International Journal of Population Data Science

Journal Website: www.ijpds.org

Towards a standardised cross-sectoral data access agreement template for research: a core set of principles for data access within trusted research environments

Rachel Brophy[1,*], Ester Bellavia[1], Maeve Groot Bluemink[2], Katharine Evans[3], Munisa Hashimi[4], Yemi Macaulay[1], Edel McNamara[1], Allison Noble[5], Paola Quattroni[1], Amanda Rudczenko[4], Andrew D Morris[1], Cassie Smith[1], and Andy Boyd[1,3]

Submission History

Submitted:	**19/05/2023**
Accepted:	**10/08/2023**
Published:	**09/10/2023**

[1]Health Data Research UK, Gibbs Building, 215 Euston Road, London, NW1 2BE
[2]Our Future Health, 2 New Bailey, 6 Stanley Street, Manchester M3 5GS
[3]UK Longitudinal Linkage Collaboration, University of Bristol, Canynge Hall, Clifton, Bristol, BS8 2PS
[4]On behalf of the HDR UK Public Advisory Board
[5]Research Data Scotland, Bayes Centre, 47 Potterrow, Edinburgh, EH8 9BT

Abstract

Introduction
Trusted Research Environments (TREs) are secure computing environments that provide access to data for approved researchers to use in studies that can save and improve lives. TREs rely on Data Access Agreements (DAAs) to bind researchers and their organisations to the terms and conditions of accessing the infrastructure and data use. However, DAAs can be overly lengthy, complex, and can contain outdated terms from historical data sharing agreements for physical exchange of data. This is often cited as a cause of significant delays to legal review and research projects starting.

Objectives
The aim was to develop a standardised DAA optimised for data science in TREs across the UK and framed around the 'Five Safes framework' for trustworthy data use. The DAA is underpinned by principles of data access in TREs, the development of which is described in this paper.

Methods
The Pan-UK Data Governance Steering Group of the UK Health Data Research Alliance led the development of a core set of data access principles. This was informed by a benchmarking exercise of DAAs used by established TREs and consultation with public members and stakeholders.

Results
We have defined a core set of principles for TRE data access that can be mapped to a common set of DAA terms for UK-based TREs. Flexibility will be ensured by including terms specific to TREs or specific data/data owners in customisable annexes. Public views obtained through public involvement and engagement (PIE) activities are also reported.

Conclusions
These principles provide the foundation for a standardised UK TRE DAA template, designed to support the growing ecosystem of TREs. By providing a familiar structure and terms, this template aims to build trust among data owners and the UK public and to provide clarity to researchers on their obligations to protect the data. Widespread adoption is intended to accelerate health data research by enabling faster approval of projects, ultimately enabling more timely and effective research.

Keywords
data access agreement; DAA, data sharing agreement; data governance; public involvement and engagement; PIE; trusted research environments; TREs; secure data environments; SDEs; health data

*Corresponding Author:
Email Address: rachel.brophy@hdruk.ac.uk (Rachel Brophy)

https://doi.org/10.23889/ijpds.v8i4.2169

Introduction

The importance of access to near real time data, linked across organisations, for research analyses to inform policy was exemplified by the COVID-19 pandemic. This requires standardisation, interoperability, and responsiveness nationally and internationally to support data access and analysis. Yet the pandemic demonstrated that urgent, efficient, and trustworthy access to data requires the simplification, streamlining and coordination of the multiplicity of governance mechanisms currently in place [1].

Considering these needs, the Goldacre Review in the UK made recommendations for the 'better, broader and safer' use of NHS data for research and analysis and emphasised the need for standardised governance approaches [2]. In recognition of this, the UK government 'Data Saves Lives' strategy made a commitment to the implementation of Trusted Research Environments (TREs), also known as Secure Data Environments [3] and Data Safe Havens [4] (hereafter, we use TREs as a consistent, although not universally accepted term), at both national and regional levels to facilitate research and emphasise the importance of coordinated work amongst all stakeholders [3]. TREs are highly secure computing facilities that provide approved researchers with access to de-identified data within a controlled computational environment and in some cases only provide access to synthetic data, thereby preventing direct access to the real data.

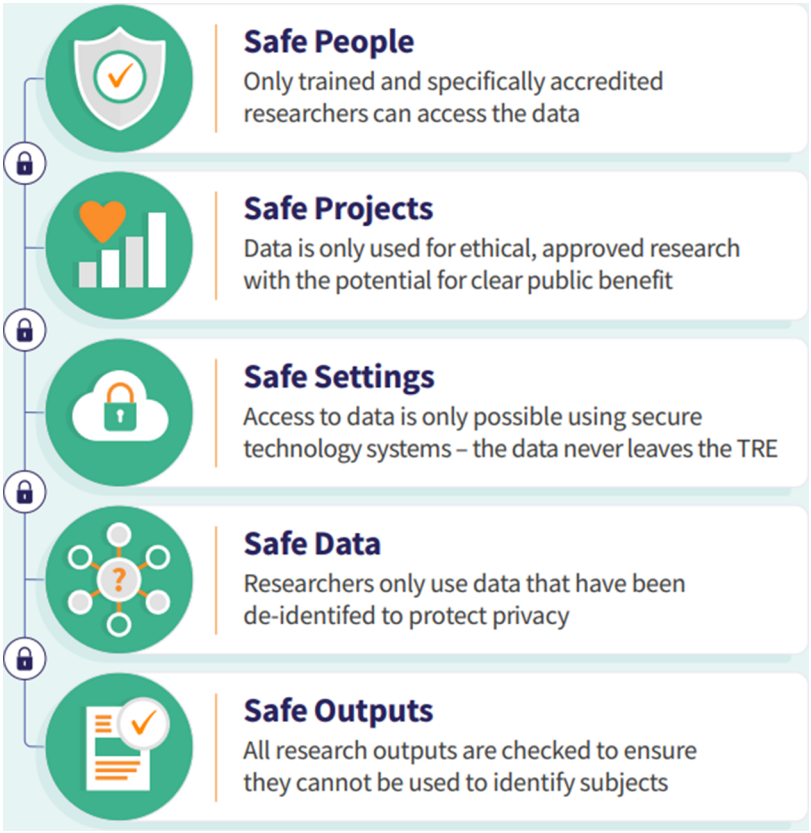
Regardless of the specific implementation, all data analysis takes place within the TRE, supported by a robust infrastructure that prevents personal data from being exported. The TRE hosts carefully curate 'research ready' datasets [5], eliminating the necessity of sharing data extracts for each project, thereby removing duplication of effort and potential risks to data security. TRE data access is not only cost and time-effective but provides opportunities for collaborative working, shared learning, and reduction of error. The UK Health Data Research Alliance White Paper 'Building Trusted Research Environments' offers best practice guidance on implementation of TREs and outlines the planned direction of travel for TREs to function as part of a federated infrastructure [6].

