International Journal of Population Data Science

Journal Website: www.ijpds.org

Common governance model: a way to avoid data segregation between existing trusted research environment

Fatemeh Torabi[1,2,*], Chris Orton[1,3,†], Emma Squires[1,2,†], Sharon Heys[1,4], Richard Hier[1], Ronan A. Lyons[1,2,4], and Simon Thompson[1,2,3,4]

Submission History	
Submitted:	29/04/2023
Accepted:	18/09/2023
Published:	08/11/2023

[1]Population Data Science, Swansea University, Swansea, SA2 8PP
[2]Dementias Platform UK, Swansea University, Swansea, SA2 8PP
[3]Health Data Research UK, 215 Euston Road, London, England, NW1 2BE
[4]SAIL Databank, Swansea University, Swansea, SA2 8PP
[†]These authors contributed equally to the work

Abstract

Background
Trusted Research Environments provide a legitimate basis for data access along with a set of technologies to support implementation of the "five-safes" framework for privacy protection. Lack of standard approaches in achieving compliance with the "five-safes" framework results in a diversity of approaches across different TREs. Data access and analysis across multiple TREs has a range of benefits including improved precision of analysis due to larger sample sizes and broader availability of out-of-sample records, particularly in the study of rare conditions. Knowledge of governance approaches used across UK-TREs is limited.

Objective
To document key governance features in major UK-TRE contributing to UK wide analysis and to identify elements that would directly facilitate multi TRE collaborations and federated analysis in future.

Method
We summarised three main characteristics across 15 major UK-based TREs: 1) data access environment; 2) data access requests and disclosure control procedures; and 3) governance models. We undertook case studies of collaborative analyses conducted in more than one TRE. We identified an array of TREs operating on an equivalent level of governance. We further identify commonly governed TREs with architectural considerations for achieving an equivalent level of information security management system standards to facilitate multi TRE functionality and federated analytics.

Results
All 15 UK-TREs allow pooling and analysis of aggregated research outputs only when they have passed human-operated disclosure control checks. Data access requests procedures are unique to each TRE. We also observed a variability in disclosure control procedures across various TREs with no or minimal researcher guidance on best practices for file out request procedures. In 2023, six TREs (40.0%) held ISO 20071 accreditation, while 9 TREs (56.2%) participated in four-nation analyses.

Conclusion
Secure analysis of individual-level data from multiple TREs is possible through existing technical solutions but requires development of a well-established governance framework meeting all stakeholder requirements and addressing public and patient concerns. Formation of a standard model could act as the catalyst for evolution of current TREs governance models to a multi TRE ecosystem within the UK and beyond.

Keywords
data governance; Trusted Research Environments; data protection

*Corresponding Author:
Email Address: fatemeh.torabi@swansea.ac.uk (Fatemeh Torabi)

https://doi.org/10.23889/ijpds.v8i4.2164
November 8, 2023 © The Authors. Open Access under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/deed.en)

Background

The General Data Protection Regulations (GDPR) along with the Data Protection Act 2018 apply to personal data [1–3] within the UK. Research organisations working with personal data, including data related to health and social care, must ensure that data protection legislation is adhered to when accessing and processing data. The principles which underpin accessing, controlling and processing personal data are set out within Article 5 and Chapter 4 of GDPR and provide the legal basis for accessing data [4]. These legal basis are met through accessing sensitive data within safe environments often referred to as Trusted Research Environments (TREs) [3, 5]. TREs provide the legal basis for safe and secure access to patient data while also ensuring that the "five-safes" principles are met [6]. Each project requesting access to data that is held within a TRE is evaluated on: scientific rationale, public benefit, researcher accreditation/training and analysis plans. Achieving adequate levels of adherence to the governance within the TREs is mostly done via an array of technical solutions, as well as procedures on data access requirements for safe researcher and safe projects. However, not all components of the procedure are standardized. TREs have invested time and effort into information security management systems (ISMS) [7]. While some of the national data providers, like the NHS England Secure Data Environment (SDE) and Office of National Statistics (ONS), mandate that platforms are managed in line with these accreditations, currently there is no standardised approach on how the architectured computing systems of TREs which are comprised of software and hardware should achieve the integrated governance procedure. All TREs holding patient level health data aim to provide a secure and trustworthy environment with technical and analytical resources enabling researchers to generate insights from individual-level health data. During the COVID-19 pandemic, secure NHS data flows within TREs played a crucial role in gaining timely answers to critical research questions to enable government and the health sector on steering the population securely out of the pandemic [8, 9].

