net









 main

Go to file








 Code

About

khoaguin import notebooks 0b1d765 · 3 days ago 


 .github/workflows	fix: update version retrieval to use ...	3 days ago
 docs	small fix	last month
 notebooks	import notebooks	3 days ago
 scripts	control encryption with env variab...	last month
 src/syft_flwr	bump: version 0.3.1 → 0.3.2	3 days ago
 tests	refactor	3 weeks ago
 .gitignore	control encryption with env variab...	last month
 .pre-commit-config.yaml	fixes	6 months ago

Privacy-preserving distributed computing framework for federated learning, analytics, and RAG with secure job submission, approval workflows, and offline capability via SyftBox file sync.

-  Readme
-  Apache-2.0 license
-  Code of conduct
-  Contributing
-  Activity
-  Custom properties
-  11 stars

Federated Learning in Practice: Training a Diabetes Prediction Model Across Distributed Datasites

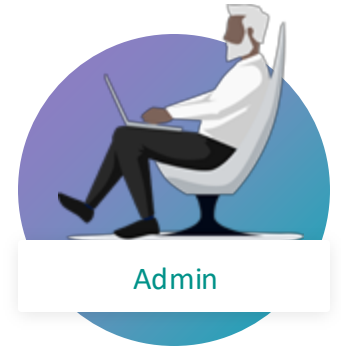
3 months ago • Product • 14 min

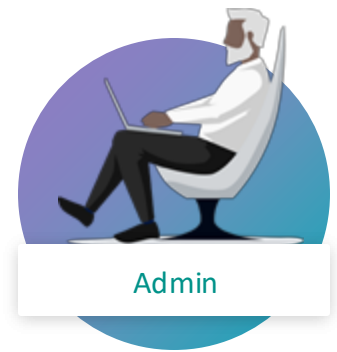
FL + 
part 1

Access tutorial



openmined.org/blog





1

Launch datasite

2

Load data

3

... goes and has a coffee...
...(or tea)





1

Launch datasite

2

Load data

3

... goes and has a coffee...
...(or tea)



1

Launch datasite

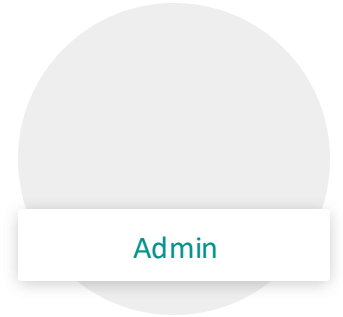
2

Load data



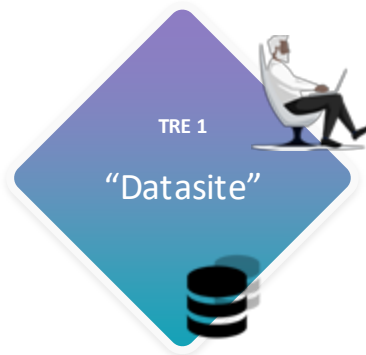
3

... goes and has a coffee...
...(or tea)



bye!

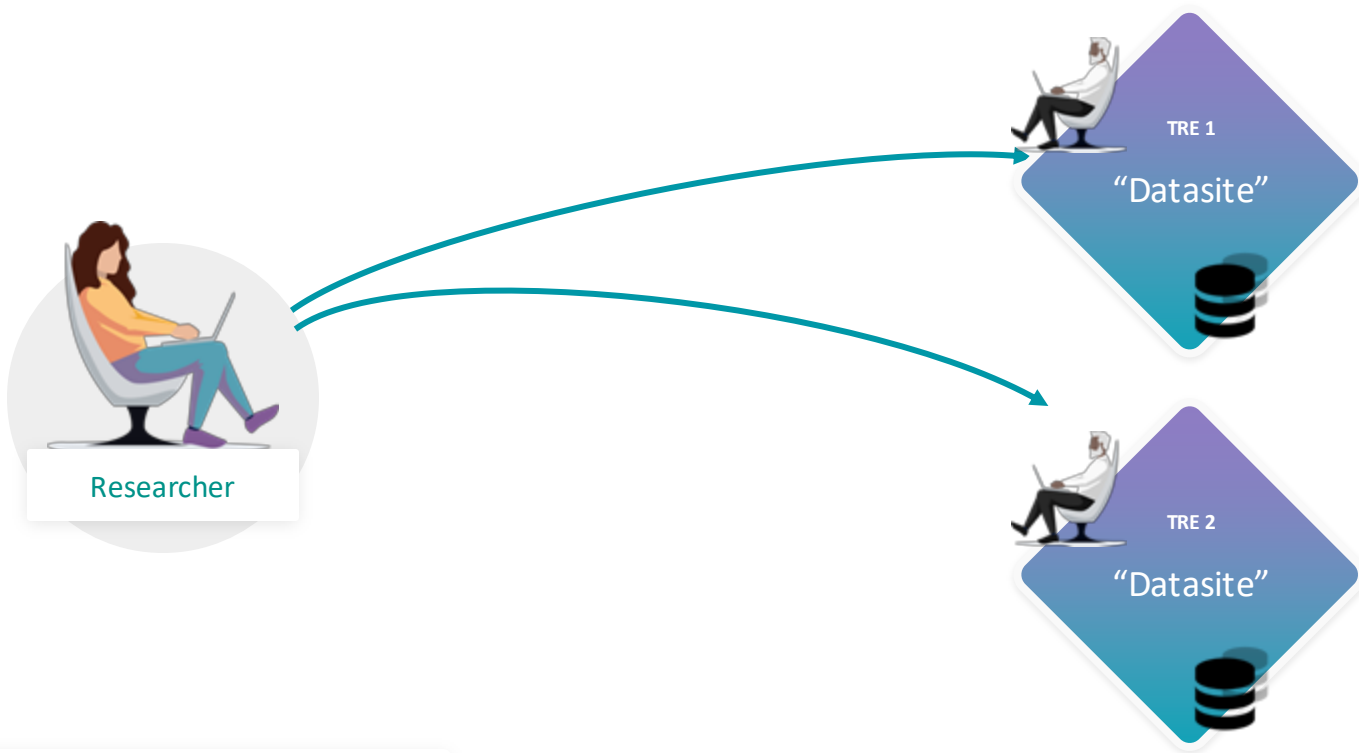




Same goes for all data owners...