TREs will rapidly become the default way to access National Health Service (NHS) data for research [7] and are already the default mechanism for accessing non-health government records in the UK [8]. Reflecting this, the UK is now developing a network of TREs due to the improved safeguards that a TRE way of working brings; this network approach provides an opportunity to review the complex governing processes that consistently delay or create roadblocks for research projects vital to public benefit [9]. TREs sit within a wider landscape of Privacy Enhancing Technologies (PETs) that are designed to enable the useful derivation and analysis of data without providing full access to the data. PETs may be beneficial for cross-national collaboration and can facilitate compliance with regulatory requirements, although they should be combined with legally binding and enforceable obligations to protect data subject rights [10].

Extensive public dialogue and stakeholder deliberations have identified that for data science to be trustworthy, a suite of rigorous controls is needed to manage risks to confidentiality and to ensure data is used for public benefit purposes [11]. Aligned with this, data owners depositing their data into research infrastructures need to be confident that their underlying terms and conditions and the confidentiality of the data will be respected by the research end users. Yet to deliver research benefits, and to be responsive to pressing research needs, the controls also need to facilitate efficient data access for researchers. The challenge for data science infrastructure providers therefore is to balance safeguards with user efficiency and to minimise procedural burden in order to support scaling to a growing user base without undue access delays and within a cost constraint. Governance processes are underpinned by contracts to ensure all parties are aware and accountable for their obligations under data protection laws and other applicable legal requirements. Data Access Agreements (DAAs) for established TREs vary widely and often require time-consuming review and negotiation or contain legacy clauses that are rendered redundant when considering the robust data security and protection provided by TREs. Success in streamlining the format of research contracts has already been demonstrated through the Brunswick template agreements [12], covering material transfer, research collaboration, and studentship between non-commercial organisations. The UK Clinical Research Collaboration has also developed a suite of model agreements [13], which are widely utilised for clinical trials to avoid delays in contract review and initiation within the NHS and wider Health and Social Care, allowing more junior contract staff to undertake the review as these agreements are used 'off the shelf' and typically unmodified [14]. These existing models suggest an opportunity to improve the efficiency of TRE data access contracting through the development of a template DAA. Although the aim of this work is to harmonise de-identified data access in the evolving network of UK TREs, UK data protection law remains based on European Union (EU) law and therefore the template could be readily adapted for TREs established in the EU. Previous work between cross-national collaborators to develop a template data use agreement for access to 'factually anonymous' data via remote access to a safe room, provides an example of how the framework can be adapted to suit national, legal, institutional and technical requirements [15].

The UK Health Data Research Alliance [16], an independent alliance of leading healthcare and research organisations united to establish best practice for the ethical use of UK health data for research at scale, commits members to using a proportionate governance approach to data access based on the 'Five Safes framework' [17] via its principles for participation [18]. Within this framework presented in Figure 1, DAAs provide a vital control to bind researchers and their contractually responsible organisation to the terms and conditions imposed by the infrastructure and its constituent data owners. Specifically, the DAA will commit researchers to maintain data confidentiality, specify which data can be used, by whom, and for which purpose, the circumstances in which data access can be revoked, and to define how outputs from the research process – including results and reusable research outputs such as derived data – are used. The DAA will also provide the basis for other controls – such as researcher audits – and to define personal and institutional liabilities for data misuse.

Figure 1: Five safes framework



The Alliance has convened a Pan-UK Data Governance Steering Group [19] with partners at the UK's Office for National Statistics, with members that are data science professionals, associated government data owners and members of the public. Members of the Pan-UK Data Governance Steering Group were invited to join a TRE Legal Toolkit Action Force and were asked to volunteer contacts within their organisation with relevant knowledge and expertise. The Action Force has members from the following organisations: Bennett Institute for Applied Data Science (University of Oxford, England), DataLoch (Scotland), Health and Social Care (HSC) (Northern Ireland), Health Data Research UK (HDR UK), NHS England, NHS Health Research Authority (HRA), Office for National Statistics (ONS), OpenSAFELY, Our Future Health, Public Health Scotland, Research Data Scotland, Secure Anonymised Information Linkage (SAIL) Databank (Swansea University, Wales), UK Longitudinal Linkage Collaboration (University of Bristol, England), and Wales Cancer TRE Project (Cardiff University, Wales). Members of the public were also invited to join the TRE Legal Toolkit Action Force, given the valuable perspective on data access and governance they can provide. The Action Force has the objective of developing a set of standardised legal agreements, including templates and related guidance for a DAA, data depositing agreement (DDA), and a data protection impact assessment (DPIA) or other forms of risk assessment, to enhance clarity and consistency in contractual arrangements and provide researchers with a user-friendly toolkit that can reduce the administrative burden and accelerate contracting within institutions. Developing a standardised DAA template for UK TREs [41] was identified as a priority to improve the efficiency of the data access process and to enable improved public understanding and scrutiny of this key safeguard.

