UK Health Data Research Alliance

# Trusted Research Environments: 12th March Workshop Report

## Background

The UK Health Data Research Alliance is an independent alliance of data providers, custodians and curators dedicated to improving human health by maximising the potential of health data at scale. One of the workstreams of the Alliance focuses on aligning approaches to Trusted (or Trustworthy) Research Environments (TREs), where the main objective is to define and agree the criteria required for analytical environments to allow safe and ethical use of health data for research and innovation.

One of the first outputs of this workstream is to develop a green paper on future options for UK Health Data TREs for organisations within and beyond the Alliance to consider. This will incorporate the outputs from a baseline exercise across existing environments, understanding of current and future user needs and considering the likely technological advances that will influence TRE provision.

## The workshop

On 12th March 2020 Health Data Research UK (HDR UK) hosted a workshop to discuss approaches to TREs and brought together a community of experts interested in improving the use of data for research in a safe and ethical way. The intention was to stimulate discussions on the importance of TREs and to establish a working group to collaborate on the development of the Alliance TRE workstream. The outputs from the workshop will provide an input to the next version of the draft green paper which will be shared widely for consultation.

Representatives from the data custodian community, TRE providers, HDR UK Public Advisory Board, and funders attended the workshop either in person or via Zoom videoconferencing. A full list of participants and the agenda can be found in the Appendix.

During interactive breakout sessions, participants rotated around three topic areas:

- Ideal requirements for TREs
- Approaches to federation across TREs
- TRE accreditation

Below we summarise the main points highlighted during the meeting.

UK Health Data Research Alliance

### *Requirements for TREs (safe setting, safe people, safe projects, safe data, safe outputs)*

- Agreement that a 'safe environment' approach should be adopted as part of the implementation of a governance based on the 'Five Safes'[1].
- Some data custodians already implement and use a subset of 'Five Safes', but a common agreed set of rules and specifications for TREs would be extremely helpful to build understanding across communities and to facilitate access to and use of health data in a safe analytical environment.
- TREs should permit software and algorithms to be imported and the underlying data would not be distributed, minimising the risks of re-identification, increasing control and safeguarding patients.
- The adoption of common standards, including accreditation and federation standards, could also facilitate the federated analysis of data across multiple TREs.
- Scope for TREs is very broad; having different TREs that manage a variety of data types and that different maturity levels should be considered. There are different types of TREs, and users need to choose what is right for them.
- Health Data Research Innovation Gateway could highlight TREs available and their functionality/features and driving users to the TRE that is most suited to them.
- Importance of involvement of patients and the public at each stage of the process:
    - Setting the rules and governance
    - Defining the standards
    - Data Access Committee membership
- Cultural change is one of the challenges to the wide adoption of TREs. Many users are still able to download data to their own settings following relevant approvals and shifting behaviour to bring software to an environment that the user does not control can be difficult.
- Need to work with the whole community, including the users, and ensure that they have the right skills and resources in place to be able to change their behaviour without undermining their research effort.
- Ensuring that only 'accredited' researchers can access the environment and controlling the level of data that they can use (data minimisation), can help mitigate the risks.

### *Accreditation of TREs*

- Need a widely accepted approach to accreditation that would meet the requirements of data custodians, regulatory bodies and patients and public representatives.
- Recognition of prior accreditation - there are existing accreditations that could be considered as examples to learn from, including the ISO27001, NHS Digital Data Security and Protection Toolkit, Digital Economy Act Accredited Processors (preparation and/or provision), Scotland Data Safe Haven.

[1] Desai, Tanvi; Ritchie, Felix; Welpton, Richard (2016). "Five Safes: designing data access for research". Bristol Business School Working Papers in Economics: Footnote 1.

- Community agreement on an accreditation system is paramount and that it would be helpful to build on what is already there and ensure alignment among stakeholders.
- In line with the requirement of TREs, accreditation should still be based on the Five Safes (i.e. not just 'setting') but how the other safes are also implemented:
    - Level of constraints/controls to ensure safe **outputs**
    - Ability to support conditions applied following **project** level assessment
    - Managing ingress and linkage of **data** that are in different places (which would determine the level of security needed in **setting**
    - Both governance and technical accreditation of setting sufficient to meet data controllers' needs
    - Authentication, authorisation and audit of safe **users.**
- Legislation changes and learning from incidents (both direct and those occurring in different sectors) may require changes to accreditation standards, so flexibility and understanding who is responsible for accreditation updates would be important.
- Perception of Safe Users influences public trust. Academic research is generally more easily accepted than health data research by multi-national pharmaceutical companies. Thus, it is particularly important to consider PPIE when thinking about TRE accreditation and capture and public and patient views.