1

Discover & connect
to datasites

2

Submit research
code

3

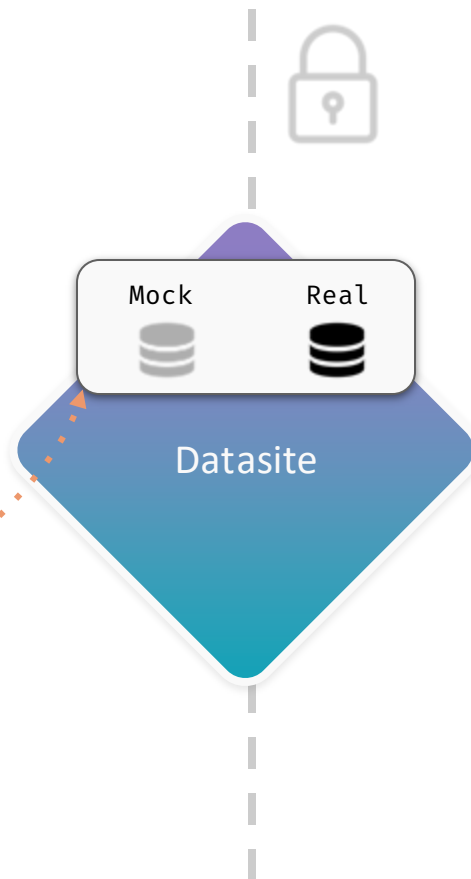
Download Answers



Open-access
mock data



Researcher



Securely hosted
private data



Admin





Researcher



1

Discover & connect
to datasites

2

Submit research
code

3

Download Answers





1

Discover & connect
to datasites

2

Submit research
code

3

Download Answers





Researcher

1

Discover & connect
to datasites

2

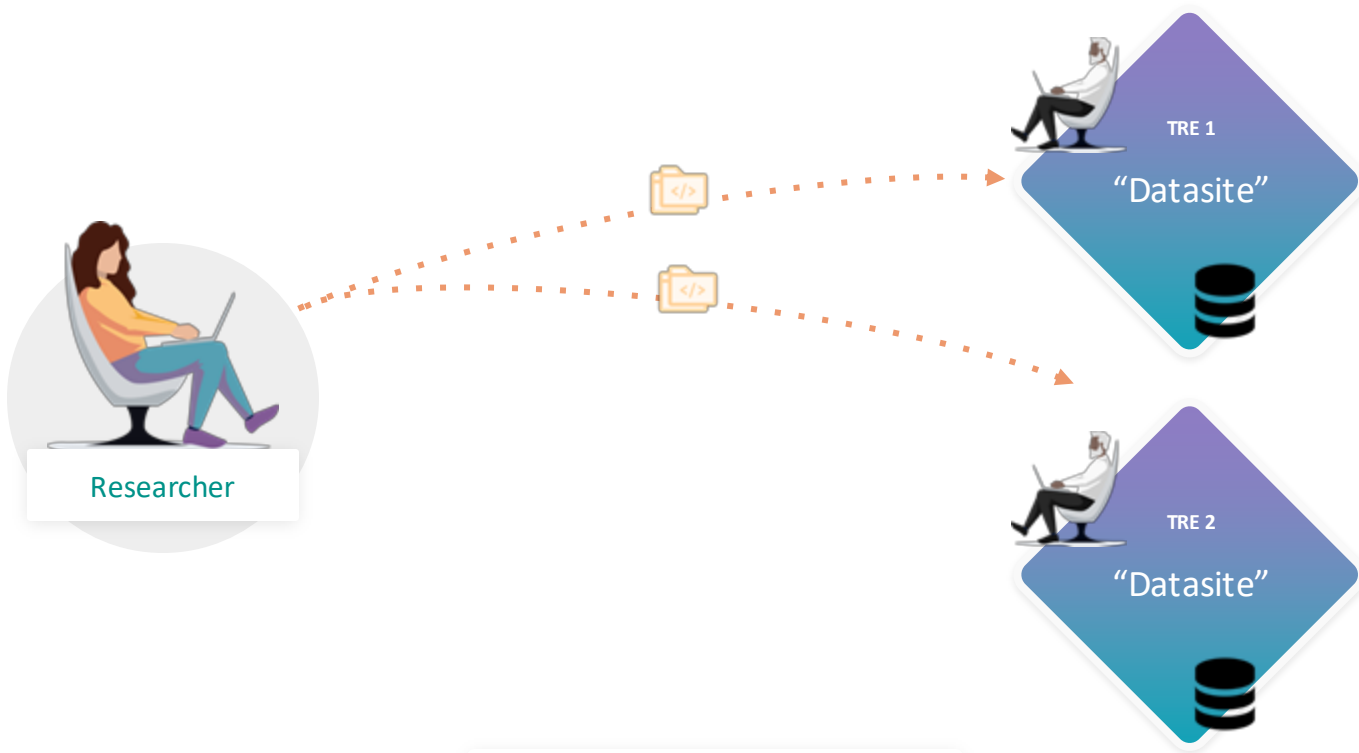
Submit research
code



3

Download Answers





1

Discover & connect
to datasites

2

Submit research
code

3

Download Answers





1

Launch datasite

2

Load data

3

Code review and approval





1

Launch datasite

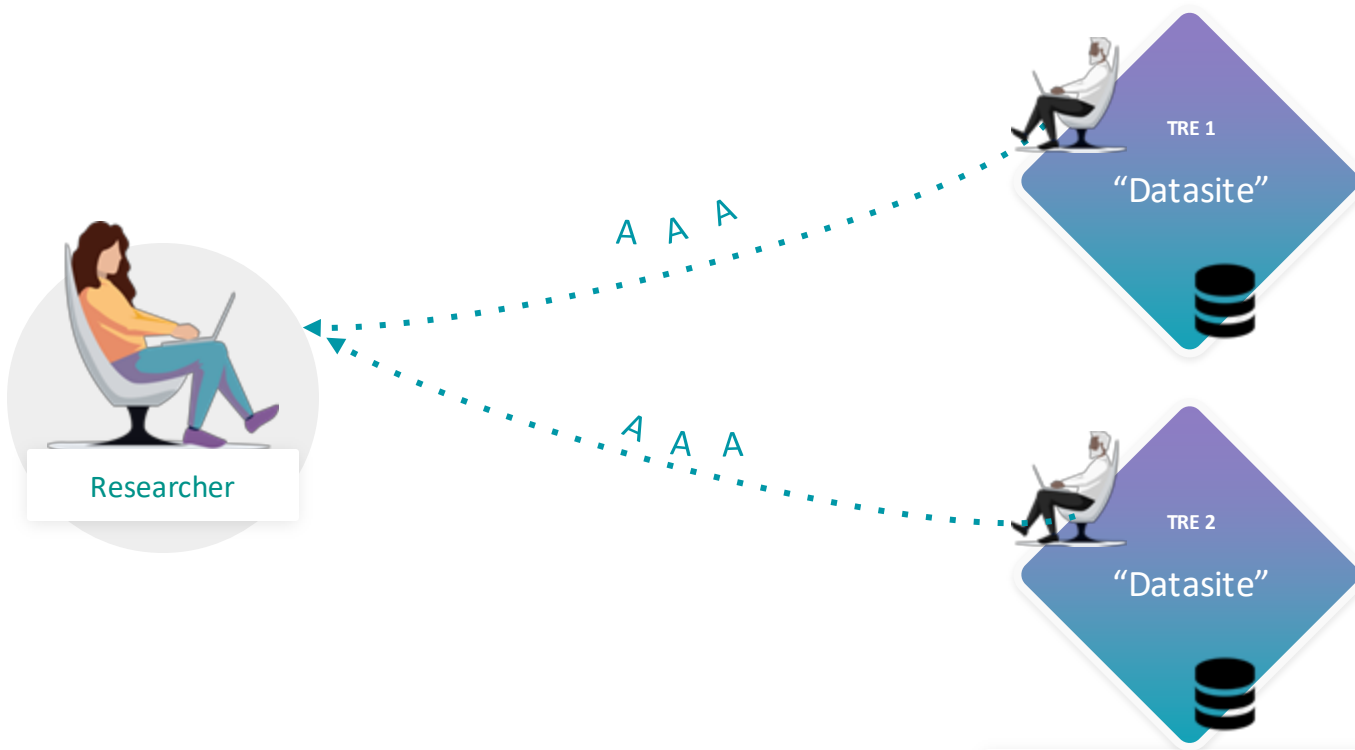
2

Load data

3

Code review and
approval





1

Login to Domain
Server

2

Get answers to
allowed questions

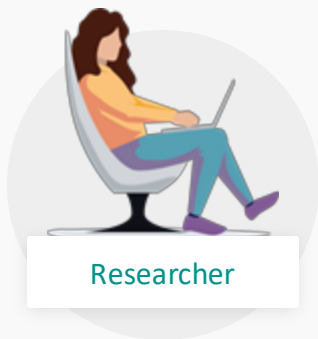
3

Download Answers



Here, a researcher can study
sensitive data remotely without
accessing it.

What if we need that information to be securely joined/linked?



1

Login to Domain
Server

2

Get answers to
allowed questions

3

Download Answers

APPROVED!



TRE 1

"Datasite"



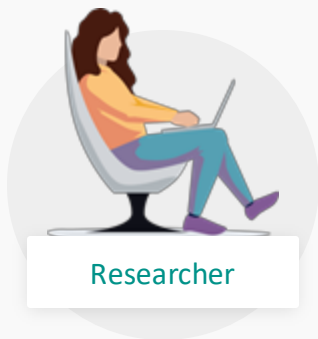
APPROVED!



TRE 2

"Datasite"





To run the code, the data
needs to be co-located
on the same computer

APPROVED!



APPROVED!



1

Login to Domain
Server

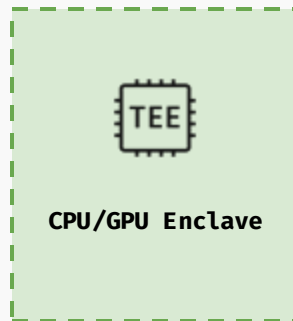
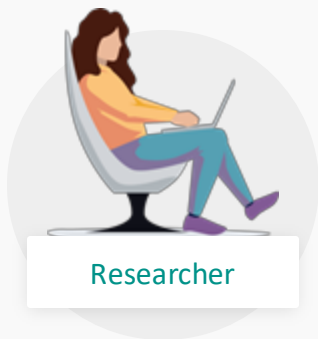
2

Get answers to
allowed questions

3

Download Answers





1

Login to Domain
Server

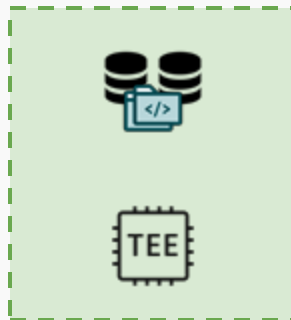
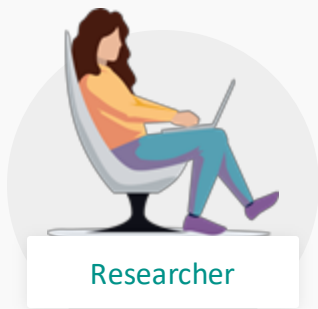
2

Get answers to
allowed questions

3

Download Answers





1

Login to Domain
Server

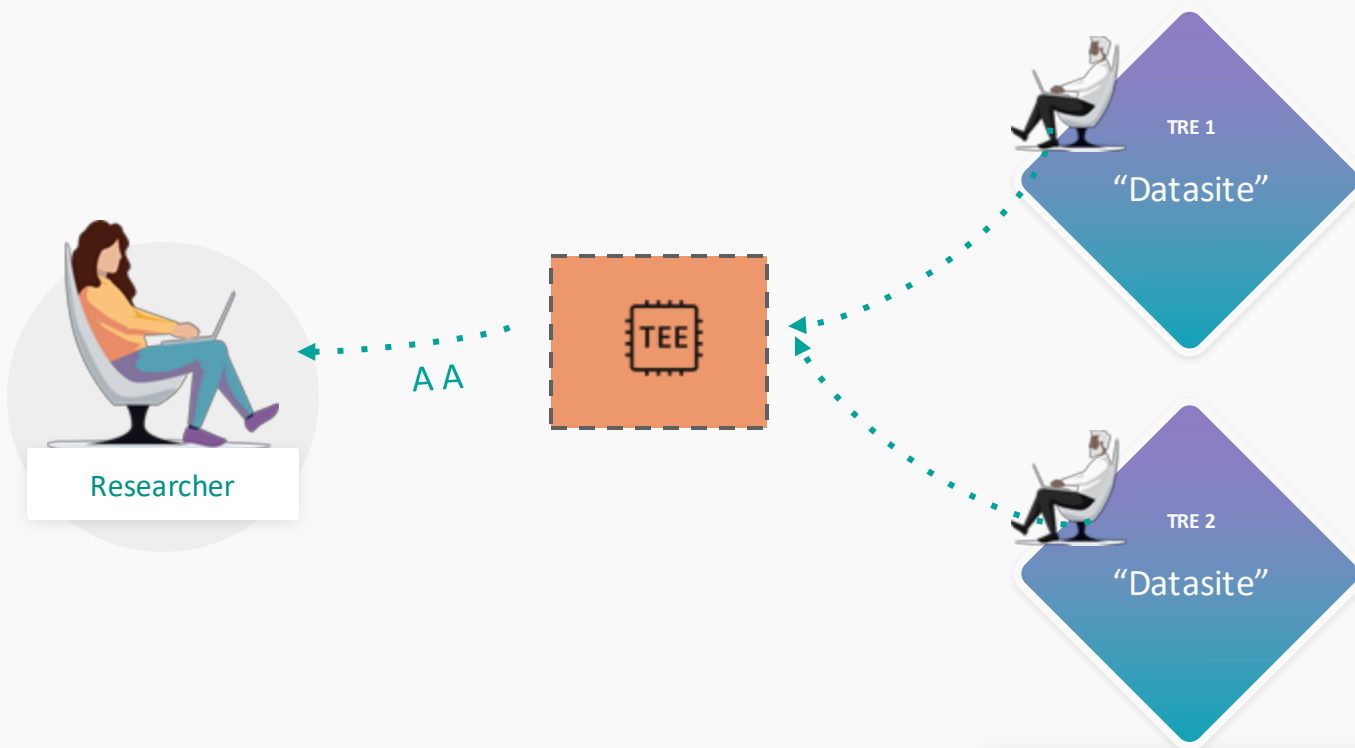
2

Get answers to
allowed questions

3

Download Answers





1

Login to Domain
Server

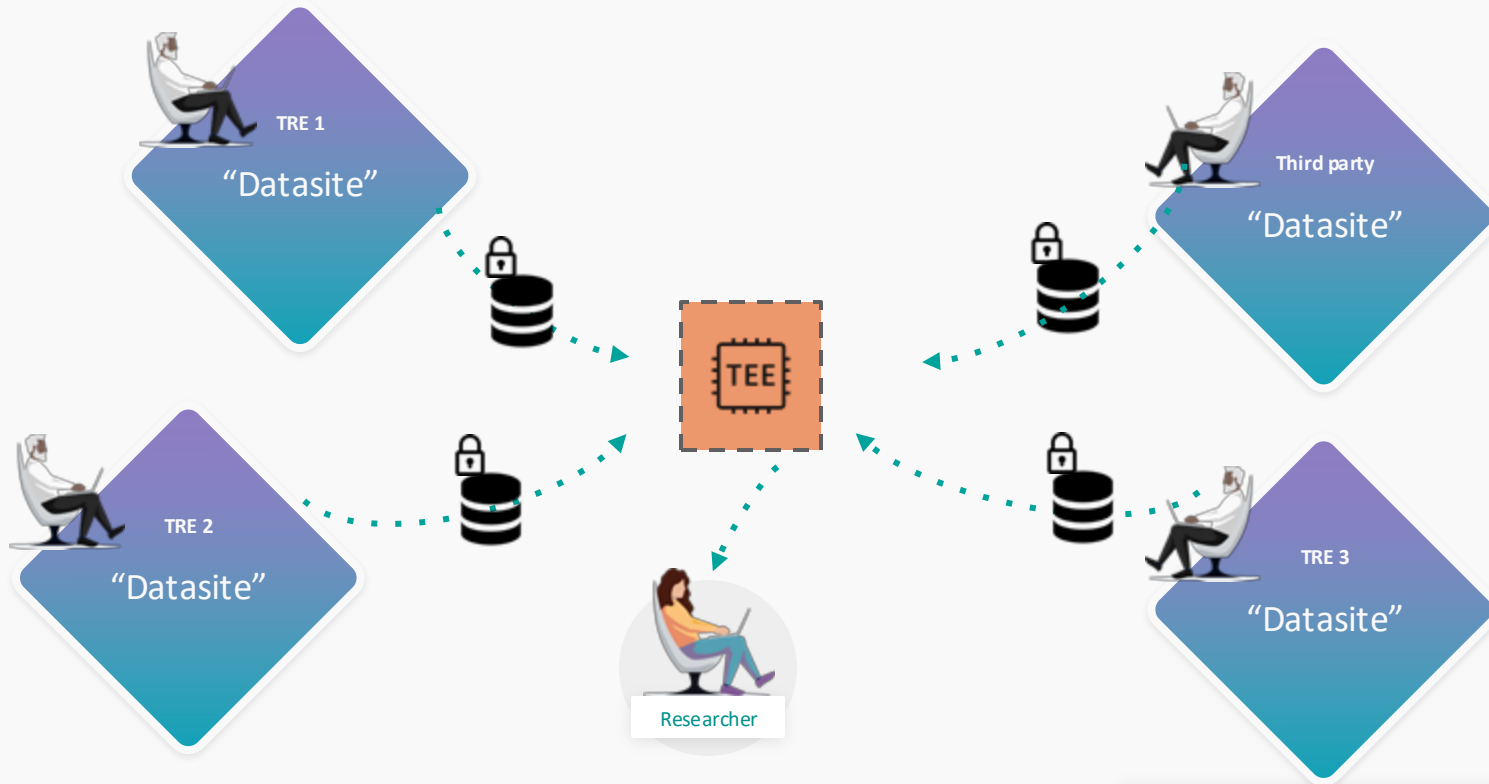
2

Get answers to
allowed questions

3

Download Answers





1

Login to Domain
Server

2

Get answers to
allowed questions

3

Download Answers



Now, a researcher can study data
from multiple data sources
without accessing it + none of the
data owners can see the other
party's assets.