Major investment into technical solutions for TREs has resulted in establishment of multiple TREs across each nation of the UK [10]. All these TREs have excelled in gaining public trust [6, 11]; however, the existence of differing routes in achieving governance can suggest that different standards are in operation, which is not the case, as an absolute baseline is met for each TRE. The difference lies in the level of information security achieved in each environment allowing a TRE to operate under an accredited ISMS as a safe and secure environment. Additionally, existing UK TREs have obtained specific data assets and data flows. These flows sometimes overlap resulting in existence of the same copy of the data in more than one TRE [12]. The disparate approaches and intersecting data streams manifest as a concern when we enter the world of joint governance models beyond single environment systems. For instance, each output generated by researchers working within a TRE environment must undergo a thorough review process prior to its secure release form the TRE. The disclosure control process entails meticulous scrutiny carried out by trained experts to ensure that there is no risk with individual identification. These rules are not standardised across existing TREs. For the same type of data and methodological approach, a set of statistical outputs may be considered safe in one TRE but may require further details before release in another TRE [12]. Mapping the existing governance models and collaborating on standardization efforts will contribute to the path towards analysis of patient-level data across multiple-TREs [13].

A standard governance model is the core pillar of what is called *data federation*. Data federation has evolved since its initial introduction as a technical solution for enabling access and analysis of data across more than one database [14]. Projects such as the European Medical Information Framework (EMIF) have explored and established federated approaches in ethical access and use of health data held in multiple secure data environments (SDEs) across European countries [15, 16]. Many of these lessons learned by EMIF on improved identification, access and reuse of health data are directly applicable to the data held within TREs, and require advance technological solutions for secure connection of TREs as well as a single common framework for secure governance of the linked network of TREs. Implementation of a federated network of TREs across the four devolved UK nations (England, Scotland, Wales and Northern Ireland) would have had considerable benefits for researchers and governments in answering policy-relevant questions. In this article, we focus on characterising 15 UK-TREs based on their existing governance models. We further focused on operational and governance requirements for cross-TRE collaboration and identified a subset of TREs which were operating on equivalent level of governance.

Methods

We have identified 15 major UK TREs based on their direct contribution to UK wide analysis, enabling secure access to patients' personal data, facilitating a swift national response to COVID-19. We have summarized three key characteristics found across these TREs: 1) the data access environment, 2) data access requests and disclosure control procedures, and 3) governance models. For each TRE, we provide detailed insights into data access requirements and governance approaches, with the goal of identifying potential areas for standardizing existing data flows and achieving a common governance model. Additionally, we present case examples of multi-institutional collaborations that are developing protocols and implementing studies across multiple TREs. In doing so, we outline the challenges faced and propose *pop-up TRE* as a solution for achieving multi TRE analysis. We categorise TREs at an equivalent level of governance if they are operating on a same level of ISMS with regards to international standard on requirements for information security management [17]. Furthermore, we have identified TREs that share common governance practices and have considered architectural strategies to meet an equivalent standard of ISMS, the equivalent standard is defined as having similar measures for data access requirements and output review procedures. Such measures are a prerequisite to facilitating multi TRE collaboration and ultimately federated analytics.

Building upon the principles of common governance practices and equivalent standards in data access and review procedures, this article elaborates on our innovative 'pop-up

TRE' approach. This approach facilitates the participation of multiple institutions in collaborative research projects that leverage data from diverse TREs, effectively addressing the operationalization of research governance requirements within multi TRE collaborations.

Results

All 15 UK-TREs use their specific web portals to provide an overview of data access environment, rules and procedures on data access requests (DARs) which are unique to each TRE (Table 1).

Data access environment

All of the TREs provide a suite of analytical tools within their data access environment. We have documented two primary approaches in the technical implementation of the data access environments: 1) online platform, and 2) virtual machines which operate similar to a normal windows or Linux machine. Data is made accessible either through direct access to data tables or databases. England, Wales and Northern Irelan TREs offer their data in a database format, empowering the analysts through structured query language to interrogate the database. Public Health Scotland provides the national data files within the Safe Haven as version copies.

Data access

Approval procedures are required for research across all 15 UK TREs. We have documented that among the 11 TREs (68.7%) providing secondary access to routinely collected health data, data access requirements (DARs) were achieved through direct approval from TRE governance panels. This was the case for national TREs such as NHS England SDE, RGCP, Turing, PIONEER, and OpenSafety in England [18–22] ; SAIL Databank in Wales [23]; Public Health Scotland's eDRIS and Scottish Safe Haven platform [24] and Honest Broker Service (HBS) in Northern Ireland [25]. Similarly, the electronic health records held within the UK Health Security Agency [26], as well as the institutional platforms like the UCL Data Safe Haven [27] and Imperial college Big Data & Analytical Unit Secure Environment [28], also followed this approach. The Dementias Platform UK [29] and Avon Longitudinal Study of Parents and Children (ALSPAC) [30] were two platforms that offered a data access approval procedure involving direct participation of data owners in the review process. There is no direct governance panel for the national UK Longitudinal Linkage Collaboration (UK LLC), the access approval is through a triage service with similar data owner approval as for DPUK and ALSPAC [31] Information regarding DARs for the Wales National Data Resource was limited to the environment's blueprint, thus not included in our evaluation of DAR procedures.