### *TREs Federation*

- For this discussion federation was defined as the ability to undertake projects using data from different TREs.
- Data can be analysed separately in different TREs and then summary outputs can be compared. However, there are other approaches to TRE and data federation that might allow linking of data within a single TRE or a virtual TRE environment, allowing better analysis of data.
- There are various possible levels of federation between adequately accredited TREs:
    - Level 1: Beacons: an API service that can be programmed externally to generate and export a summary dataset
    - Level 2: Deployment of same algorithm on multiple TREs, each generating and exporting a summary dataset
    - Level 3: Execution of algorithm on multiple TREs that communicate with each other (e.g. to update a shared mixing matrix in AI)
    - Level 4: Transfer of individual data between TREs.
- TRE accreditation will be necessary to establish a common model of trust across a federation of TREs with common user identity.
- However, accreditation would not be sufficient to allow data federation between TREs, but that rules of governance of data should drive it.
- The TRE might only be a data processor and the data controller would need to be satisfied.

- The concept of accredited researchers or safe people might also be different for different TREs, where individuals may have expertise with different data types and data users may be regarded as safe for one TRE and not for another.
- Potential risk of data leakage if a single TRE hosts data from different projects with different access rules.
- Even if governance criteria are satisfied, it will be important to hold data in common formats and ensure that the same software packaging system is supported in each TRE.
- Suggested that the Alliance might adopt multiple levels of accreditation rather than just one, with the highest-level allowing data exchange for instance.
- There is a need for a TRE community oversight on accreditation and it was felt that HDR UK could be in a good position to present a sort of service catalogue across the Alliance to showcase what functionality and types of data each TRE may offer [see above link to role of Gateway].

## Next steps

- Further comments and input on the current version of the green paper on TREs that can be found here[2].
- Gather and share existing TRE / safe haven specifications and accreditation standards
- Incorporate the feedback received and the outputs from the workshop into a new version of the draft green paper that will be shared widely for consultation in the next couple of weeks.
- Establish working groups to support the design of TRE options.
- Continue to align with UKRI Digital Research Infrastructure developments and implementation of Digital Economy Act accreditation processes.
- Produce the next iteration of the approach by 16 June 2020 for the HDR UK One Institute virtual event.

[2] https://drive.google.com/file/d/15ZfT_6WL4la5zuWPdDPDKEpqZTe7zSvv/view?usp=sharing

UK Health Data Research Alliance

# Appendix

## Agenda

| Approx. Timing | Agenda item | Lead / approach |
|---|---|---|
| 9:30 | Registration & coffee | |
| 10:00 | Workshop overview and round of introductions<br><br>Case for using Trusted Research Environments | Tim Hubbard & Gerry Reilly |
| 10:15 | Breakout discussions, all groups rotating around same three topics [15 mins per group per topic]:<br><br>• Discussion 1 – Requirements for a TRE (safe people, safe projects, safe settings, safe outputs, safe data)<br>• Discussion 2 – Federation<br>• Discussion 3 – Accreditation of TREs | 3 groups (Tim Hubbard, Paola Quattroni and David Seymour facilitate / scribe) |
| 11:00 | Feedback from groups and discussion | Breakout Leads |
| 11:20 | PPIE questions and consultation | Amanda White |
| 11:45 | Conclusion and next steps | Tim Hubbard & Gerry Reilly |
| 12:00 | Meeting Close & Lunch in ground floor atrium | |

UK Health Data Research Alliance

## List of attendees (either in person or via Zoom)

| Name | Organisation |
|---|---|
| Antony Shimmin | AIMES |
| Dennis Kehoe | AIMES |
| Richard Spragg | AIMES |
| Alicia Gibson | Aridhia |
| Martin O'Reilly | Alan Turing Institute |
| Sebastian Vollmer | Alan Turing Institute |
| Tim Hubbard | Genomics England |
| Natasha Spiridou | GOSH |
| Michael Chapman | HDR UK Cambridge |
| Kush Kanodia | HDR UK Public Advisory Board |
| Garry Coleman | NHS Digital |
| Laura Sato | NHS Digital |
| Jonathan Pilgrim | NIHR BioResource |
| Neil Walker | NIHR BioResource |
| Richard Hier | Swansea University |
| Simon Thompson | Swansea University |
| James Hetherington | UKRI |
| Monika Maini | University of Leicester |
| Umar Riaz | University of Leicester |
| | |
| Amanda White | HDR UK |
| Andrew Morris | HDR UK |
| David Seymour | HDR UK |
| Gerry Reilly | HDR UK |
| Neil Sebire | HDR UK |
| Paola Quattroni | HDR UK |
| Sinduja Manohar | HDR UK |
| Susheel Varma | HDR UK |