Engage with us



Email:

- dave@openmined.org

Danny Silk

No slides, audio only

George Svarovsky

One year of UCL Advanced Research Computing's new TRE

Lightning Talk

The ARC TRE – Recap of Goals

Compute platform agnostic

- Built to “TREify” ARC’s research computing platforms & services

Maximise automation & self-service

Focus on reproducible science

Reusable TRE-in-a-box

*General availability of the ARC TRE for UCL research is planned for
was delivered in November 2024*

BYO Data *and* Software

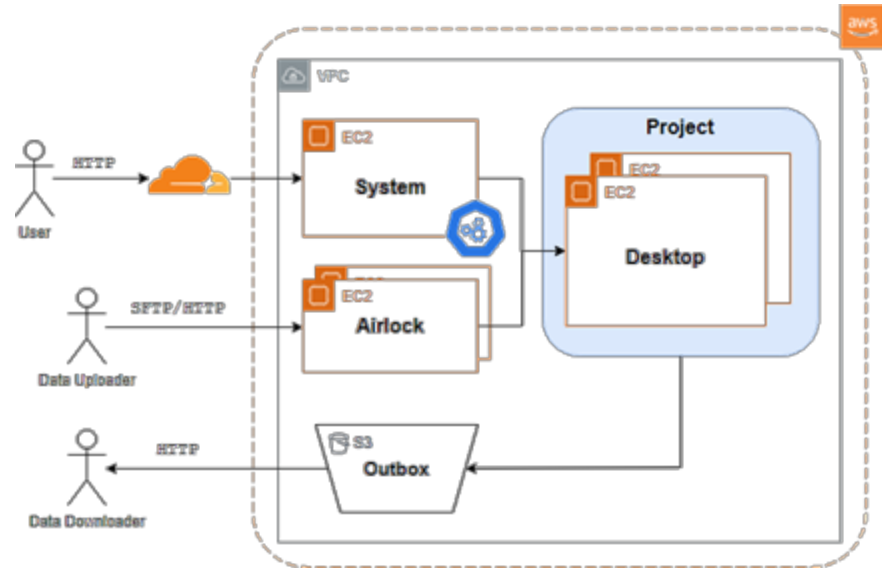
Not a Data Provider (“SDE”)

cf. NHS, CPRD, OFH

Instead, a multi-tenant TRE of TREs
 (“Projects”)

Shared Responsibility Model with
 researchers, platform & organisation

Choose-your-own compute (priced)



Status

“Overall everything is working very well, it’s been a good experience and I’m pleased to be able to get things up and running so quickly.”

- 👍 ISO 27001, DSP Toolkit through shared ISMS
- 🚢 Agile: soft launch, 16 active research Projects, active DevOps
- ☀️ On AWS; private cloud soon (test environment this year)
- 📖 Open-source portal for sign-up & IG
- 📄 Exemplar for K8TRE specification compliance
- 🤝 Shared-source – come and chat!

Search ucl arc tre

arc.tre@ucl.ac.uk

<https://isms.arc.ucl.ac.uk/>

<https://github.com/ucl-arc-tre/portal>

Jamie Kidd

Assessing ML Models in the Scottish National Safe Haven using Medical Images

Risk-based governance of machine learning models trained on real-world medical images

Dr Jamie Kidd

Scottish Medical Imaging (SMI) Project Overview

Comprehensive Imaging Archive

SMI project has created a vast archive of over 57 million radiologic exams with 2.4 billion images.

Diverse Imaging Modalities

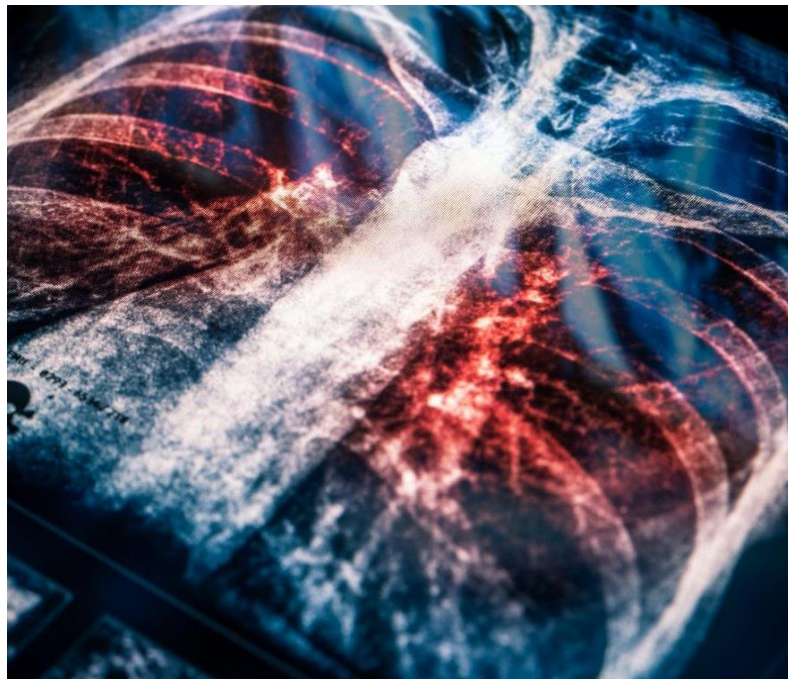
The archive includes a wide variety of imaging types such as CT scans, MRI, and X-rays for extensive research use.

Secure Data Storage

Images are securely stored in the Scottish National Safe Haven ensuring data privacy and protection.

Advancing Machine Learning Research

This infrastructure supports building clinically relevant machine learning models using population-scale imaging data.



Retinal Imaging Repository

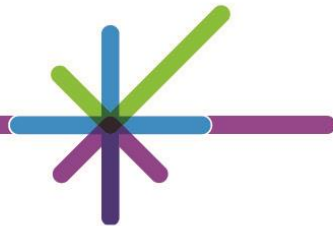


Collaboration for Retinal Data

Public Health Scotland and University of Edinburgh have created a large, dedicated retinal image repository.

Routine Retinal Imaging

Retinal scans are routinely captured during eye exams across Scotland, providing a near-universal data source.



Risk Evaluation

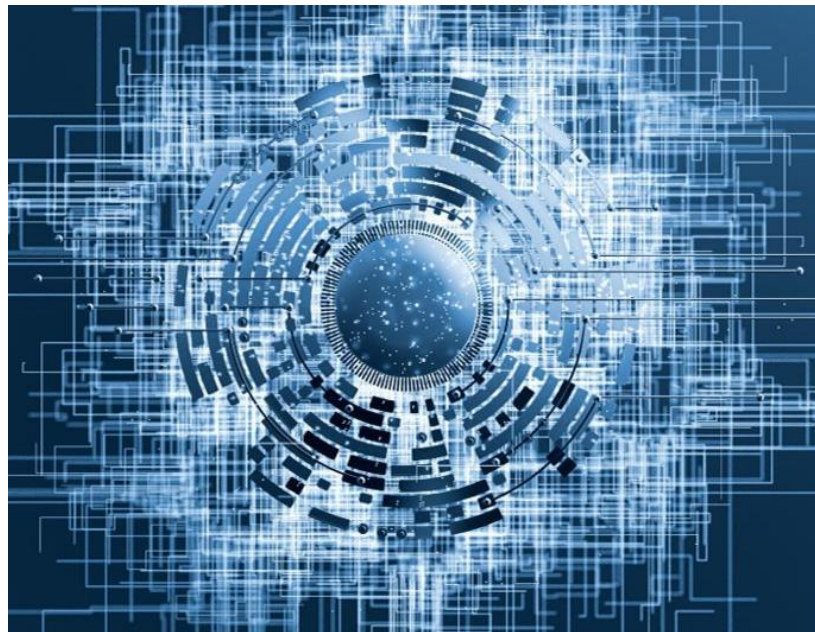
Risk Evaluation Framework

Public Health Scotland's eDRIS uses a structured risk matrix assessing likelihood and impact of adverse events.

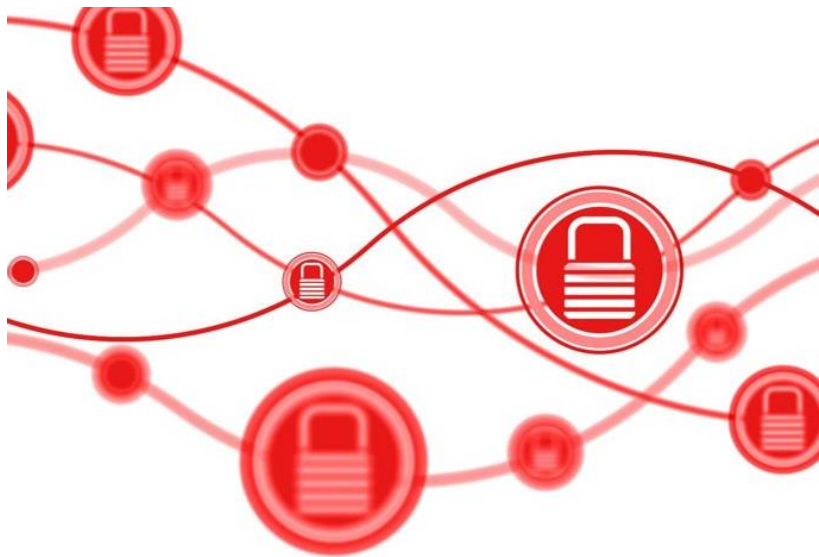
Privacy and Security Scrutiny

Models are carefully reviewed for privacy and security risks before release from the Scottish National Safe Haven.

Initial benchmark: would we release this if it were tabular data? The same principles apply



Technical Risk Considerations



Decoder Component Security

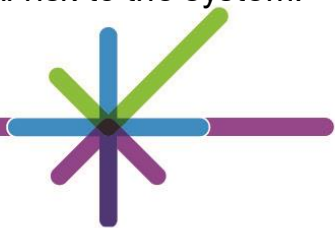
The model's decoder is kept within a secure environment to prevent unauthorised reconstruction of image features.

Preprocessing Risk Reduction

Preprocessing steps are implemented to minimise reverse engineering risks before data leaves the secure environment.

Membership Inference Attack Risks

Membership inference attacks require prior access to original images, limiting practical risk to the system.



Disease Classification and Identifiability

Generalised Retinal Models

Foundation models trained on retinal images may seem low-risk without disease-specific tuning or classification

Retinal vs Brain Scans

Retinal scans are routine and widely used, while brain scans often imply clinical concerns with higher identifiability risk.

Impact of Disease Classification

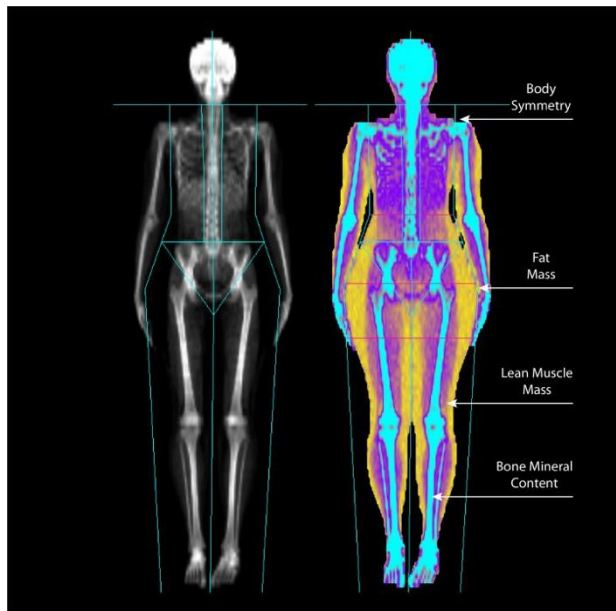
Including disease classification in models significantly changes the risk profile related to patient privacy and data sensitivity.

Contextual Identifiability

Identifiability is not binary; it depends on context and evolves with advancements in medical imaging and technology.



Why a Universal Approach Doesn't Work



DEXA Scan Results

Limitations of Universal Models

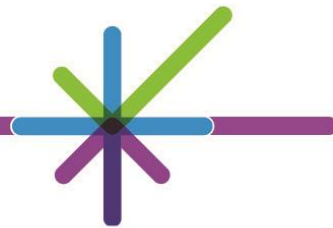
A single model assessment approach fails to address the unique characteristics of diverse medical imaging modalities.

Significance of DEXA Scans

DEXA scans are less common but provide critical insights into bone density and health risks like osteoporosis.

Context and Privacy Risks

Image context can reveal sensitive health information, necessitating specialized risk assessments for each modality.