Disclosure control procedures

Major national TREs across devolved nations (England, Wales, Scotland, and Northern Ireland) have published their disclosure control procedures required for accessing and analysing their data on the TRE platform's websites (Table 1).

Governance model

None of the TREs outline a requirement to apply through the integrated research application system (IRAS) established by the National Health System (NHS) England [32]. The legal basis of all governance approaches is set according to GDPR, DPA and S251 regulations [1–3]. However, across nations, each TRE utilizes a different set of criteria for their governance revisions. For instance, the NHS England SDE, and OpenSAFELY, were originally established for accessing COVID-19 datasets for pandemic response, have now undergone reassessment to support ongoing data flows addressing similar questions in the context of long COVID [33, 34]. The COVID-19 response models within some TREs have been structured using a project access model with a predefined end date. Consequently, new investigations for long COVID and or other disease areas are reviewed separately by governance panels [29, 35, 36]. We also found examples of disease focused TREs [22, 37] as well as institutional-level TREs [21, 27, 38]. There were also disease focused hubs established within existing TREs, for example SAIL Databank, NHS Digital and OpenSafely; all followed a project approach to establish secure access within hubs supported by a consortium project agreement [19, 39]. Consolidating all existing governance models ands data access requirements within TREs, along with harmonising the established research projects, would significantly enhance data access, review, and analysis processes for future projects (refer to Table 1 for a summary of TRE characteristics).

Case examples of multi TRE research

Six national TREs enabled and supported UK wide multi-nation analysis [19, 20, 23–25, 33] across various areas of COVID-19 investigations. A four-nation approach facilitated investigation of rare adverse reactions post COVID-19 vaccination [41] where individual nations had reported their limited ability to reach a reliable conclusion due to small sample size [42, 43]. Similarly, this multi TRE collaboration supported investigation on vaccine booster analysis within a pooled population of 30 million individuals [44]. Associations between COVID-19 and cardiovascular medication usage, along with cardiovascular outcomes, were assessed using data from multiple TREs [45, 46]. All of these collaborations were based on combining aggregated level data from TREs [45, 46]. In a unique instance, row-level data from multiple TREs were pooled through special permissions for the analysis of rare outcomes post COVID-19 vaccination [41] (Table 2).

TRE access costs models

One of the important aspects that affects the resilience of a TRE in implementing the requirements for joining a broader network of TREs is the costing models applied to research projects. We identified two main funding models across existing TREs: 1) the cost recovery model and 2) central funding model. In the first model, cost recovery occurs by charging each project against services provided to them while the second model receives centralised support and funding for the services provided, in most cases without providing an onward cost to the end user.

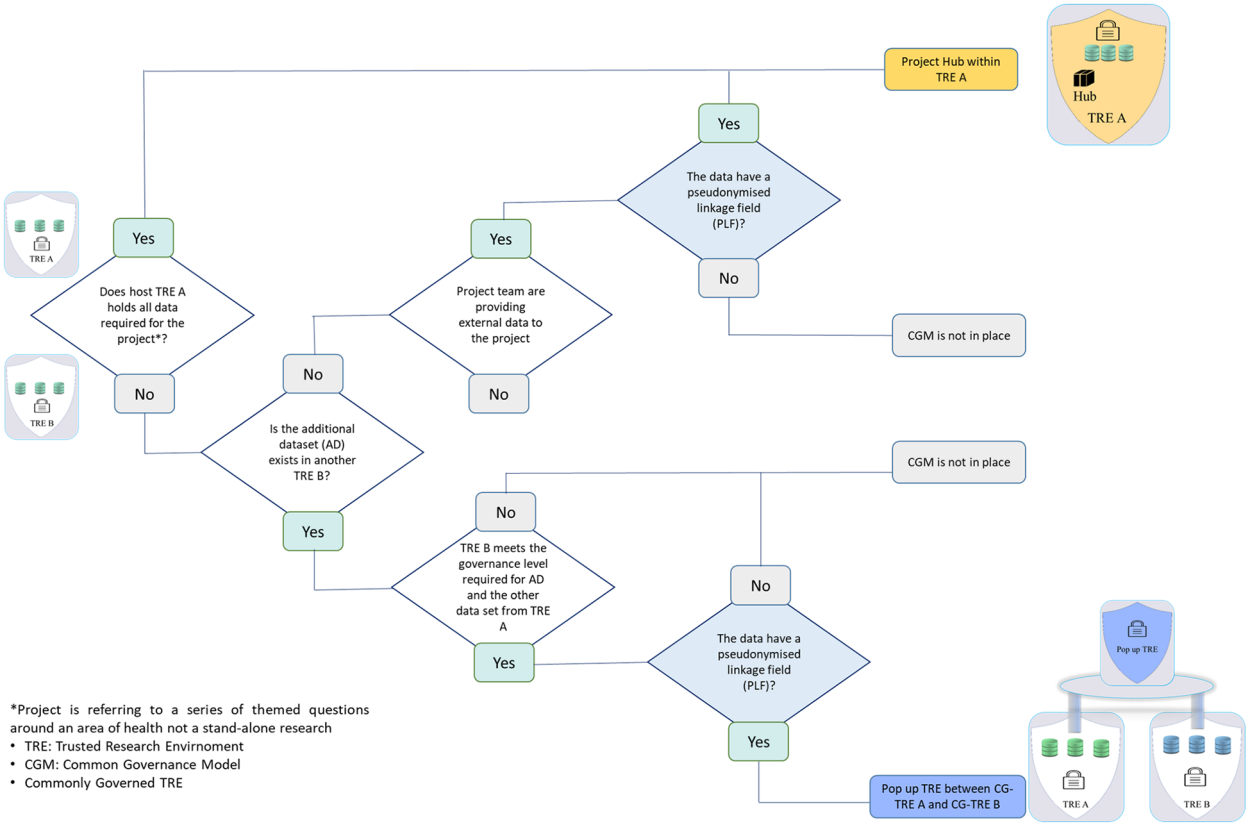
Table 1: Summarising existing TRE and infrastructures for the populations of England, Scotland, Wales and Northern Ireland