The ability to describe controls and governance processes in a manner that promotes public understanding is essential. A review of public attitudes towards administrative data sharing for research, covering studies over a decade to 2018, revealed that the public is generally supportive when three core conditions are met: privacy and security, public interest, and trust and transparency [20]. Additionally, a separate review underscored the public's willingness to share health data for public good research, subject to the assurance of addressing security concerns and prioritising transparency and inclusivity of stakeholder perspectives [21]. While these studies encounter some differences in findings, they both highlight the close link between increased public understanding and trust in data research, lending credibility to the Five Safes framework as a valuable conceptual structure.

Concerns about potential individual harm from commercial access to data, such as discrimination by insurance companies, or collective harm, such as the sale of data for profit-making purposes or questionable agendas, can be addressed through transparent processes and clear communication. While quantitative studies have indicated that the public, in principle, do not favour sharing their health data with private companies [22], further qualitative research into the nuances of commercial involvement has shown greater support, particularly if there is thought to be societal value to the sharing [23]. The National Data Guardian's guidance on evaluating public benefit in research recognised that the public are in support of a 'net good' accruing to the public.

This may include instances of commercial profit-making, where this is proportionate and leads to demonstrable improvement in NHS services, knowledge, and insights, and is underpinned by an assessment of 'fairness' [24]. The Association of the British Pharmaceutical Industry principles of analysis and use of health data commit pharmaceutical industry members to: transparency; fairness with an appropriate balance of commercial and public benefit and return to the researcher and patient communities; legal and regulatory compliance; and patient and public involvement and engagement (referred to as public involvement and engagement [PIE] in this article) [25]. Aiming for 'trust' may not be sufficient when it comes to commercial involvement as the concept suggests that the individual must be depended on to demonstrate 'trustworthiness'. Rather the system within which they operate should be confidence-worthy [26], and with effective communication of the workings of TREs and the associated governance processes (including the DAA), the public have the opportunity to assess whether this is the case. PIE is essential to ensure that research is driven by a commitment to delivering public benefits and is worthy of public trust and confidence. Accessing and using people's data is a privilege, and including public members in data-driven research processes can ensure these are transparent, open, and accessible [27]. Consistency in information governance and contractual standards enhances protection for data subject rights and effective PIE is required so that the public can be assured that standards are not being lowered in the interest of speed of access.

In this paper, we present a core set of principles for data access within TREs. We discuss the value of aligning these principles to established frameworks and explore the benefits of creating a DAA template in collaboration with key stakeholders and members of the public.

Methods

Development of data access principles

We conducted a benchmarking exercise where DAAs from nine established UK TREs (Table 1) were compared, including those from Clinical Practice Research Datalink (CPRD), DataLoch, eDRIS (Public Health Scotland), Genomics England, Honest Broker Service (HSC Northern Ireland), OpenSAFELY, SAIL Databank, UK Data Service and UK Longitudinal Linkage Collaboration. We identified that for some TREs, DAAs were supplemented by additional documents such as end-user terms and policies. Where used, these documents were included in the benchmarking to capture the full set of terms and conditions. Author RB reviewed all materials and categorised the terms and conditions into the following themes: 'Data available', 'Access', 'Outputs', 'Use of Data', 'Intellectual property', 'Liability for data accuracy and availability', 'Compliance with Data Protection Legislation and Liability', 'Commercial use', 'Onward linkage', 'Open sharing of analysis, code and derived data', 'Freedom of information' and 'Others'. Any terms that did not appear to fit in to one of the categories initially were included in the 'Others' section and these were reviewed for additional themes, then allocated to the most suitable existing category. Each category was reviewed for areas of commonality, which were determined by identifying the most prevalent approach. Areas where variance between agreements was discovered were labelled 'alternative/additional' terms and subsequently brought to the Action Force for discussion.

To provide clarity on the inclusion of each principle and resulting DAA term, and to create alignment with the wider governance systems surrounding data access, the principles of data access were categorised and mapped against the Five Safes framework (Figure 1) by CS and RB. This was achieved by matching the principle to the most appropriate of 'people', 'projects', 'settings', 'data' or 'outputs' to align with the framework.

Public involvement and engagement

We invited public members from HDR UK's Public Advisory Board [28] to contribute to developing the DAA template, by participating in an activity designed according to the UK Standards for Public Involvement [29]. We provided each contributor with materials describing the purpose, responsibilities, and expectations for the exercise (see Supplementary Appendix 1). The contributors were provided with the initial draft of the DAA principles, including explanatory account and a key point of contact to address any queries. Each of the principle subject areas had questions to stimulate thought and discussion. An online workshop was held to discuss the materials and questions, to evaluate the accessibility of the material shared, and to clarify expectations. As the public members had varying experience of data access governance and research infrastructures, the workshop also gave an opportunity to build their understanding of TREs. While public members were offered the option to provide feedback via email, video conference, or phone call, they preferred sharing their comments via email. An honorarium payment was issued to all public contributors, in line with National Institute for Health and Care Research guidance [30]. An overview of the feedback received is provided in the results section, detailing how this has informed the development of the DAA template. The public contributors were given the opportunity to review this article prior to submission to validate and agree to its content. Some of them accepted an invitation to contribute as co-authors.

Results

The principles of data access

Through the benchmarking exercise of DAAs in place for established TREs, we found many areas of commonality that allowed the development of the core principles of data access. To help describe our findings we refer to three parties to the principles, the Research User's Organisation (UO), the TRE Host Organisation (HO) and the Approved Researcher (AR), where the UO is responsible for individuals accessing the TRE, the AR is affiliated with the UO and the HO hosts and controls a TRE and is typically Data Controller of the data within this environment. In many cases, these data are deposited within the TRE by a third-party data owner under a separate legal agreement.