Risk Mitigation vs. Model Performance



Privacy-Preserving Techniques

Adding noise or distorting features reduces identifiability risk but may impair model accuracy and effectiveness.



Risk and Performance Trade-off

Lower-risk scenarios allow reduced privacy measures, enabling better model performance and utility.



eDRIS Framework Approach

eDRIS framework balances adverse event risks against benefits to ensure privacy without compromising model utility.



Current Evaluation and Release Rates

Model Evaluation Process

Each model is assessed individually using a risk matrix to ensure thorough evaluation before release.

Approval Rate Statistics

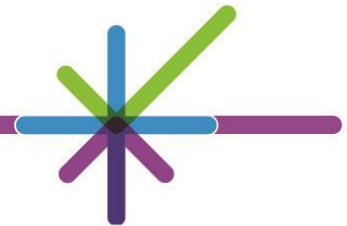
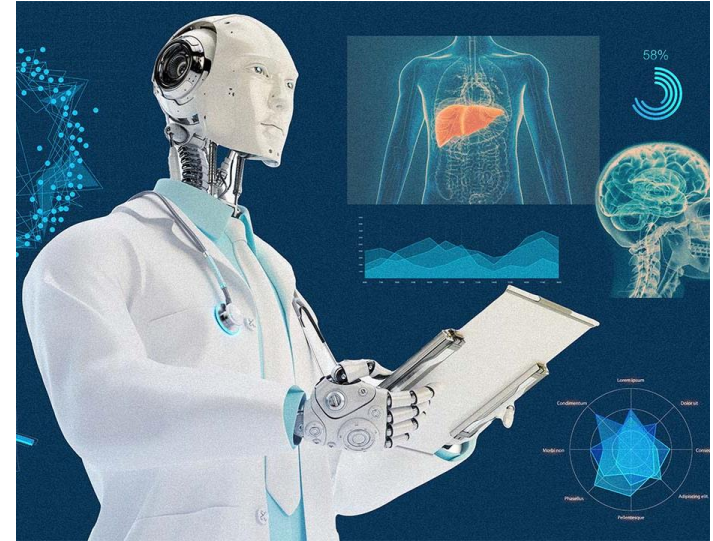
Currently, 50% of models that reach the release request stage are approved, reflecting cautious progress.

Balancing Innovation and Privacy

The framework balances enabling technological innovation with safeguarding user privacy effectively.

Future Risk Anticipation

Framework is designed to adapt and anticipate future machine learning risks as technology evolves.



Jenny Johnston



Health Informatics Centre
University of Dundee

SATRE

Short Life Working Group

Control Alignment - Mapping Exercise

Presented by:
Jenny Johnston



INTRODUCTION

This project set out to address the fragmented assurance landscape that Trusted Research Environments (TREs) face across the UK. By using SATRE Framework as the anchor, the project developed a SATRE-Based Control Alignment Table to map controls from five key assurance and regulatory frameworks:

- ISO/IEC 27001:2022
- Digital Economy Act (DEA) Accredited Processor Requirements
- NHS Data Security and Protection Toolkit (DSPT)
- SDE Self-Assessment Framework
- Cyber Essentials



Members

Short Life Working Group

Project Sponsor

Chris Cole / Gordon Milligan

Project Lead

Jenny Johnston

Subject Matter Expert(s)

Symone Sheane, Health Informatics Centre

Emily Almond, Daniel Higgins, Department of Health and Social Care

Rhys Nadin, Jo-Anna Hagen, UKStats

Anne Whiting, EPCC

Tom Billins, Health Innovation East



PROJECT OBJECTIVES



Identify Overlap

To identify and analyse overlapping assurance requirements across frameworks, revealing opportunities for streamlining. Laying the groundwork for a more efficient and scalable governance model for research.



Identify Differences

To highlight the unique value and intent behind each accreditation or standard, emphasising their role in maintaining a secure, ethical, and trusted research landscape.



Control Alignment Table

To develop a shared control alignment table that promotes transparency, reduces duplication, and promote community-wide agreement on common governance measures, ensuring we don't reinvent the wheel but build collaboratively toward consistent, trusted assurance pathways.

SCOPE OF WORK

ISO 27001

DSPToolkit

SDE Self Assessment

DEA

Cyber Essentials

CONTRIBUTION & COLLABORATION



Collaborative Alignment

The group helped identify both conceptual overlaps and terminological differences across standards enabling us to have a common understanding.



Cross-Framework Expertise

Members brought experience with implementing and interpreting ISO 27001, DEA, DSPT, Cyber Essentials and the SDE Self-Assessment Framework within TRE settings.



Gap Analysis Support

Through expert dialogue, the SLWG helped articulate where SATRE uniquely adds value and where other standards leave implementation gaps.



Accelerated Mapping

The SLWG's domain knowledge allowed for swift identification and interpretation of controls across standards..

BENEFITS AND IMPACT



Strong Alignment

A high degree of alignment was found between SATRE and ISO 27001 and DSPT in areas such as access control, information security governance, and incident management.



DEA Specifications

The DEA introduces specific data processor and accountability provisions that are flagged in the alignment table for SATRE-based implementations and will support the evolution of the Data Capability within DEA.



SDE Synergies

The SDE framework shares several thematic areas with SATRE, especially around data lifecycle and user accountability.



SATRE Builds on Existing Frameworks

SATRE builds on existing standards rather than replacing them, adding TRE-specific expectations that enhance operational effectiveness, maturity, and confidence.



Control Alignment Table

A comprehensive resource aligning controls from five major frameworks against SATRE's structure. This alignment table can now be published for TRE teams to utilise and feedback onto support continuous improvement which will increase usability and implementation.



CONTROL DATA





SUMMARY

✅ Strong Alignment with Existing Standards

- Most SATRE controls align with standards like ISO 27001, DSPT, SDE and DEA
- Confirms shared foundations in: Information Governance; Security; Data Management
- Organisations already compliant with major frameworks are largely SATRE-ready

🌱 SATRE Adds Unique Value

- 22 unmatched controls address practical gaps:
 - Project-specific costing
 - Dev/test environments
 - Automation portals
- Introduces:
 - Operational maturity expectations
 - Emphasis on transparency, public trust, and collaborative functionality (critical for TREs)

NEXT STEPS

Step 1

Dissemination

Share Control Alignment Table with TRE's nationally via existing networks (e.g. SATRE community, UKRI, HDR UK).

Step 2

Tooling

Develop current digital tooling to support TRE's interactively using table for assurance planning.

Step 3

Maintenance

Establish lightweight digital peer review assessment governance process to keep the alignment table and tooling current as framework evolves.

Step 4

TRE Toolkit

Collection of materials, templates, standard processes and policies for implementing a TRE aligned with accreditation.





CONCLUSION

- SATRE builds on existing standards: not a replacement
- Enhances assurance through TRE-specific expectations
- This mapping is a starting point for good practice and will evolve with the TRE community

BIG THANKS



Kostas Kavoussanakis

HOW TO PURSUE FEDERATED ACCESS TO DATA

Connect4

Kostas Kavoussanakis,
EPCC,

The University of Edinburgh

k.kavoussanakis@epcc.ed.ac.uk

What did we do?

The problem: Processing data across national TREs is beyond the practical reach of researchers

- Metadata are not descriptive enough of such datasets, so the researcher cannot assess whether a combination is feasible or leads to a sensible analysis
- The TRE policies that govern access are not compatible with each other and generally do not consider federated access.

Connect4

- Surveyed potential users and developed software that creates enhanced metadata for datasets in TREs, including summary statistics, proportion of missing data and maps showing simple geospatial analyses.
- Attempted a use case using federated data from the ONS Integrated Data Service and the Scottish National Safe Haven and proposed a nine-prong roadmap and a shared-service model to streamline federated access to sensitive data.

Metadata generator: <https://github.com/epcced/connect4-metadata>

```
PLAYING_PER_1000: {+ ...},
PSYCHIATRIC_CONDITION_PER_1000: {+ ...},
ROAD_TRAFFIC_ACCIDENT_PER_1000: {-
  count: 136,
  mean: 4.140919571250001,
  std: 1.8118658130475336,
  min: 0.96324461,
  25%: 3.033335715,
  50%: 4.00634288,
  75%: 5.22810703,
  max: 9.37323546,
  range: 8.40999085
},
RTA_PEDESTRIAN_PER_1000: {+ ...},
SPORTING_INJURY_PER_1000: {+ ...},
SURGICAL_CONDITION_PER_1000: {+ ...},
UNSPECIFIED_PER_1000: {+ ...},
NO_OF_TOP_5S_ADJUSTED: {+ ...},
SHAPE__AREA: {+ ...},
SHAPE__LENGTH: {+ ...},
SHAPEAREA: {+ ...},
SHAPELEN: {+ ...},
INTERZONE: {+ ...},
NAME: {+ ...},
ACORN_GROUP: {+ ...}
},
```

```
data_quality: {-
  missing: {-
    OBJECTID: 0,
    INTERZONE: 0,
    NAME: 0,
    TOTPOP2011: 0,
    RESPOP2011: 0,
    HHCNT2011: 0,
    STDAREAHA: 0,
    STDAREAKM2: 0,

    ...

    UNSPECIFIED_PER_1000: 0,
    NO_OF_TOP_5S_ADJUSTED: 86,
    ACORN_GROUP: 116,
    SHAPE__AREA: 0,
    SHAPE__LENGTH: 0,
    SHAPEAREA: 0,
    SHAPELEN: 0,
    geometry: 0
  },
},
```

```
completeness: {-
  OBJECTID: 100,
  INTERZONE: 100,
  NAME: 100,
  TOTPOP2011: 100,
  RESPOP2011: 100,
  HHCNT2011: 100,
  STDAREAHA: 100,
  STDAREAKM2: 100,

  ...

  UNSPECIFIED_PER_1000: 100,
  NO_OF_TOP_5S_ADJUSTED: 36.76470588235294,
  ACORN_GROUP: 14.705882352941178,
  SHAPE__AREA: 100,
  SHAPE__LENGTH: 100,
  SHAPEAREA: 100,
  SHAPELEN: 100,
  geometry: 100
},
```


Roadmap: <https://zenodo.org/records/15516094>

- *Maintain a labelled and visible data catalogue*: as clearly tagged in the respective metadata catalogues.
- *Agree standard project criteria*: aligned with the Research Code of Practice and the guidance of the Research Accreditation Panel.
- *Provide one front-door for federated studies*: Either single or local, federated TRE.
- *The host TRE is the approver*: its approval pathway will be used, with mutual acceptance of pathways.
- *The host TRE supports the researcher*.
- *Manage a unified indexing service*: An agreed cross-nation collaborative approach or development of a UK-wide index held and maintained in one TRE to successfully match, index and enable linkage of federated data.
- *Minimise data movement*: A new technical solution to allow access to federated data with minimal data movement is encouraged over data deposits to the host TRE.
- *Establish parameters for automated output disclosure checks*.
- *Cost recovery at access*: The front-door service is responsible to create quotes using per-TRE, per-federated study cost models.



THANK YOU

Funded by the Economic and Social Research Council, under the “Future data services: pilots to enhance data services for the future” programme (ES/Z502972/1).



Chris Wood



THE UNIVERSITY
of EDINBURGH

Metadata Enhancement

Increasing visibility of datasets stored in TREs

KARA MORAW, CHRIS WOOD

EPCC, UNIVERSITY OF EDINBURGH

Reminder of metadata!



THE UNIVERSITY
of EDINBURGH

- Metadata is either:
 - *findable* (e.g. filename, author, last modified date)
 - *structural* (e.g. variable names, variable datatype)
- *Enhancing* metadata:
 - increase richness of data catalogues
 - improving access for researchers
 - reduce effort for everyone involved in the approvals process

Technical overview



THE UNIVERSITY
of EDINBURGH

- Extensible framework written in Python
 - easy for other developers to contribute
- Methods to create:
 - numerical summary statistics (e.g. mean, median, quartiles)
 - (some) geospatial metadata (e.g. spatial bounds; work ongoing to display on a suitable basemap)
 - LLM-powered keyword generator locally-hosted instance of the Llama LLM; no additional training; main challenge: prompt-engineering!

More thoughts



THE UNIVERSITY
of EDINBURGH

- Testing implemented using Pytest and Hypothesis
- Output to relevant formats (JSON and images)
- Currently a prototype: SDC required for exporting and adding to data catalogues
- A production version would require:
 - code running (automatically) within TREs
 - export of files (*is SDC always needed?*)
 - upload to data catalogue
 - decision on vocabularies to describe the metadata

Thanks!