No.	Nation	TRE name	Data access environment (DAE)	Data access requests (DARs) and output disclosure control procedure	Governance model
1	**England**	NHS England Secure Data Environment (SDE) [40]	Secure web-based environment accessible via two step authentication for approved users - passed required trainings (more info)	**DARs**: Project application submission – costing done by case officer **Output check**: human based	ISO 27001 certified research environment Data Sharing Agreement
2	**England**	Imperial college - Big Data & Analytical Unit Secure Environment (BDAU SE) [28]	Secure web-based platform with access to code builder and study design platform for approved users (more info)	**DARs**: Institutional review process **Output check**: human based	ISO 27001 certified research environment Compliant with NHS Data Security and Protection Toolkit
3	**England**	UK Health Security Agency protected data [37]	Secure web-based platform for approved users (more info)	**DARs**: Project application submission and review based on UKHSA data application form and standards **Output check**: human based	Compliant with NHS Data Security UKHSA approval standards
4	**England**	Oxford-Royal College of General Practitioners (RCGP) Research and Surveillance Centre (RSC) database [20]	Secure web-based platform for approved users – required to provide the date they've passed Information Governance training (more info)	**DARs**: Online application form and project review **Output check**: human based	Information Governance review
5	**England**	PIONEER [22]	Secure web-based platform for approved users -passed required trainings	**DARs**: Application form and project review **Output check**: human based	National Data Guardian for health and care
6	**England**	UCL Data Safe Haven (DSH) [27]	Web-based platform for approved users (more info)	**DARs**: Account access through UCL secure sign in (More info) **Output check**: human based	Not specified
7	**England**	OpenSafely [19]	Web-based platform for approved projects and users (more info)	**DARs**: Contact with the team **Output check**: human based	OpenSafely oversight board acting on behalf of the NHS England
8	**England**	Turing Data Safe Haven [21]	Web-based platform for approved projects and users – institutionalized internal access (more info)	**DARs**: Internal institutional access **Output check**: human based	Not specified
9	**England**	UK Longitudinal Linkage Collaboration (UKLLC) [31]	Secure access via virtual machine with two-factor authentication for approved users (https://ukllc.ac.uk/apply/)	**DARs**: Any UK-based research can apply for this data via: https://ukllc.ac.uk/apply/ **Output check**: human based	ISO 27001 certified research environment https://ukllc.ac.uk/apply
10	**England**	Avon Longitudinal Study of Parents and Children (ALSPAC) [30]	Secure access via virtual machine with two-factor authentication for users.	**DARs**: Application form and project review outlined: http://www.bristol.ac.uk/alspac/researchers/access/ **Output check**: human based	http://www.bristol.ac.uk/alspac/researchers/access/
11	**Wales**	Secure Anonymised Information Linkage Databank [23]	Secure access via virtual machine with two-factor authentication for approved users – passed required trainings (more info)	**DARs**: Application form and project review **Output check**: human based	ISO 27001 certified research environment Information Governance Panel approval procedure
12	**Wales**	Dementias Platform UK [29]	Secure access via virtual machine with two-factor authentication for approved users (more info)	**DARs**: Online application form and project review **Output check**: human based	Cohort data owners will review each application individually
13	**Wales**	Wales National Data Resource [47]	National Data Resource (more info)	**DARs**: Under development - Each partner TRE reviews applications under their own information governance **Output check**: human based	Information Governance Panel
14	**Scotland**	Public Health Scotland's Scottish Safe Haven [24]	Secure access via public health Scotland for approved users (more info)	**DARs**: Application form and project review **Output check**: human based	Reviewed by NHS Scotland Public Benefit and Privacy Panel for Health and Social Care (HSC-PBPP)
15	**Northern Ireland**	Honest Broker Service (HBS) Governance Board [25]	Secure access for approved projects and users incorporated data access via the SeRP tenancy as well as direct access via facilities directly housed by the HBS (more info)	**DARs**: NHS research ethics committee. The HBS process does not require separate National Research Ethics Service governance approval. **Output check**: human based	ISO 27001 certified research environment Honest Broker Governance Board