Table 1: UK TREs included in the DAA benchmarking process

TRE name (Acronym)	Data holdings	Geography	Population coverage	URL
CPRD	Primary care data (Patient electronic GP health records)	UK	All of England, Scotland, Northern Ireland, and Wales	cprd.com
DataLoch	Routinely collected health care data from primary and secondary care	Lothian region, Scotland	Routinely collected data as part of daily interactions with health and social care services, Lothian region	dataloch.org
eDRIS	Secondary care data, administrative data	Scotland	The population of Scotland	isdscotland.org/products-and-services/edris/
Genomics England	Genomics data, secondary care data, mortality data, omics	England	Consented participants across England	genomicsengland.co.uk
Honest Broker Service	Secondary care health data, data from the integrated health and social care system, mortality data	Northern Ireland	The population of Northern Ireland	hscbusiness.hscni.net/services/2454.htm
OpenSAFELY	Primary and secondary care electronic health records	England	The population of England	opensafely.org
SAIL Databank	Secondary care health data, administrative data	Wales	The population of Wales	saildatabank.com
UK Data Service	Economic, population and social research datasets	UK	UK population (census), participants of multiple national and cross-national surveys, longitudinal studies (over 6000 datasets in TRE)	ukdataservice.co.uk
UK Longitudinal Linkage Collaboration	Research Study, NHS health records, socio-economic records, environmental exposures	UK	Participants of 24 UK longitudinal population studies.	ukllc.ac.uk

As a pre-requisite to the DAA being issued, the TREs all required the intended AR to submit an application form. This triggered the HO application assessment process, where ARs are required to demonstrate that their proposal is an appropriate and ethical use of the data, that it will deliver clear public benefits and that they will publish their results to enable use, scrutiny, and further research. 'Data' means the data fields and datasets to which the AR has been approved access.

The core principles of data access in TREs are listed under their assigned Five Safes category below (Tables 2–6). The 'Alternative/additional' DAA terms were discussed by the Action Force, and it was recognised and acknowledged that there will always be the requirement for areas of variance between different TREs and in some cases variation between different UOs ('Customisable controls relevant to some TREs' column).

Public members' feedback

The concept of a core set of principles and associated DAA received positive support, with some public members recognising the potential to promote efficiency and standardisation in data access processes.

> *"Overall, I think it is a fantastic idea which could potentially streamline data access, but also ensure that a universal system and standards are in place."*

While some members of the public were fully in support over the proposed distribution of responsibility, there were some concerns. For instance, some highlighted the difficulty of overseeing a large number of researchers working on a single TRE.

> *"An institution may have 100 researchers working on a single TRE, the oversight of all these researchers would be nearly impossible."*

Others shared their concerns over placing too much responsibility on the UO, which could slow down the data access process.

> *"I believe that placing so much responsibility on the User Organisation may have contradictory effect on streamlining the process. Admin processes in User Organisations may elongate data access for the researcher as opposed to when the researcher can sign documents on their own behalf."*

Table 2: Safe people

Safe people	Customisable controls relevant to some TREs
The DAA is entered into by the UO (via signature by an authorised signatory) rather than by the individual AR(s). • The UO shall ensure: ○ AR(s) are aware of their obligations ○ AR(s') compliance with the DAA terms. (AR(s) shall be advised of their obligations when accessing the TRE via 'Terms of Use', which may present as a "click through" set of terms accepted at the point of data access.) ○ that access credentials are not shared by its ARs, so that access to and use of the Data in the TRE is by AR(s) and not by any other persons. ○ that departures of any AR(s) are reported to the HO, and the AR(s) do not attempt to access the Data or the TRE after termination or expiry of the Agreement. ○ that AR(s) are affiliated with them and warrants that the AR(s) are appropriately trained and skilled in data protection, confidentiality, governance, and security. • The HO: ○ issues credentials to the ARs provided by the UO, and revokes permissions on notification that an AR is leaving the UO or should no longer have access. ○ will impose restrictions or suspension of TRE access to the UO and/or ARs if they are subject to an investigation, incident, or breach. • The AR(s) shall keep confidential the Data, and any access credentials to the Data and shall report any incidents or breaches to the UO and HO as soon as possible.	• Researcher accreditation/information governance and data protection training requirements. • TRE policy regarding penalties and remediation required for non-compliance, offences, and breaches.

Table 3: Safe projects

Safe projects	Customisable controls relevant to some TREs
• AR(s) shall only be permitted to access the TRE and use the Data for purposes defined in the Approved Project, with public good criteria. • The HO shall publish accurate and up to date details of the Approved Project and associated AR(s) in a publicly available data use register. • Commercial use will be permitted only if stated in approvals granted prior to project starting. Any unauthorised commercial or non-commercial use will result in termination. • Further research requires new approval. Non-compliance results in termination.	• The cost recovery policy for data access. • The process for submitting amendments to the scope of the project, data requested, or project research team, or extensions to the term of the Approved Project.

Finally, some discussed UO liability and that researchers should be made aware of their responsibilities (as planned with the researcher 'Terms of Use') to promote good practice.

"I understand from a legal standpoint that the user organisation should be liable, and it is probably difficult to have legal action against an individual. However, in the interest of best practice, it should

Table 4: Safe settings

Safe settings	Customisable controls relevant to some TREs
• The TRE infrastructure provides data protection and security assurances, demonstrated with appropriate accreditations. • AR(s) must access data on a device that meets the security requirements of the HO and UO (typically not a personal device), shall not leave it unattended while accessing the TRE and shall protect the screen from onlookers. • No remote access to the TRE from outside permitted locations defined in the project approval process.	• End user security requirements specified by the HO. • Terms for researcher access via Virtual Private Network (VPN) at other locations within UK. • Restrictions on international access and additional terms required for international access. • The protocol on monitoring and audit of access and use of the TRE.

always be made abundantly clear to the researcher of their responsibilities."

Since prospective datasets are often subject to periodic data quality and cleaning exercises and overall, it is often unclear whether the available datasets will produce valuable results or answer the research question, the HO includes clarification to this effect in the contract. However, public members expressed concerns about conducting research on incomplete or inaccurate data, with one member commenting:

"As a member of the public, it sounds as though research can be based on faulty data."