THE UNIVERSITY
of EDINBURGH

 <https://github.com/epcced/connect4-metadata> (currently read-only)

CONTACT: c.wood@epcc.ed.ac.uk

ACKNOWLEDGEMENTS: Chiara Asselborn (PHS), Kostas Kavoussanakis (UoE), Jen Muir (RDS), Emily Symmons (ONS),

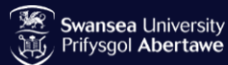
FUNDING: Economic and Social Research Council, under the “Future data services: pilots to enhance data services for the future” programme (ES/Z502972/1)

Lewis Hotchkiss

DARE UK Community Group

The UK Synthetic Data Community Group

www.syntheticdata.uk



Funded By

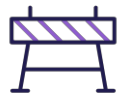


Current Challenges with Data Access in TREs



Lengthy Approval Times

Gaining access to data can involve lengthy approval processes, often taking months to be able to access. This delays research projects and creates significant inefficiencies.



Governance & Restrictions

Training and education are hindered by these restrictions, as students and early-career researchers often struggle to access real data for learning and skill development.



Limited Data Discovery

Making decisions about dataset selection without seeing the actual data can lead to wasted time and resources if, after gaining access, the dataset proves unsuitable for researchers intended analyses.



Federation

Some federated systems like TREX are “Eyes-Off” which relies on metadata to write queries. Plus, testing technological capabilities on real data can be difficult prior to governance being established.

INTRODUCTION

Meet the Co-Chairs



Lewis Hotchkiss
Dementias Platform UK



Emma Squires
Dementias Platform UK,
SeRP



Simon Thompson
Dementias Platform UK,
SAIL Databank, SeRP



John Gallacher
Dementias Platform UK,
University of Oxford



Tim Rittman
University of Cambridge



Emily Oliver
Administrative Data
Research UK



Cristina Magder
UK Data Service



Fiona Lugg-Widger
Centre for Trials Research



Robert Trubey
Centre for Trials Research



Sophie McCall
Research Data Scotland



Anmol Arora
University College London



Steve Moore
Public Representative

Stakeholder Workshops



Researcher Workshop

34 attendees



Expert Workshop

37 attendees



Data Owner Workshop

49 attendees

120 Participants across 33 Organisations Represented:

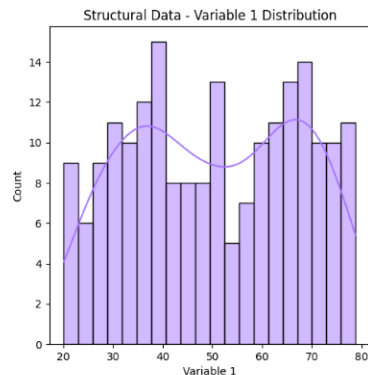
Administrative Data Research UK
Brunel University London
Cardiff University
Dementias Platform UK
Energy Systems Catapult
GSK.AI
Imperial College London
LifeArc
MHRA – Clinical Practice Research Datalink
NHS England
Public Health Wales

Queen Mary University of London
Queens University Belfast
Research Data Scotland
SAIL Databank
Scottish Government
Scottish Longitudinal Study
Swansea University
The Alan Turing Institute
UK Biobank
UK Data Service
UK Government Department for Education

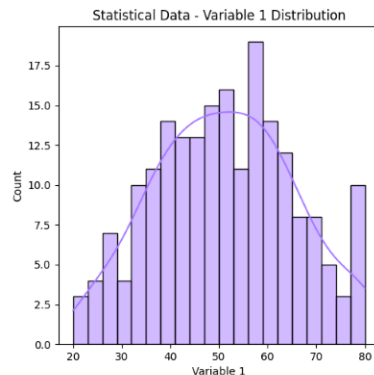
UK Government Department for Work & Pensions
UK Longitudinal Linkage Collaboration
University College London
University of Birmingham
University of Bristol
University of Cambridge
University of Edinburgh
University of Exeter
University of Manchester
University of Oxford
Welsh Government

User Requirements for Synthetic Data

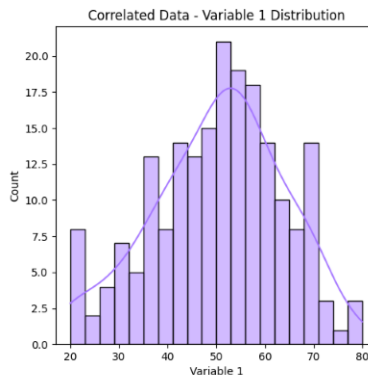
L1 Structural



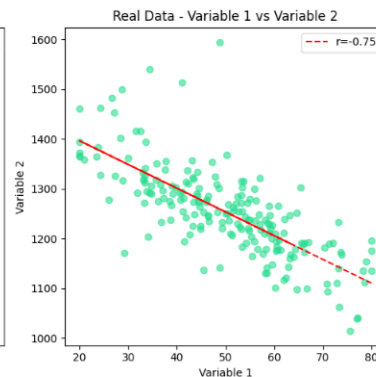
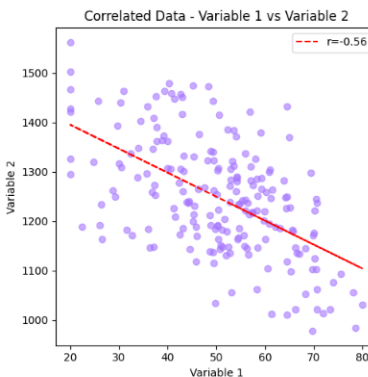
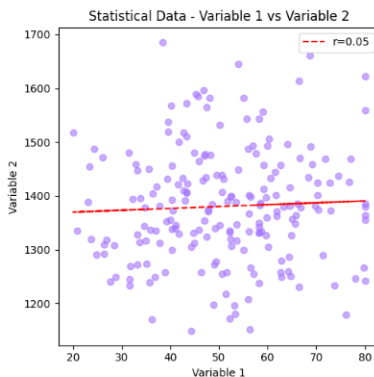
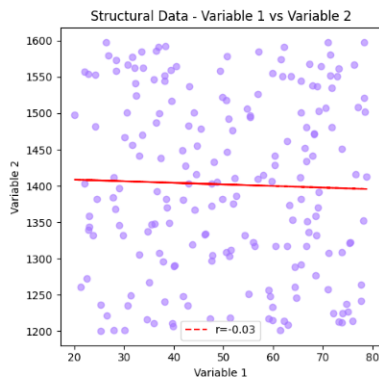
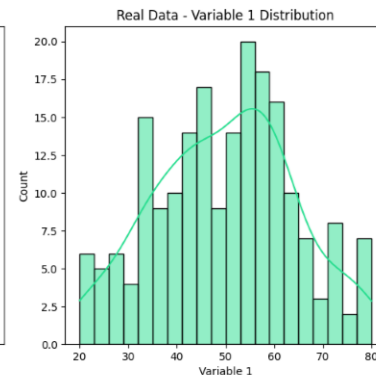
L2 Univariate



L3 Multivariate



Real

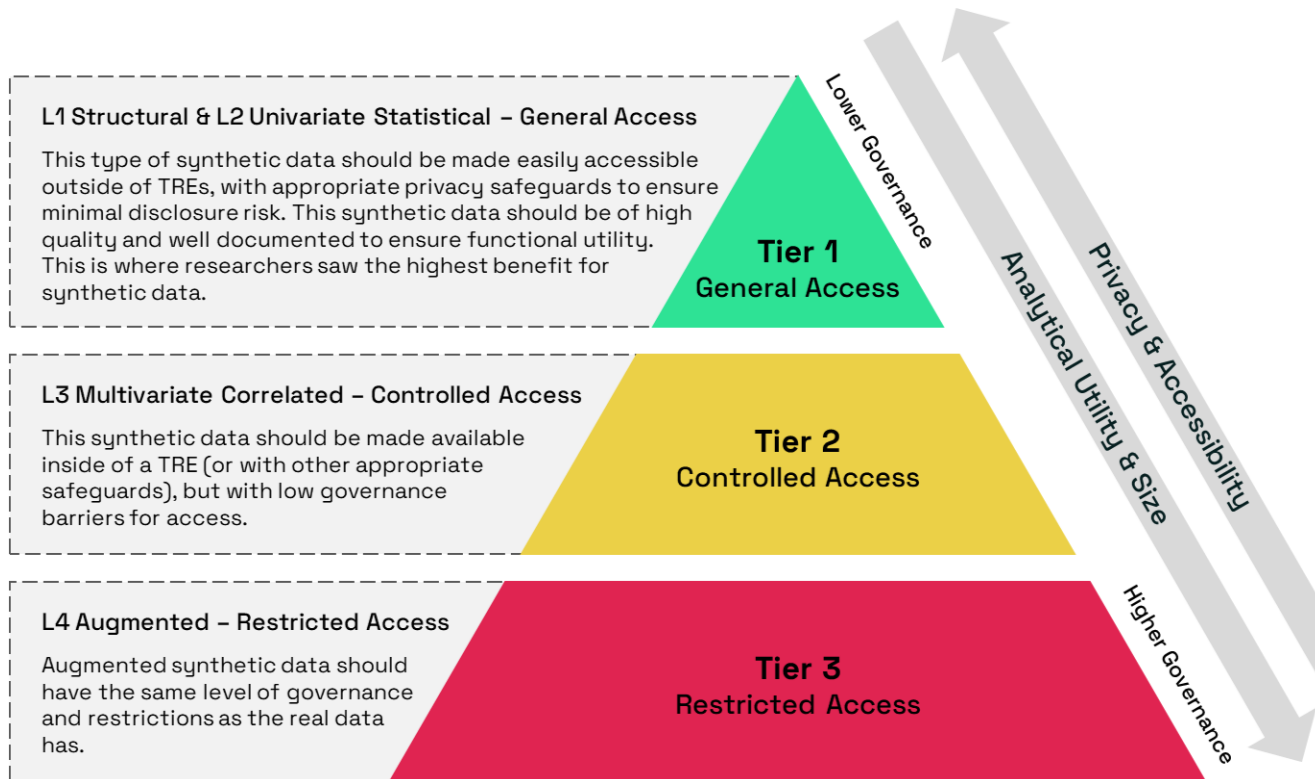
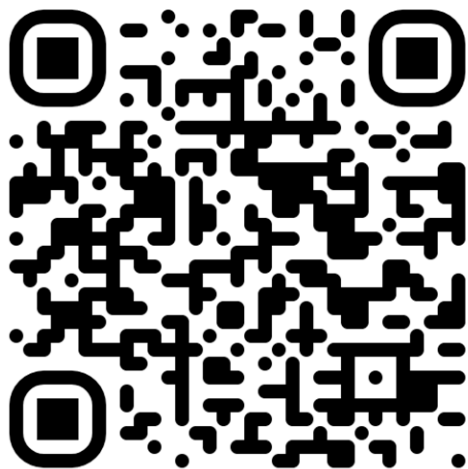


User Requirements for Synthetic Data

	Code Development	Data Discovery	Grey Area		Privacy-Preserving AI	Enhancing Analysis
			Training & Teaching	Running Federated Queries		
Level	Structural	Univariate	Univariate	Multivariate	Multivariate	Multivariate
Size	Medium	Medium	Medium	★ High	High	★ High
Access Control	★ Low	★ Low	★ Low	Medium	Medium	Medium
Documentation	★ High	★ High	Medium	★ High	★ High	★ High
Quality	★ Medium	★ Medium	★ Medium	★ Medium	High	High
Utility	Medium	Medium	★ Medium	Medium	★ High	★ High
			Fidelity			