Table 2: Case examples for multi-nation collaborative efforts

Title	Author	Nations
Association of COVID-19 With Major Arterial and Venous Thrombotic Diseases: A Population-Wide Cohort Study of 48 Million Adults in England and Wales [46]	Knight et al. 2022	England & Wales
Severe COVID-19 outcomes after full vaccination of primary schedule and initial boosters: pooled analysis of national prospective cohort studies of 30 million individuals in England, Northern Ireland, Scotland, and Wales [44]	Agrawal et al. 2022	England, Northern Ireland, Scotland, and Wales
First dose ChAdOx1 and BNT162b2 COVID-19 vaccinations and cerebral venous sinus thrombosis: A pooled self-controlled case series study of 11.6 million individuals in England, Scotland, and Wales [41]	Kerr et al. 2022	England, Scotland, and Wales
The impact of the COVID-19 pandemic on cardiovascular disease prevention and management [45]	Dale et al. 2023	England, Scotland, and Wales

Figure 1: Governance classification for joint collaboration across multiple TREs



A solution for multiple-TRE collaboration

TREs provide an overarching governance framework and a set of technologies to support the operation of research within the defined parameters. The data sensitivity, governance requirements and capabilities of a popup TRE will determine if it can be established as an entire stand-alone environment or if it needs to exist inside an environment providing the baseline secure environment (inside an existing TRE – Figure 1). A pop-up TRE should implement appropriate responses to both these aspects while aligning or providing equivalence for the governance framework of the contributing TRE's / data providers. There is a class of federated analysis that enables only the aggregate findings to be released from a TRE to be combined. However, this is a narrow subset of use cases. A pop-up TRE should be able to facilitate combining datasets in a common neutral place where all of the governance arrangements for the datasets can be fulfilled. It is not the intention of a pop-up TRE to carry the full set of the datasets needed to support a research question; however it is inevitable

that data will "move" into the pop-up TRE and may remain for the duration of the research project being undertaken, within the governance boundaries agreed by the data-providing TREs involved in any one project. Therefore, considerations about location of data, who can access the data and at what level, careful review of research protocols, analysis plans and disclosure control checks to ensure conformation to agreed governance standards are key for the successful operation of pop-up TREs to act as a safe platform for multi TRE analysis. These multi TRE analyses require careful consideration of governance models of each collaborating TRE. A pop-up TRE operates only within the lifespan of the project, using a very similar governance model to the project hubs that are already in operation within some of national TREs. (Figure 1 outlines the classification in governance models for establishing a hub or a pop-up TRE).

Discussion

In this article, we conducted a comprehensive review of governance models across 15 prominent UK TREs. These TREs were selected based on their direct involvement in collaborative analysis spanning the entire UK. Importantly, all of these TREs have met the legal basis for secure access to patient data, underscoring the crucial foundation of their operations. Our examination unveiled notable variations in the current approaches to data governance models. However, we have also highlighted that despite differing interpretations of the governance requirements as stipulated in GDPR, DPA and Section 251 [1–3], a substantial majority of these TREs exhibit the potential to converge towards standardising their existing governance frameworks. Furthermore, the preparedness of the legal basis for UK-TREs becomes even more evident by the establishment of multi TRE collaborations dedicated to COVID-19 investigations. Our findings strongly advocate for the pursuit of standardisation as the way forward to achieve multi TRE collaboration and the realization of federated analytics.

Our evaluation underscores that standardisation, particularly in four key domains, offer the means to break down existing barriers and eliminate segregation: a) **data access environment**: harmonising the infrastructure and protocols for establishment of TRE computing systems, b) **data access requirements**: streamlining and aligning the criteria and prerequisites for data access, c)**file out review procedures**: enabling uniform processes for the review of data extractions and d) **researcher accreditation/training**: promoting consistency in accreditation and trainings required by TREs.