In relation to the DAA terms and conditions, public members acknowledged that they will have been developed with extensive consultation. Therefore, any changes to these terms should only be made if there is a good reason to do so.

"Ultimately, these terms and conditions have been best considered for all stakeholders so there would need to be an extremely important reason to customise them."

They also believed that if members of the public have been involved in developing and finalising these terms and conditions, any subsequent changes made without consulting them would undermine the value of their input.

"If there has been PIE involvement in finalising the terms and conditions and then these were to be changed, this would go against the meaningfulness of PIE."

A suggestion put forward from a member of the public was to ensure that each of the customisable annexes meet a set level of acceptability.

"Each of the above [customisable annexes] needs to be explicitly covered to a core minimum standard. Customisation applies to additional terms and conditions beyond this minimum."

The concept of researcher training acting as a security assurance was supported by public members, who recognised the possibility of mistakes, which could be addressed by the TRE being a 'safe setting' and the principles of 'safe outputs'.

"The likelihood of data misuse is minimal, unless someone is being malevolent. This should be picked up by being an 'approved researcher'."

"I believe that most errors in data safety occur by mistake (human error) and so all the training in the world will not prevent mistakes from occurring. I think that's one of the safety features of the TRE system, is that human error can occur in a controlled environment and be monitored closely."

Responses to commercial involvement in TRE data access was met with mixed responses, highlighting the need for caution and clear communication.

"I am not sure what I think/feel about commercial use. It would probably depend on the individual case."

"Personally, I do not have any issues with commercial use of results from approved projects, if and when approached correctly. Although commercial use requires a more detailed explanation and an outline of the processes in place to protect the data."

"Need to consider forms of licencing so that the public benefit from commercial products based on public data."

Regarding the commercial use of data, particular attention was also given to the importance of involving members of the public in developing criteria that inform decision-making about granting access to data, with some public members questioning the ownership of intellectual property (IP) and whether researchers who develop results with the data should have the choice to move forward with commercial use if they desire.

"What are the criteria for approving commercial use? Who has decided these? There needs to be PIE involvement in developing these criteria and in making these approvals."

Table 5: Safe data

Safe data	Customisable controls relevant to some TREs
• Individual-level linked/linkable datasets will be available in the TRE. Data is de-identified before access is granted. • Only anonymised aggregate data may be downloaded following quarantine and screening prior to release from the TRE. • The AR(s) shall not, and shall not attempt to, link or combine the Data with other information or data (including any information relating to an identified or identifiable natural person) available to the UO. • Although the possibility to do so is extremely low in TREs, the AR(s) shall not, and shall not attempt to, and the UO is responsible for ensuring that the AR does not: (i) identify individuals from the Data; or (ii) contact any data subject. • Each party shall comply with their respective obligations under data protection law, including UK General Data Protection Regulation (UK GDPR) and the Data Protection Act 2018. • The UO acknowledges and agrees that it has sole responsibility, and the HO takes no responsibility, for interpretation or further analysis of the Data. • The UO shall be responsible for ensuring that the HO is informed without delay, and in any event within 12 hours of the UO or AR becoming aware of: i. any unauthorised access, disclosure, loss, damage, or alteration of the Data ii. any element within the Data that might permit the identification of a data subject, iii. anything that may impact or compromise the confidentiality, integrity, or availability of the Data, iv. any complaints from an individual or supervisory authority in relation to the Data; and v. any request from a research participant to exercise their rights in respect of the Data. • It is generally not possible to know with certainty at the outset of research whether access to and use of the Data will answer the relevant research question or produce valuable outputs. Therefore, to the fullest extent permitted under Applicable Laws, the HO makes no warranty as to the quality of the data and is not liable for data unavailability.	• Specific data field names for all (linked/linkable) datasets available. [no actual data] • Conditions relating to the linkage of datasets held within the TRE and any permitted datasets that may be ingested into the TRE for linkage. • Additional terms for NHS data.

"If the researcher develops results with the data and the IP changes from the host to the researcher, is it not then the researcher's choice to move forward with commercial use if they desire? Overall, a better explanation of commercial use for example in the glossary would be beneficial."

Members of the public involved underscored the importance of open sharing of syntax and code as a way to improve accuracy in research and enhance transparency.

"It is important to make the methods known so that if errors have occurred this can be flagged up and false results are not being published."

"Methodologies should also be shared for the best interest of the public. It is important for both public and patients to understand how their data is being used."

A common theme in the feedback received was the requirement for further standardisation in monitoring the

Table 6: Safe outputs

Safe outputs	Customisable controls relevant to some TREs
• Research outputs must not include personal Data. • The AR(s) shall not: i. use the Data or any research output for any purpose contrary to the applicable laws or Approved Project purposes. ii. download, extract, transmit, transfer, remove, share, copy, or publish any of the Data from the TRE. • There is no transfer of intellectual property rights of the Data or the TRE	• HO terms and conditions/protocol on output approval. • Any acknowledgement of source/ copyright statement requirements in publications or other forms of dissemination. • The HO may specify intellectual property ownership for source data, linked data, derived data, metadata, researcher analyses and research outputs. • Detail on ownership of syntax and methodology and open-source sharing (where ownership remains with UO, but they shall grant a licence for use for other research and non-commercial purposes).

use of the TRE and in handling breaches in order to earn public trust. Public members emphasised the importance of clarity and transparency around the consequences of a breach of agreement, identifying the need for standardised access restrictions for non-compliance.