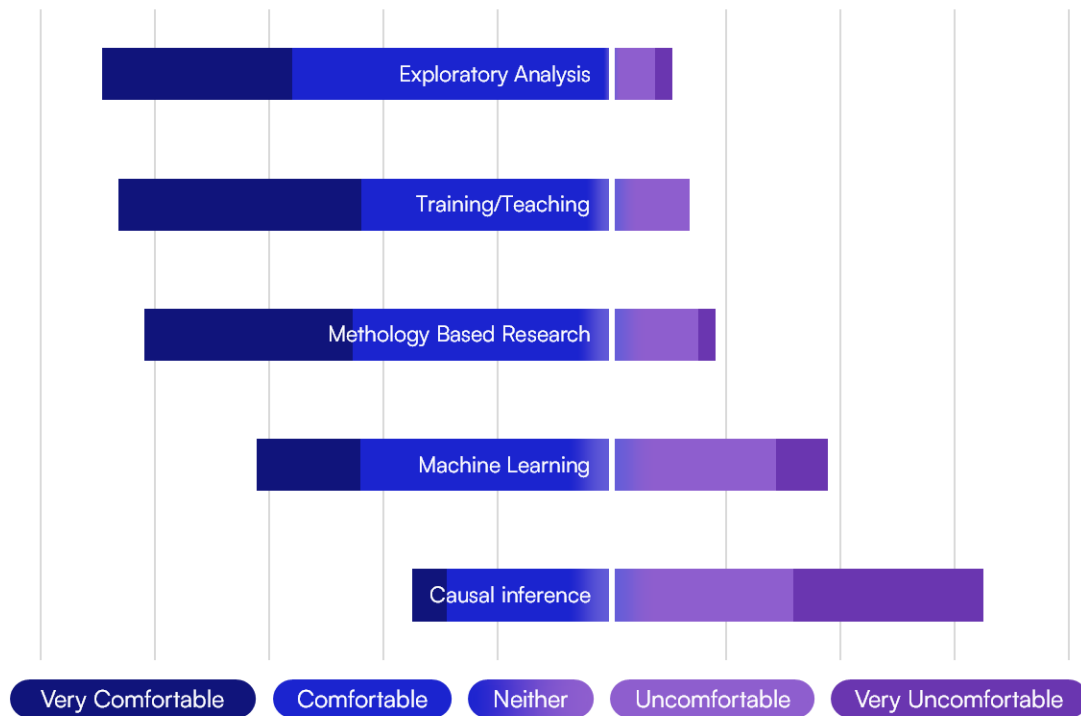
User Requirements for Synthetic Data

Report :

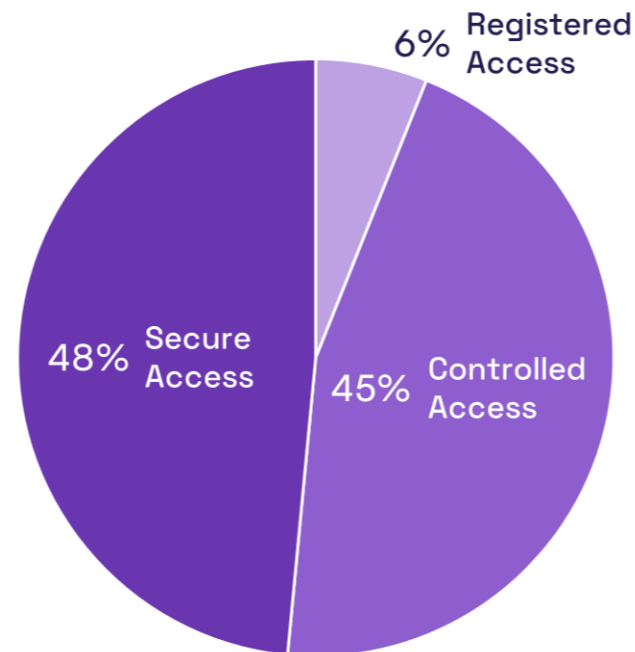


Data Owners Perspectives

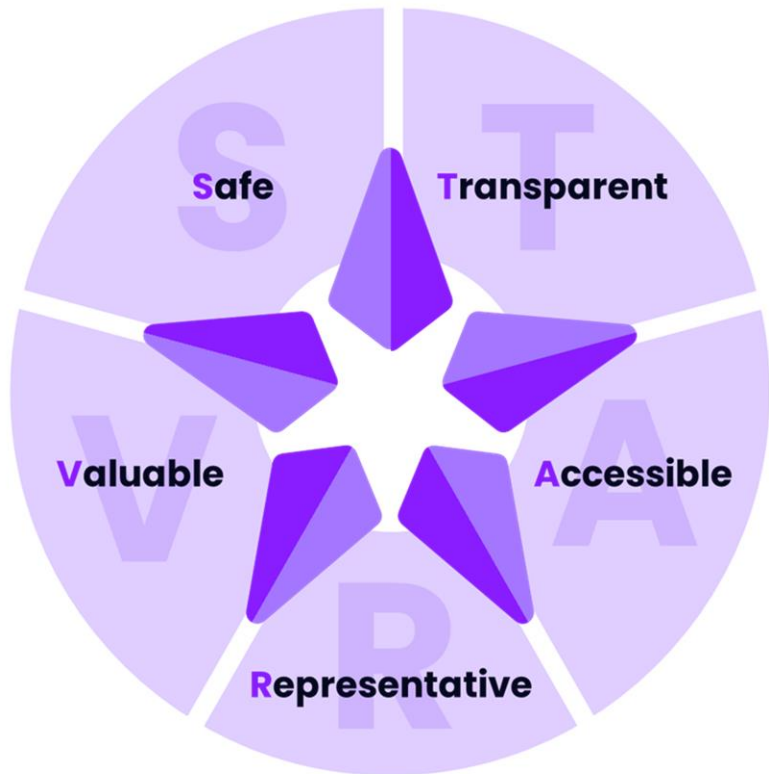
How comfortable would you be with high-fidelity synthetic data being used for the following:



How should high-fidelity synthetic data be made available?



The VSTAR Framework for Synthetic Data



V

Valuable : Synthetic data should be fit-for-purpose and demonstrate clear utility for a defined use case. From the start, the intended application should guide data generation, with end-stage evaluations confirming that the synthetic data meets expectations and supports the original objective.

S

Safe : Synthetic data should minimise the risk of re-identification through techniques like differential privacy and undergo rigorous testing for singling out, linkage, and inference risks. It must align with legal standards such as GDPR and demonstrate that individuals' data remains protected.

T

Transparent : The synthetic data generation process should be clearly documented and easy to understand, including the methods used, assumptions made, and any limitations. Evaluations should also be clearly communicated so that stakeholders can make informed decisions.

A

Accessible : Synthetic data should have lower governance barriers and simplified approval processes compared to real data. Safe sharing can be enabled through user agreements, making data more readily available while still maintaining responsible oversight.

R

Representative : High-quality synthetic data should reflect the structure, diversity, missingness, irregularities, and patterns of the original data. It must be realistic, plausible, and fair, avoiding the amplification of biases and supporting trustworthy, equitable outcomes in downstream tasks.

Next Steps...

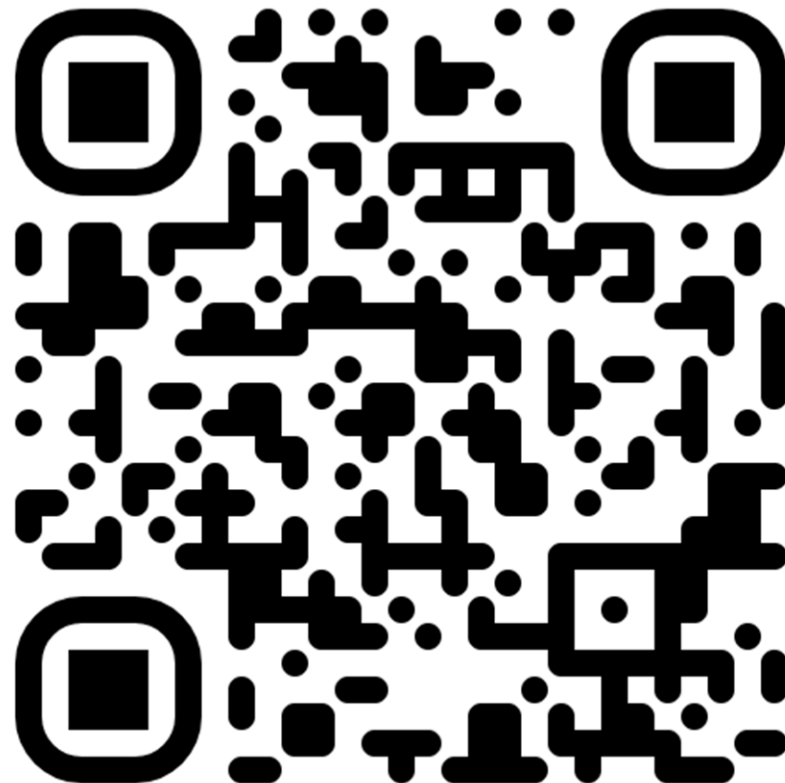
Website: www.syntheticdata.uk

Public Workshop to get feedback into our recommendations and guidance from members of the public – October.

Workshop with the Information Commissioners Office (ICO) to create principles and guidance around privacy evaluations - December.

Documentation Templates to help synthetic data developers record the generation and evaluation process to be transparent.

Open-Source Tools to enable Trusted Research Environments to develop varying levels of synthetic data fidelity.



DARE UK Community Group

The UK Synthetic Data Community Group

www.syntheticdata.uk



Funded By



Martin O'Reilly

The Alan Turing Institute



Federated Research Infrastructure by Data Governance Extension

Martin O'Reilly | Director of Research Engineering, The Alan Turing Institute

23 September 2025 | TRE Community Conference 2025



What is FRIDGE?

- A Trusted Research Environment capability for supercomputers ...

What is FRIDGE?

- A Trusted Research Environment capability for supercomputers ...
- ...supporting the extension of an existing TRE's information governance model to the supercomputer

Why is FRIDGE needed?

Why is FRIDGE needed?

- TREs generally aren't supercomputers

Why is FRIDGE needed?

- TREs generally aren't supercomputers
- Supercomputers generally aren't TREs

Why is FRIDGE needed?

- TREs generally aren't supercomputers
- Supercomputers generally aren't TREs
- Establishing trust between systems is hard

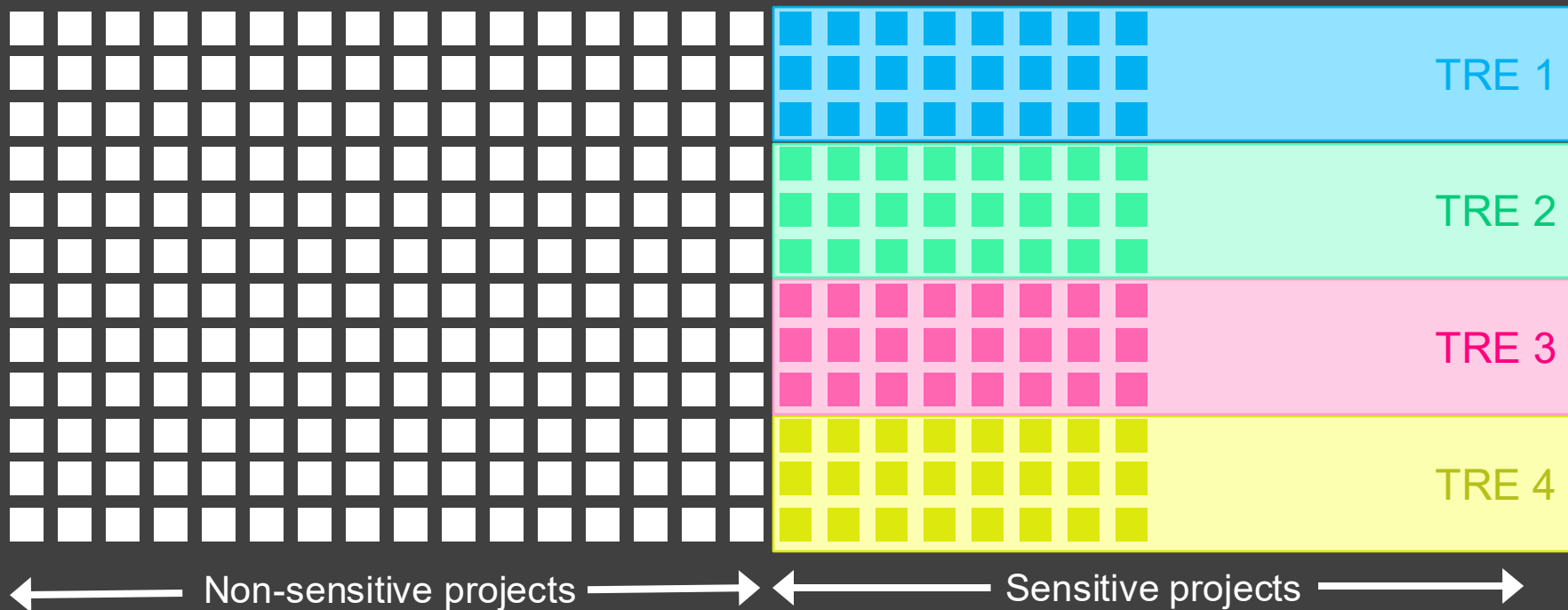
Why is FRIDGE needed?