While strides have been made towards harmonising data access and governance requirements, there also exist diverse range of technical solutions designed to enable federated analysis across multiple TREs. Engaging in a multi TRE approach for health data research, necessitates a steadfast commitment to data protection. We observed the presence of formal indicators of information security certifications, such as ISO-20071 across TREs [17]. This observation supports the notion of establishing a pop-up TRE, serving as an intermediary environment bridging multiple TREs, while concurrently emphasizing the importance of an overarching governance framework. The case examples we provided in this article were executed on the aggregated outputs that had passed the information governance requirement of each TREs and been analysed openly between collaborators; however, conducting pooled analysis on a comprehensive dataset comprising individual level multi-national data remains limited due to required governance complexities across multiple UK TREs. A pop-up TRE offers an organised solution for the pooled analysis of patient level data. While it is not the intention of a pop-up TRE to carry the full set of the datasets needed to support a research question, it is inevitable that data will 'move' into the pop-up TRE and may persist for the duration of the research project. Therefore, joint effort of contributing TREs is required in initial assessment of the projects as well as disclosure of outputs. Having a common framework in place will therefore act as a common language for governance requirements between TREs and can greatly contribute to achieving multi TRE analysis.

Requirements for an effective approach: the medical community is supportive of the idea of accessibility of data across multiple jurisdictions. Practically, a blended approach in technological implementation as well as public and patient engagement is needed [48]. Demonstrating that a common governance model can be achieved across the current UK TRE ecosystem has unveils great potential for optimising the use of siloed data by implementing appropriate governance frameworks. The siloed data flows can then join into mainstream flows of data and contribute to achieving diverse multinational insights as well as providing novel approaches with data for our sample validation and more generalizable findings.

Funders and data providers: we documented the two primary funding streams that sustain the operation of UK TREs: core funding and cost recovery. It is worth noting that some of these TREs operate under a hybrid model, blending elements of both core funding and cost recovery. The motivations for enabling multi TRE data governance are strong amongst the research community. However, this noble objective is also accompanied by well-recognized challenges inherent in the existing models of funding for each national TRE and the provision of data access through any single TRE. Implementing changes in the existing funding models requires careful considerations of several critical factors. These include the appropriate acknowledgement of data providers, protecting data provider ownership rights, as well as the established sovereignty of each TRE's resources by the main stakeholders. Furthermore, the significant investments and unwavering commitment of TREs' to ISMS serves as badges of honour to represent the "security" and "management" of the platforms. Transforming these attributes into quantifiable and comparable features across TREs is essential to enable a meaningful cross comparison of existing governance models. As we have shown, the two main existing costing models have a direct impact on resources available to TREs. Consequently, both funders and data providers should consider these models when assessing preparedness of TREs to participate in creating of multi TRE network. The cost recovery model permits operations on a project basis, albeit without guaranteeing the long-term existence of the infrastructure post project life span. In contrast, direct funding models, supported by core grants,

facilitate the sustainability of funded infrastructures. This approach effectively alleviates the operational cost burdens placed on researchers, thereby ensuring the enduring viability of the TREs.

In conclusion, the goal of achieving UK wide TRE standards must include a comprehensive mechanism to demonstrate adherence to GDPR principles and other legal bases for data sharing while enabling multi TRE research projects to continue to receive established data flows. Multi TRE analytics could be achieved via federation or alternatively within a pop-up TRE a key part of which is an appropriate and legally compliant governance wrapper.

Statement on conflicts of interest

ST, RAL, FT, ES declare funding from Dementias Platform UK 2 - Integrated Dementia Experimental Medicine (MR/T033371/1). RH and ST declare funding from SeRP and all organisations and research programmes using the SeRP platform. ST declared funding from SAIL Databank. Others had no competing interests.

Funding

This work has been funded by Dementia Platform UK 2 – integrated Dementia Experimental Medicine MR/T033371/1.

Ethics and dissemination

This project uses data openly available and all resources used are appropriately cited.

Authors' contributions

ST, FT, EM developed the proposal and the main conceptual idea. FT conceived and developed the research approach and manuscript of the paper and lead major revisions in each round. FT, CO, EM, SH, RH and ST contributed to development of the idea. FT, EM, SH, RH, CO, RAL and ST have discussed, reviewed and contributed to the final manuscript.

Public and patient involvement

This project is undertaken with a view of protecting public and patient data while enabling researchers to securely access and analyse valuable patient data. We have not used patient level data in this article. The ongoing work in this project will involve engagement of multiple public and patient representatives.