> *"Monitoring of the use of the TRE is not clear. There needs to be a template covering this."*

> *"I think that there needs to be some kind of standard expectations that user organisations and researchers comply with. What are the consequences of non-compliance from the researcher?"*

> *"User Organisation/Researcher access restrictions if subject to an investigation/breach, will this be made publicly available?"*

> *"Access restrictions if there is a breach – who decides what these will be? Should these be standardised so that user organisations agree to them?"*

Discussion

The principles as building blocks to the DAA

The principles and by extension the associated DAA clauses are intended to be concise but comprehensive. They are risk-proportionate in line with the inherent data security and protection assurances provided in TREs by design and function, and therefore remove outdated terms which reflects the progression from data dissemination to data access. For example, data destruction clauses are not included and not relevant to the AR or UO where no personal data will ever leave the TRE. The template DAA will follow the data access principles, whereby clauses will be mapped against the Five Safes framework to create a familiar structure and to aid ease of use. The advantage of mapping principles to the Five Safes framework is the ability to achieve successful data access while maintaining ethical oversight [31]. The framework has been used widely across public sector projects due to its reputation of encouraging safety by design in a way that is agile and proportionate [17]. Public trust is a key component to the process of successful data access and individuals have a right to revoke access to personal data. A further advantage of using the Five Safes structure is the effective messaging this can provide to aid public understanding of how data security and processing are managed [32]. Although the PIE feedback suggests the need to clearly communicate the role and safeguards inherent in the DAA in the context of wider controls implemented elsewhere in TRE processes (e.g. the application review process).

The TRE infrastructure provides a 'safe setting' [6] and the DAA allows the HO to specify access restrictions based on either end user security requirements, location, or both, and how access will be monitored and audited. The AR is made aware of, and acknowledges understanding of, the practical requirements to keep access credentials protected and the UO has a responsibility to ensure they are appropriately trained and aware of their obligations. Only ARs can access the TRE and the HO has a responsibility to restrict access and deal with breaches. A combination of all three parties' responsibilities ensure 'Safe people' is achieved. 'Safe projects' is met by ensuring the AR works only to the approved project and purposes in the TRE, the requirement not to act outside this and to preserve the confidentiality of the data. Not only is the AR made aware of this, but the UO becomes contractually responsible for compliance. 'Safe data' is met by the provision of de-identified data in the TRE by the HO, with set limitations on linkage. Anything that compromises the security or confidentiality of the data must be reported, and these responsibilities on the AR and UO will be contained within the DAA, along with managing requests from individuals to exercise their rights. No personal data ever leaves

the TRE, which keeps the data safe and only 'safe outputs' are therefore released following the HO process, appended to the DAA. The outputs must only be used in a way that is compliant with the terms of the DAA, and further to this intellectual property rights and open-source sharing can be detailed in the annexes.

Implementation and acceptability of the DAA

For the DAA to succeed in speeding up the contracting process and present as a trusted and recognised template, the core DAA will come with a strong recommendation that it should not be modified by any party. Modifications may also impact the integrity of the TRE and any accreditations so this must be emphasised. The customisable annexes can be modified, but this should fall within the remit specified in the guidance and best practice recommendations that will be included with each, and not used as an opportunity to copy and paste old agreement terms or add overly cautious extensive clauses.

Introducing standardisation of the UO (rather than all ARs) acting as signatory to the DAA template may be met with mixed responses. The effectiveness of this accountability is heavily dependent on the oversight of those acting as signatory on the running of research projects and the interaction with researchers, which can be determined by both the infrastructure within institutions and the number of active researchers. The public member feedback demonstrates concern in that the process with the UO will add an extra step, rather than the efficiency that we are aiming to achieve, and that it is far quicker to request signature from the researcher. However, the hope is that with introduction of the standardised template, this will encourage harmonisation and clear delineation of roles as well as the opportunity for research-active institutions to have full awareness of projects being run from the outset, and to develop efficient oversight processes where required. The most impactful practical step within this streamlining initiative will be in making the researcher aware of their individual responsibilities via the Terms of Use. The implementation of a standardised template will result in efficiencies for both controllers and researchers in the early stages of data access processes. This may carry forward benefits with regards to researcher experience and their project outputs. Researchers face a number of pressures when undertaking data-led work, including funding and project timelines. Therefore, the standardising of processes can potentially eliminate scenarios where researchers may face dead spaces of research time while awaiting approval from institutions due to procedural inefficiencies [33].

Established TREs may be reluctant to change processes in place, therefore emphasising the importance of familiarity of a contract structure and content and the impact this will have on improving time to data access from approval will be important here. The strong public involvement in the development of the principles and the template should also act as driver for adoption. Where an established TRE HO may not immediately have the resources to manage the adoption of a new DAA operationally, alignment with the principles discussed within this paper is a move towards standardisation and will be encouraged.

Commercial research

The principles of data access are intended to apply to both non-commercial and commercial party involvement and give assurance that both types of organisations are working to the same standards. Feedback from industry representatives has consistently highlighted the requirement for flexibility in terms of open sharing and intellectual property rights, where commercial protection can be justified where there has been significant investment into research and development activities and when it would not be appropriate to share these. Where there may be a more standardised approach in non-commercial research, room for commercial flexibility has been accounted for in the DAA structure, where this permitted by the TRE or its constituent data owners, by making open sharing (of code, syntax, or methodology) and intellectual property rights customisable annexes. Larger commercial organisations frequently have affiliates across the globe; therefore, the TRE may specify restrictions and allowances in the international access annex. Reference to national laws and regulations in DAAs can discourage researchers from applying to access data [34]. By making the AR aware of their obligations via Terms of Use makes it clearer for the individual to acknowledge what this means in practice, and it is the UO's centralised responsibility to ensure compliance with laws and regulations.

The value of PIE

The development of core principles and a DAA received positive support from public members, recognising the potential for efficiency and standardisation in data access processes. The feedback raised by the public members on the distribution of responsibility identifies that it is very rare to take legal action against an individual in these circumstances. In promoting organisational-level responsibility there will be accountability and formal review of the DAA rather than just signature from an individual. Concerns over clarity of roles and obligations for the UO and AR highlight the need to clearly communicate how this process works to both these parties and the public. Public members stressed the importance of researchers being aware of their responsibilities and the need for good practices and the concept of researcher training as a security assurance in data access. Open sharing of syntax and code was seen as crucial for research accuracy and transparency. Additionally, standardisation in monitoring the TRE and handling breaches, as well as clear consequences for non-compliance, were deemed necessary to build public trust. Concerns expressed over the HO making no warranty as to the accuracy and quality of the data has led to a revision of the wording in the principles to place more emphasis on data quality rather than implying careless inaccuracy, which will be reflected in the DAA.