- TREs generally aren't supercomputers
- Supercomputers generally aren't TREs
- Establishing trust between systems is hard
... even when both systems are TREs

How FRIDGE works

How FRIDGE works

FRIDGE will let **groups of computers** within a supercomputer be **strongly isolated** from each other in a way that complies with the governance rules for **different existing TREs**



How FRIDGE works

- Adapt **shared responsibility** model used for TREs deployed in public cloud

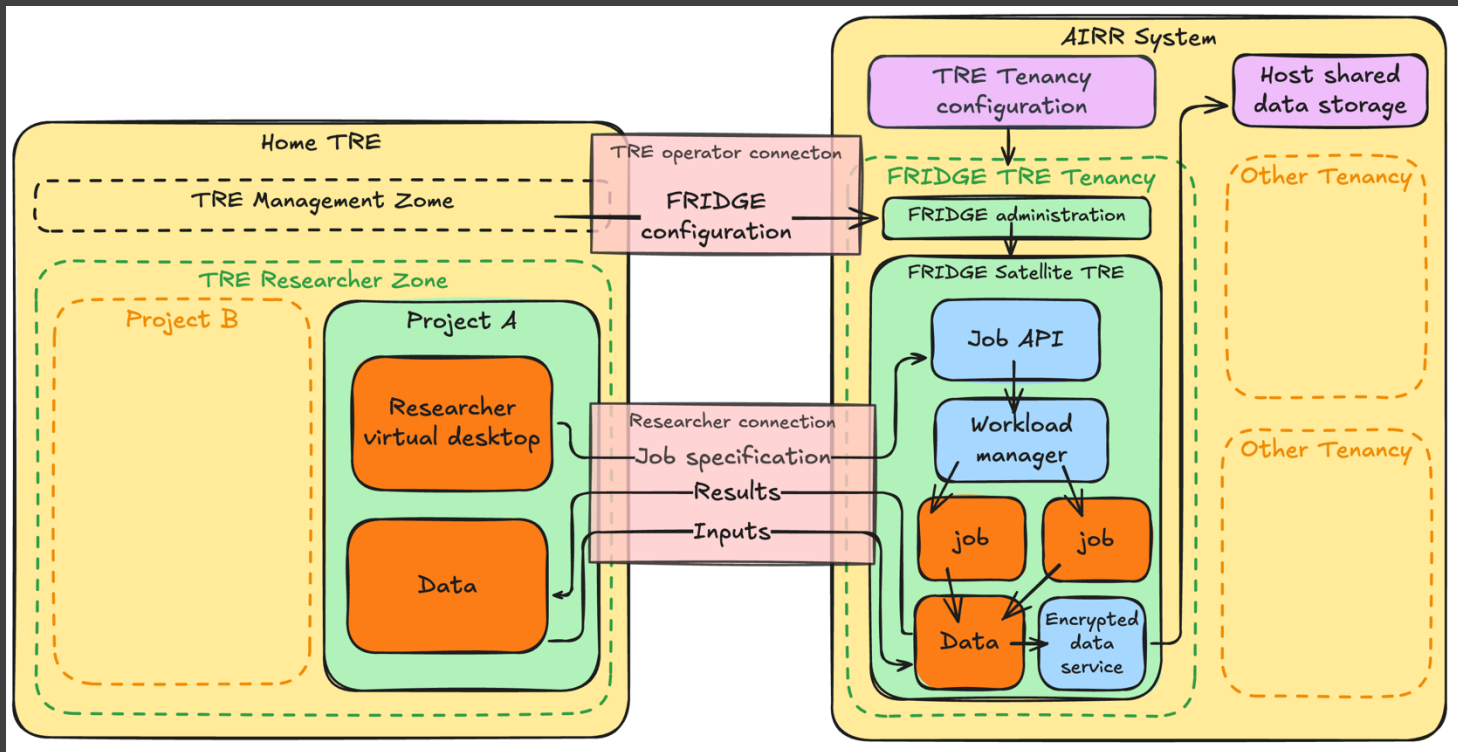
How FRIDGE works

- Adapt shared responsibility model used for TREs deployed in public cloud
- Extend **existing TRE's** information governance model to the supercomputer

How FRIDGE works

- Adapt shared responsibility model used for TREs deployed in public cloud
- Extend existing TRE's information governance model to the supercomputer
- Standardised **shared responsibility** TRE tenancy **hosting agreement**

How FRIDGE works



How FRIDGE works

Supercomputer host

Existing TRE operator

How FRIDGE works

Supercomputer host

- TRE tenancy isolation

Existing TRE operator

How FRIDGE works

Supercomputer host

- TRE tenancy isolation
- Secure connections to existing TRE

Existing TRE operator

How FRIDGE works

Supercomputer host

- TRE tenancy isolation
- Secure connections to existing TRE

Existing TRE operator

- Role-base access control

How FRIDGE works

Supercomputer host

- TRE tenancy isolation
- Secure connections to existing TRE

Existing TRE operator

- Role-base access control
- Pod security standards

How FRIDGE works

Supercomputer host

- TRE tenancy isolation
- Secure connections to existing TRE

Existing TRE operator

- Role-base access control
- Pod security standards
- Network policy

How FRIDGE works

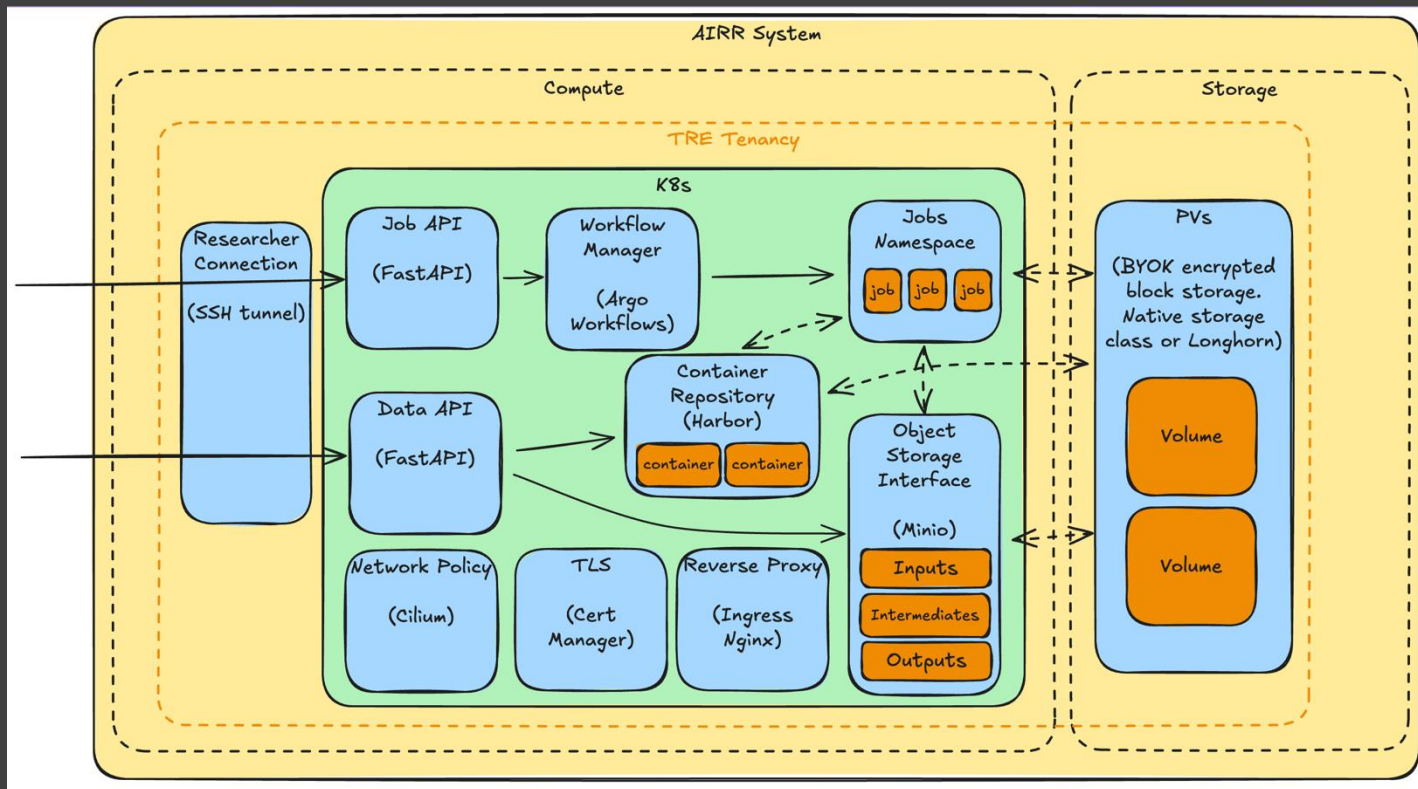
Supercomputer host

- TRE tenancy isolation
- Secure connections to existing TRE

Existing TRE operator

- Role-base access control
- Pod security standards
- Network policy
- Automatic data encryption

How FRIDGE works



What is success for FRIDGE?

- Deployment of real-world sensitive dataset to a third-party supercomputer for real-world research

What is success for FRIDGE?

- Deployment of real-world sensitive dataset to a third-party supercomputer for real-world research
- **Turing** project and researchers working on **Bristol** and **Cambridge** managed supercomputers (AIRR) under **Turing** information governance control




What is success for FRIDGE?

- Deployment of real-world sensitive dataset to a third-party supercomputer for real-world research
- Turing project and researchers working on Bristol and Cambridge managed supercomputers (AIRR) under Turing information governance control
- Providing **tools and processes** that are **easily adopted and adapted** by others to do the same

Project progress






Project progress

Dawn

- GPU passthrough 
- On-demand K8s 
- Development deployment 






Project progress

Dawn

- GPU passthrough 
- On-demand K8s 
- Development deployment 
- TRE tenancy 
- Volume encryption 

Project progress

Dawn






- GPU passthrough 
- On-demand K8s 
- Development deployment 
- TRE tenancy 
- Volume encryption 

Isambard-AI




- GPU passthrough 

Project progress

Dawn

- GPU passthrough 
- On-demand K8s 
- Development deployment 
- TRE tenancy 
- Volume encryption 

Isambard-AI

- GPU passthrough 
- TRE tenancy 
- Volume encryption 

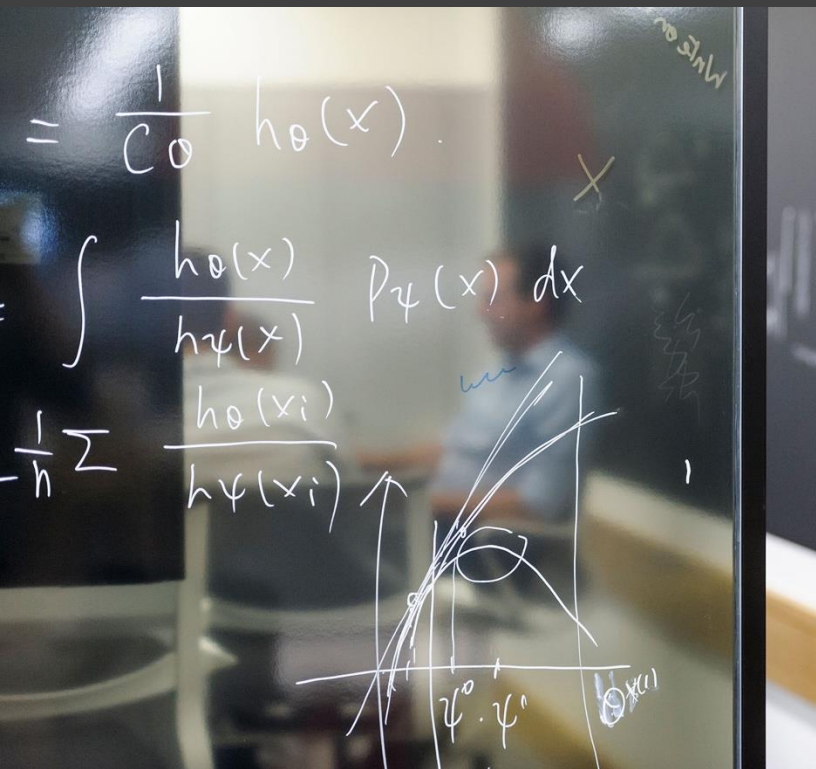
Getting involved

Getting involved

- **Data providers / TRE administrators:** What would make you comfortable extending your existing governance to a third-party hosted TRE tenancy?