References

1. ICO, "What is personal data?," 2022, Accessed: Apr. 19, 2023. [Online]. Available: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/key-definitions/what-is-personal-data/.
2. "Section 251 documents - GOV.UK." https://www.gov.uk/government/collections/section-251-materials (accessed Apr. 29, 2023).
3. "Data Protection Act 2018." https://www.legislation.gov.uk/ukpga/2018/12/contents/enacted (accessed Apr. 29, 2023).
4. Gov.uk, "Art. 5 GDPR – Principles relating to processing of personal data - General Data Protection Regulation (GDPR)." https://gdpr-info.eu/art-5-gdpr/ (accessed Apr. 19, 2023).
5. U. H. D. R. Alliance and NHSX, "Building Trusted Research Environments - Principles and Best Practices; Towards TRE ecosystems," Dec. 2021, https://doi.org/10.5281/ZENODO.5767586
6. UKDS, "What is the Five Safes framework? — UK Data Service," 2017. https://ukdataservice.ac.uk/help/secure-lab/what-is-the-five-safes-framework/ (accessed Mar. 06, 2023).
7. "Information Security Management System (Pre-configured ISMS) Solution." https://www.isms.online/information-security-management-system-isms/ (accessed Apr. 29, 2023).
8. UKRI, "Impact Report." https://saildatabank.com/wp-content/uploads/2023/04/FINAL-NCSi-Impact-Report-2023.pdf (accessed Apr. 19, 2023).
9. "National Core Studies." https://www.hdruk.ac.uk/wp-content/uploads/2022/08/National-Core-Studies-July-2022-update-.pdf (accessed Apr. 19, 2022).
10. "Government invests £200 million in health data research - BHF." https://www.bhf.org.uk/what-we-do/news-from-the-bhf/news-archive/2022/march/government-invests-200-million-in-health-data-research (accessed Apr. 19, 2023).
11. T. Royal Society, "Learning data lessons: data access and sharing during COVID-19", Accessed: Feb. 14, 2023. [Online].
12. B. Goldacre, J. Morley, and N. Hamilton, "A review commissioned by the Secretary of State for Health and Social Care," 2022.
13. D. Pastor-Escuredo, A. Gardeazabal, J. Koo, A. Imai, and P. Treleaven, "Multi-scale governance and data for sustainable development," *Front. big data*, vol. 5, Dec. 2022. https://doi.org/10.3389/FDATA.2022.1025256
14. D. Heimbigner and D. McLeod, "A federated architecture for information management," *ACM Trans. Inf. Syst.*, vol. 3, no. 3, pp. 253–278, Jul. 1985, https://doi.org/10.1145/4229.4233
15. L. Floridi *et al.*, "Key Ethical Challenges in the European Medical Information Framework," *Minds Mach.*, vol. 29, no. 3, pp. 355–371, Sep. 2019, https://doi.org/10.1007/S11023-018-9467-4/METRICS

16. "EMIF-PLATFORM |EMIF." http://www.emif.eu/emif-in-practice/ (accessed Apr. 24, 2023).

17. "ISO/IEC 27001 Standard – Information Security Management Systems." https://www.iso.org/standard/27001 (accessed Apr. 29, 2023).

18. N. Digital, "TRUD." https://isd.digital.nhs.uk/trud3/user/guest/group/0/home (accessed Oct. 20, 2019).

19. "OpenSAFELY: Home." https://www.opensafely.org/ (accessed Feb. 15, 2023).

20. "ORCHID ::: Oxford-RCGP RSC." https://orchid.phc.ox.ac.uk/ (accessed Apr. 24, 2023).

21. M. T. O'Reilly, "The Turing Data Safe Haven: An open, scalable, reproducibly deployable, cloud-based Trusted Research Environment for working safely with sensitive data," Feb. 2023. https://doi.org/10.5281/ZENODO.7646620

22. "Home - Pioneer." https://www.pioneerdatahub.co.uk/ (accessed Apr. 24, 2023).

23. R. A. Lyons *et al.*, "The SAIL databank: Linking multiple health and social care datasets," *BMC Med. Inform. Decis. Mak.*, vol. 9, no. 1, 2009. https://doi.org/10.1186/1472-6947-9-3

24. "ISD Services | Electronic Data Research and Innovation Service (eDRIS) | Use of the National Safe Haven | ISD Scotland." https://www.isdscotland.org/products-and-services/edris/use-of-the-national-safe-haven/ (accessed Aug. 31, 2023).

25. "Honest Broker Service." https://hscbusiness.hscni.net/services/2454.htm (accessed Apr. 24, 2023).

26. UKHSA, "Approval standards and guidelines: data flow diagram - GOV.UK." https://www.gov.uk/government/publications/accessing-ukhsa-protected-data/approval-standards-and-guidelines-data-flow-diagram (accessed Aug. 31, 2023).

27. "Sensitive Data and Trusted Research Environments |Advanced Research Computing - UCL – University College London." https://www.ucl.ac.uk/advanced-research-computing/expertise/sensitive-data-and-trusted-research-environments (accessed Apr. 24, 2023).

28. "Research data environments |Administration and support services |Imperial College London." https://www.imperial.ac.uk/admin-services/ict/self-service/research-support/research-support-systems/research-data-environments/ (accessed Apr. 29, 2023).

29. "Welcome — DPUK." https://www.dementiasplatform.uk/ (accessed Mar. 07, 2023).

30. "Access data and samples |Avon Longitudinal Study of Parents and Children | University of Bristol." http://www.bristol.ac.uk/alspac/researchers/access/ (accessed Apr. 29, 2023).

31. "Apply | UK Longitudinal Linkage Collaboration." https://ukllc.ac.uk/apply/ (accessed Apr. 29, 2023).