Overall, the feedback received generated further reflection in several areas and highlighted the value of including the public in decision-making processes around data access.

The need for a robust 'social contract' on national and international level becomes increasingly crucial as personal data collection rates escalates [35]. This is particularly relevant in health data research, where the individuals providing the data can be far removed from the data processing. Recent experiences, such as the closure of the Care.data programme

[36] and the postponement of the General Practice Data for Planning and Research Direction (GPDPR) programme, have underscored the repercussions of an inadequate social contract with criticism centred on insufficient public engagement and information dissemination [37]. While PIE traditionally focuses on the development of research projects, approving access, and dissemination of findings [38], in this case, the public was invited to actively participate in improving research-related contracts. The inclusion of PIE in data governance is vital to ensure that data is used safely and for public benefit. Building a trustworthy and transparent framework for data access and utilisation requires active involvement of the public, recognising their role in shaping the governance processes and establishing a strong social contract that upholds rational social cooperation.

Ongoing work and further standardisation

The nature of developing a DAA template that is trusted, accepted, and proportionate requires incorporation of wide stakeholder and public participation. Research Contracts Leads from universities across the United Kingdom and industry representatives have reviewed the principles and will input to the DAA template to ensure maximum acceptability and functionality. The public members consulted expressed that easy-read publicly accessible versions of all documents should be produced, along with input to any guidance. Bringing PIE to the development stage was thought to strengthen the process further and it was raised that there is a need for more PIE in governance processes in general. There is a commitment to embed PIE throughout, with public members joining the Action Force to oversee and input to all developments and to develop lay guidance.

This work has highlighted the potential for further standardisation, which will aid transparency, essential for gaining public trust. Key to this area, is the use or misuse of the customisable annexes to the DAA, where room has been allowed for TRE-dependent clauses to be included. Although essential to include this level of flexibility, it is vital that the customisable annexes are restricted to their defined purpose, and not used as a mechanism to include unnecessary clauses or caveats that will defeat the purpose of all adhering to the core principles and agreement. Particularly of concern in this area is that researcher accreditation is standardised so that all researchers have had the same quality of training ahead of accessing data. This is a notion supported by the Goldacre Review [2] where the recommendation is that a single accreditation scheme should be in place that mirrors the Office for National Statistics (ONS) accredited researcher scheme [39]. The same recommendation has been added as best practice guidance in the DAA appendix on researcher training and is an area that would benefit from widespread consensus and adoption. The function of the HO in monitoring or auditing use of the TRE should meet a standard and the process should be publicly available. A well-maintained data use register provides details to the public on who is accessing the data and the purposes for access [40], but assurance that there are restrictions for access to only authorised persons and audit of data use also needs to be provided, with a requirement that this is actively monitored, and the process made publicly available. Similarly, assurance that breaches and non-compliance will be dealt with to a defined and appropriate standard will aid public confidence and is an area that could benefit from regulatory alignment [2], collaboration and effective communication. The DAA template is being developed to harmonise data governance across UK TREs and can be readily adapted for the EU given post-Brexit UK data protection laws remain largely aligned with EU laws. TRE HOs will have the opportunity to add terms of international access in a customisable annex where this is already permitted and there is potential for further work to ensure adaptability and alignment with international requirements, involving representative public input to understand variance in public sensitivities and perceptions. The DAA template may be further tailored to be used on a global scale, and this will be informed by endeavours of the Pan-UK Data Governance Steering Group to reach consensus in this area.

Conclusion

DAAs governing data access for research in UK TREs are varied, complex, and cause delays to approved research projects while they undergo legal review. With the widening network of TREs, an opportunity to streamline presents itself. The Pan-UK Data Governance Steering Group, a working group of the UK Health Data Research Alliance, has developed a set of core principles of data access for research in TREs, to underpin a template DAA. These will apply to non-commercial and commercial access and the aim is for widespread adoption of the template to provide clarity and transparency on roles and obligations of all parties involved. Variability between TREs and organisations has been accounted for in the development of customisable annexes. Clear and consistent language, with PIE in the development of both the principles and the template DAA, will rightly offer the public the ability to assess their own levels of confidence and trust in the security assurances offered by TREs and associated governance processes. This has been further assisted by mapping both the principles and the DAA template to the widely trusted Five Safes framework. The principles provide a mechanism by which all TRE access can meet the same standards with equivalent delineation of roles, avoiding undue delays and introducing clarity to the research community. This work highlights opportunities for further collaborative work with key stakeholders to achieve streamlining, standardisation, and transparency in data access governance processes, both nationally and internationally.

Acknowledgments

This work was conducted by the TRE Legal Toolkit Action Force, as part of the Pan-UK Data Governance Steering Group of the UK Health Data Research Alliance and supported by HDR UK. HDR UK is funded by UK Research and Innovation, the Medical Research Council, the British Heart Foundation, Cancer Research UK, the National Institute for Health and Care Research, the Economic and Social Research Council, the Engineering and Physical Sciences Research Council, Health and Care Research Wales, Health and Social Care Research and Development Division (Public Health Agency, Northern Ireland), Chief Scientist Office of the Scottish Government

Health and Social Care Directorate. We would like to express our grateful thanks to the Steering Group, co-convened by Professor Sir Ian Diamond (UK National Statistician), for their commitment to streamlining and discussions on this topic, and the members of the Action Force for their work, including representatives from: Bennett Institute for Applied Data Science (University of Oxford, England), DataLoch (Scotland), Health and Social Care (HSC) (Northern Ireland), HDR UK, NHS England, NHS HRA, ONS, OpenSAFELY, Our Future Health, Public Health Scotland, Research Data Scotland, SAIL Databank (Swansea University, Wales), UK LLC (University of Bristol, England), and Wales Cancer TRE Project (Cardiff University, Wales). We would also like to thank the following organisations for making their DAAs available for us to review: CPRD, DataLoch, eDRIS (Public Health Scotland), Genomics England, Honest Broker Service (HSC Northern Ireland), OpenSAFELY, SAIL Databank, UK Data Service and UK LLC.