Getting involved

- Data providers / TRE administrators: What would make you comfortable extending your existing governance to a third-party hosted TRE tenancy?
- **Supercomputer administrators:** Can you support the TRE tenancy model and how would you do so?



Getting involved

- Email:
moreilly@turing.ac.uk
jmadge@turing.ac.uk
- GitHub:
<https://github.com/alan-turing-institute/fridge>
- Project webpage:
<https://tinyurl.com/fridge-website>

Tim Machin



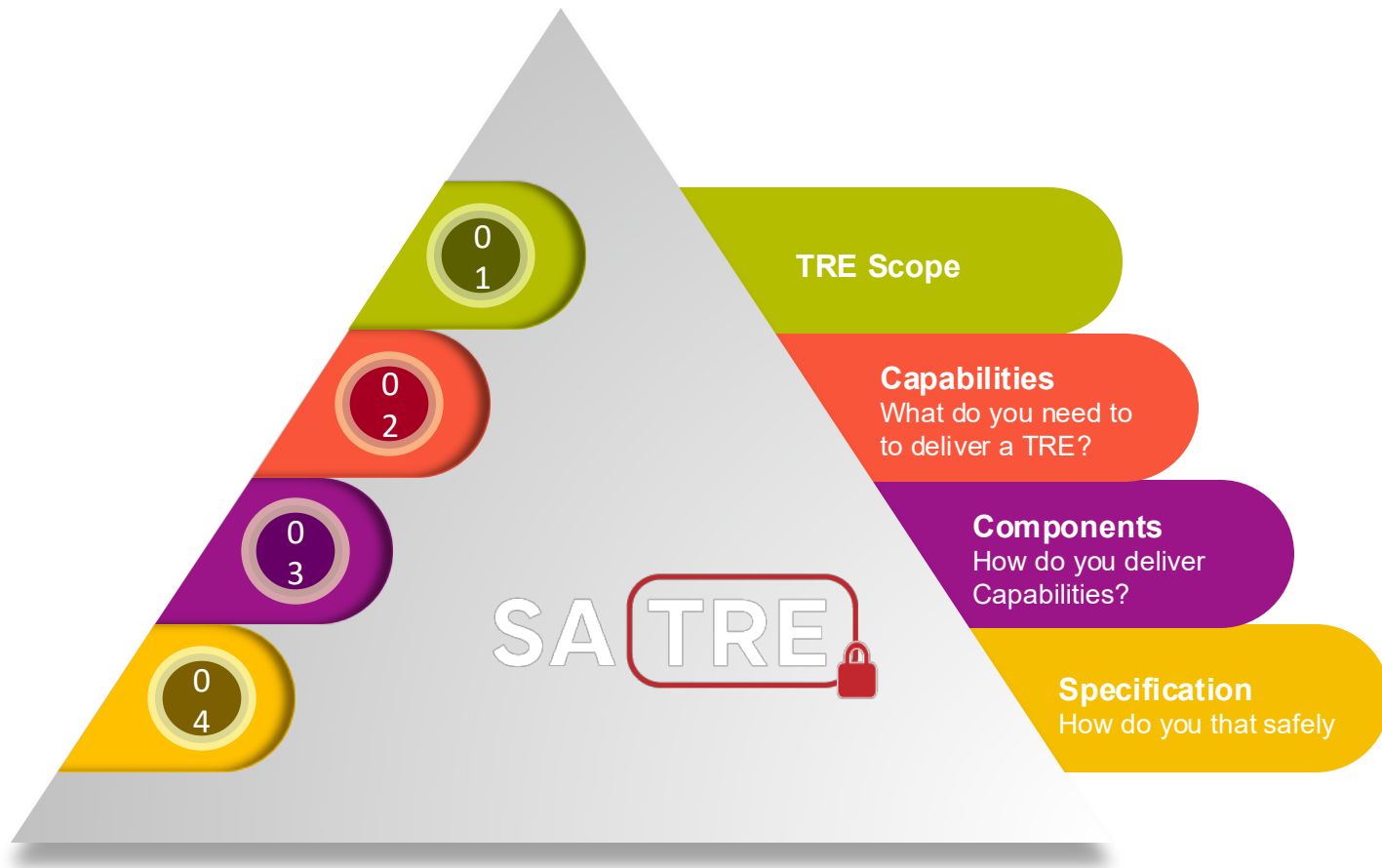
SATRE 2.0 Update

Tim Machin

UCL Centre for Advanced Research Computing



SATRE Structure



SATRE 2.0

- Re-org and Refresh
- Build in Federation
- High-level Design Pattern Library
- Mapping Standards

[SATRE 1.1](#) published as an explorable
architecture



Simplify structure

- Machine readable
- Flat structure
- Links to architecture and standard mapping

1.1. Governance Requirements

Requirements Gathering and Monitoring

This *business process* involves collecting, documenting, and managing the functional and non-functional requirements for the TRE based on the TRE organisation's goals and data assets.

	Statement	Guidance	Importance
1.1.1.	You must gather and monitor the information governance requirements needed to fulfil any legal, regulatory and ethical standards.	Requirements will come from a variety of sources including legislation, contractual obligations and ethical standards. Requirements must be monitored to ensure the TRE controls remain appropriate.	Mandatory

Controls

This *business process* involves measures, safeguards, or mechanisms implemented to manage or mitigate risks associated with your organisational requirements.

	Statement	Guidance	Importance
1.1.2.	You must ensure controls are implemented to ensure the requirements are met.	Control implementation should be systematic and directly aligned to the internal and stakeholder requirements.	Mandatory

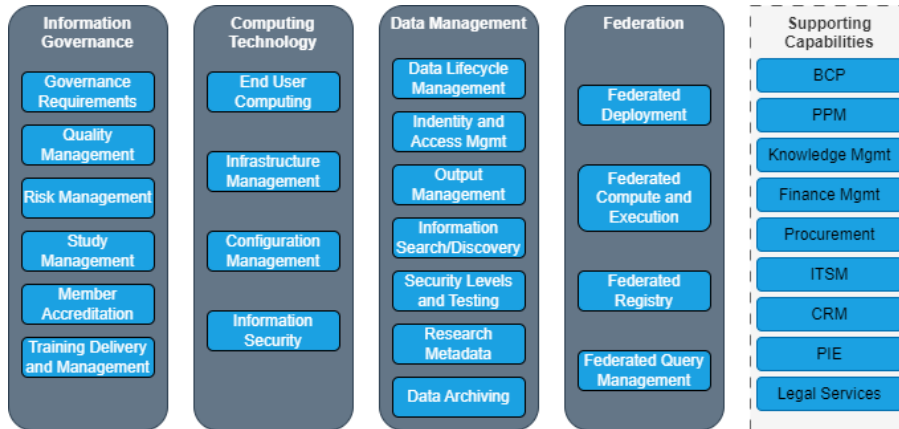
Resource Allocation Process

This *business process* involves assigning, distributing, and managing resources (such as personnel, finances, equipment, or time) within the TRE organisation to meet information governance requirements

No	Capability	Importance	Statement	Guidance
1.21	Quality Management	Mandatory	What the organisation does to measure and control quality of processes, documentation and outputs.	It is important to ensure that policies and SOPs are relevant, up-to-date and carefully controlled to maintain the integrity and security of your TRE organisation.
1.2.2	Quality Management	Mandatory	You must use versioning and a codified change procedure for all policies and standard operating procedures.	This includes recording dates of changes, person responsible for carrying out changes, and summary of changes.

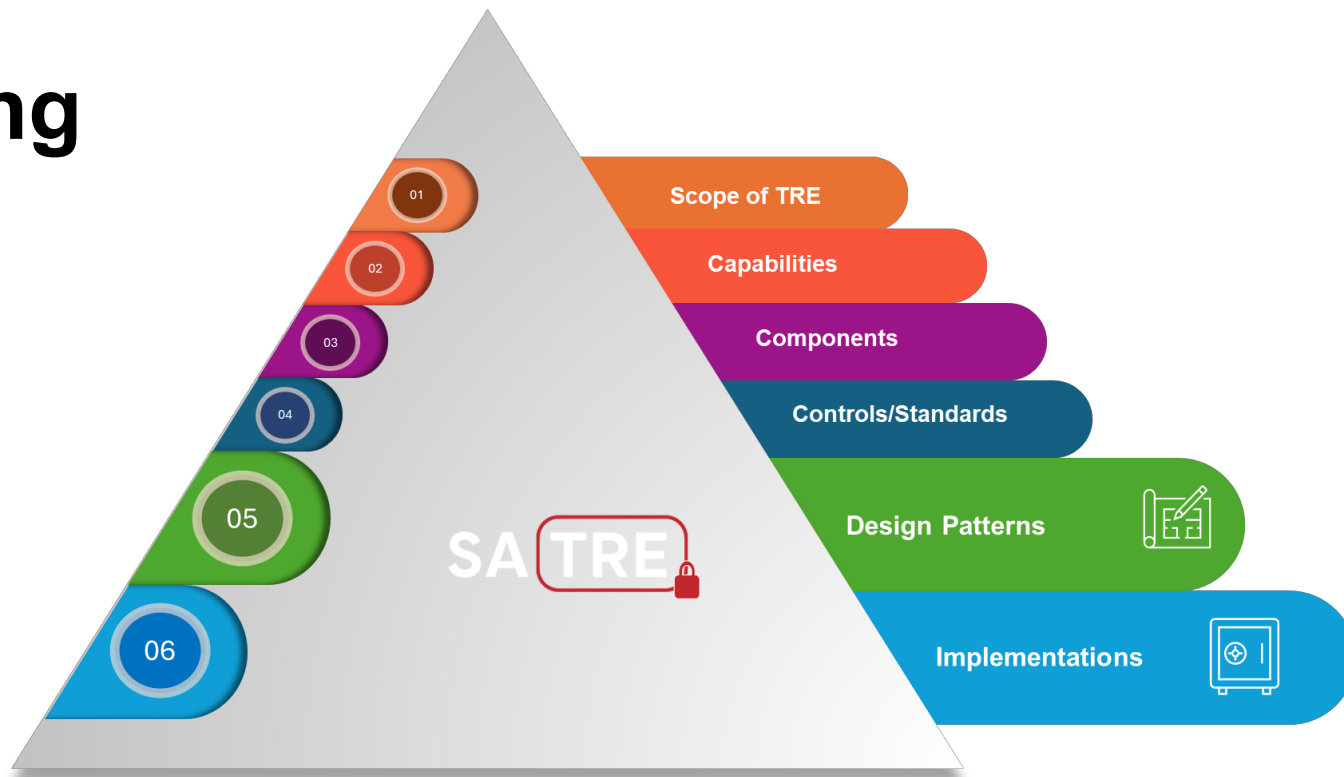
Building in Federation to SATRE

What must a Federation agree?



- A federation shall/should/could...
- A member TRE of a federation shall/should/could..
- A front door TRE shall/should/could...

Extending SATRE



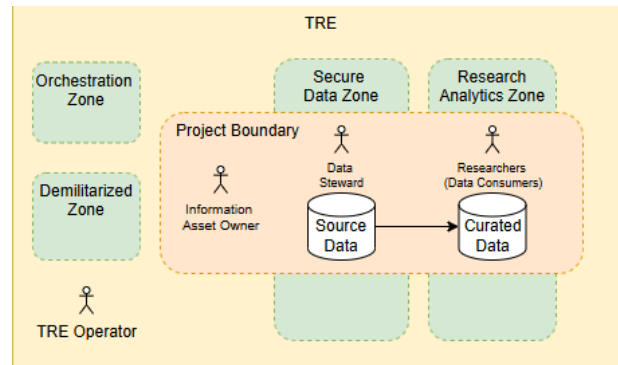
Example Pattern

“Bring your own data” archetype

- Common university pattern
- Scope of a “project”
- Research teams responsibility

Backlog

- Data federation patterns
- Planes and weaves (Federation network designs)
- TRE Zones
- Environment tiering



Role Class	Responsibilities	Primary Zone(s)	5 Safes
Project Roles	Conduct of research and the team	Research Analytics Zone	Safe People
Data Management	Source data management Provision of curated data Output checking	Secure Data Zone	Safe Data and Safe Outputs
Information Governance			
Governance	Risk ownership Quality Management	Overall TRE	Safe Projects
Infrastructure Management	Hardware and software Cyber security	Orchestration, Demilitarised Zone	Safe Setting