32. "Integrated Research Application System." https://www.myresearchproject.org.uk/ (accessed Apr. 24, 2023).

33. "Guidance for organisations on processing of confidential patient information when the COPI Notices expire - NHS Digital." https://digital.nhs.uk/services/data-access-request-service-dars/copi-guidance (accessed Apr. 24, 2023).

34. "Scottish Parliament Long COVID Inquiry |The University of Edinburgh." https://www.ed.ac.uk/usher/eave-ii/informing-policy/scottish-parliament-long-covid-inquiry (accessed Apr. 24, 2023).

35. "Accessing UKHSA protected data - GOV.UK." https://www.gov.uk/government/publications/accessing-ukhsa-protected-data (accessed Apr. 24, 2023).

36. A. Wood, R. Denholm, S. Hollings, J. Cooper, and S. Ip, "Linked electronic health records for research on a nationwide cohort of more than 54 million people in England: data resource on behalf of the CVD-COVID-UK consortium," vol. 6. https://doi.org/10.1136/bmj.n826.

37. H. Abbasizanjani *et al.*, "Harmonising electronic health records for reproducible research: challenges, solutions and recommendations from a UK-wide COVID-19 research collaboration," *BMC Med. Inform. Decis. Mak.*, vol. 23, no. 1, pp. 1–15, Dec. 2023. https://doi.org/10.1186/S12911-022-02093-0/FIGURES/4

38. S. P. C. Brand *et al.*, "Forecasting the scale of the COVID-19 epidemic in Kenya," *medRxiv*, p. 2020.04.09.20059865, Apr. 2020. https://doi.org/10.1101/2020.04.09.20059865

39. "SAIL Datasets." https://saildatabank.com/saildata/sail-datasets/ (accessed Aug. 07, 2020).

40. "Secure data access for authorised users - NHS Digital." https://digital.nhs.uk/data-and-information/securedata-access-for-authorised-users (accessed Apr. 25, 2023).

41. S. Kerr *et al.*, "First dose ChAdOx1 and BNT162b2 COVID-19 vaccinations and cerebral venous sinus thrombosis: A pooled self-controlled case series study of 11.6 million individuals in England, Scotland, and Wales," *PLOS Med.*, vol. 19, no. 2, p. e1003927, 2022. https://doi.org/10.1371/JOURNAL.PMED.1003927

42. F. Torabi *et al.*, "Risk of thrombocytopenic, haemorrhagic and thromboembolic disorders following COVID-19 vaccination and positive test: A self-controlled case series analysis in Wales," *Nat. Sci. Reports*, 2022.

43. E. Vasileiou *et al.*, "Interim findings from first-dose mass COVID-19 vaccination roll-out and COVID-19 hospital admissions in Scotland: a national prospective cohort study," *Lancet*, vol. 397, no. 10285, pp. 1646–1657, May 2021. https://doi.org/10.1016/S0140-6736(21)00677-2/ATTACHMENT/0E9EEB7D-B073-4425-B336-8350C5A53BF8/MMC1.PDF

44. U. Agrawal *et al.*, "Severe COVID-19 outcomes after full vaccination of primary schedule and initial boosters: pooled analysis of national prospective cohort studies of 30 million individuals in England, Northern Ireland, Scotland, and Wales," *Lancet*, vol. 400, no. 10360, pp. 1305–1320, Oct. 2022. https://doi.org/10.1016/S0140-6736(22)01656-7

45. C. E. Dale *et al.*, "The impact of the COVID-19 pandemic on cardiovascular disease prevention and management," *Nat. Med. 2023 291*, vol. 29, no. 1, pp. 219–225, Jan. 2023. https://doi.org/10.1038/s41591-022-02158-7

46. R. Knight *et al.*, "Association of COVID-19 With Major Arterial and Venous Thrombotic Diseases: A Population-Wide Cohort Study of 48 Million Adults in England and Wales," *Circulation*, vol. 146, no. 12, pp. 892–906, Sep. 2022. https://doi.org/10.1161/CIRCULATIONAHA.122.060785

47. "National Data Resource (NDR) - Digital Health and Care Wales." https://dhcw.nhs.wales/national-data-resource/ (accessed Apr. 25, 2023).

48. T. Greenhalgh, L. Morris, J. C. Wyatt, G. Thomas, and K. Gunning, "Introducing a nationally shared electronic patient record: Case study comparison of Scotland, England, Wales and Northern Ireland," *Int. J. Med. Inform.*, vol. 82, no. 5, pp. e125–e138, May 2013. https://doi.org/10.1016/J.IJMEDINF.2013.01.002

Glossary of terms

ISMS:	Information Security Management System
TRE:	Trusted Research Environment
EMIF:	European Medical Information Framework
IRAS:	Integrated Research Application System
GDPR:	The General Data Protection Regulations
DPA:	Data Protection Act
DARs:	Data Access Requests
SDE:	Secure Data Environment