We would like to thank Tony Plant and Roger Gibb, who, as members of the HDR UK's Public Advisory Board, contributed to the development of the principles, in addition to Munisa Hashimi and Amanda Rudczenko.

We would like to thank the Research Contracts Leads from the following organisations for their review and input to the principles and ongoing work with the DAA template: University of Bristol, University of Nottingham, University of Oxford, Queen's University Belfast, University of Southampton, representatives from Astra Zeneca for their input towards commercial acceptability, and Bird&Bird for their legal input.

Statement on conflicts of Interest

None declared.

Ethics statement

This paper reports no original data and is therefore exempt from ethical review. We have permission from the public members to use verbatim quotes in this paper.

References

1. Sir Patrick Vallance – Frontiers Meeting – 4 November 2022. Available from: https://www.youtube.com/watch?v=k94BYv_uAlo [Accessed 04.05.2023]

2. Goldacre B, et al. Better, Broader, Safer: Using Health Data for Research and Analysis. A review commissioned by the Secretary of State for Health and Social Care. 2022;April. Available here: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1067053/goldacre-review-using-health-data-for-research-and-analysis.pdf [Accessed 04.05.2023].

3. Department of Health and Social Care. Data saves lives: reshaping health and social care with data. 2022;Jun. Available here: https://www.gov.uk/government/publications/data-saves-lives-reshaping-health-and-social-care-with-data/data-saves-lives-reshaping-health-and-social-care-with-data [Accessed 04.05.2023].

4. Burton PR, Murtagh MJ, Boyd A, Williams JB, Dove ES, Wallace SE, Tasse AM, Little J, Chisholm RL, Gaye A, Hveem K. Data Safe Havens in health research and healthcare. Bioinformatics. 2015 Oct 15;31(20):3241-8. https://doi.org/10.1093/bioinformatics/btv279

5. Mc Grath-Lone L, Jay MA, Blackburn R, Gordon E, Zylbersztejn A, Wijlaars L, Gilbert R. What makes administrative data "research-ready"? A systematic review and thematic analysis of published literature. International Journal of Population Data Science. 2022;7(1) https://doi.org/10.23889/ijpds.v6i1.1718

6. UK Health Data Research Alliance, & NHSX. Building Trusted Research Environments - Principles and Best Practices; Towards TRE ecosystems 2021;(1.0). Zenodo. https://doi.org/10.5281/zenodo.5767586

7. Department of Health and Social Care. Secure data environment for NHS health and social care data – policy guidelines. 2022. Available here: https://www.gov.uk/government/publications/secure-data-environment-policy-guidelines/secure-data-environment-for-nhs-health-and-social-care-data-policy-guidelines [Accessed 04.05.2022].

8. ONS Secure Research Service https://www.ons.gov.uk/aboutus/whatwedo/statistics/requestingstatistics/secureresearchservice/aboutthesecureresearchservice [Accessed 12.05.2023].

9. Taylor JA, Crowe S, Espuny Pujol F, *et al*. The road to hell is paved with good intentions: the experience of applying for national data for linkage and suggestions for improvement. *BMJ Open* 2021;**11:**e047575. https://doi.org/10.1136/bmjopen-2020-047575 https://bmjopen.bmj.com/content/11/8/e047575.

10. The Royal Society. From privacy to partnership. The role of privacy enhancing technologies in data governance and collaborative analysis. 2023;Jan. Available here: https://royalsociety.org/-/media/policy/projects/privacy-enhancing-technologies/From-Privacy-to-Partnership.pdf?la=en-GB&hash=4769FEB5C984089FAB52FE7E22F379D6 [Accessed 04.05.2023].

11. Aitken M, McAteer G, Davidson S, Frostick C, Cunningham-Burley S. Public Preferences regarding Data Linkage for Health Research: A Discrete Choice Experiment, International Journal of Population Data Science, 2018 3(1). https://doi.org/10.23889/ijpds.v3i1.42

12. Brunswick Agreements. Available from Association of Research Managers and Administrators (ARMA) https://arma.ac.uk/updated-brunswick-agreements/ [Accessed 04.05.2023].

13. Contracts and study agreements. Available from IRAS https://www.myresearchproject.org.uk/help/hlptemplatesfor.aspx#Contracts-Agreements [Accessed 04.05.2023].

14. Model Agreements. UK Clinical Research Collaboration. https://www.ukcrc.org/regulation-governance/model-agreements/ [Accessed 04.05.2023].

15. Wollard M, Lichtwardt B, Bishop EL, Müller D. d5.9 Framework and contract for international data use agreements on remote access to confidential data (v1.0). 2021. Zenodo. https://zenodo.org/record/4534286 [Accessed 21.07.2023].

16. UK Health Data Research Alliance https://ukhealthdata.org/ [Accessed 04.05.2023].

17. Desai T, Ritchie F, Welpton R. Five Safes: designing data access for research. 2016. Available here: https://www2.uwe.ac.uk/faculties/BBS/Documents/1601.pdf.

18. The UK Health Data Research Alliance Principles for Participation. Available here: https://ukhealthdata.org/wp-content/uploads/2023/03/Alliance-principles-for-participation_Mar2023-1.pdf [Accessed 12.05.2023].

19. Data Access and Governance. Pan UK Data Governance Steering Group.https://ukhealthdata.org/projects/data-access-and-governance/ [Accessed 04.05.2023].

20. Waind, E. Trust, security and public interest: Striking the balance: A review of previous literature on public attitudes towards the sharing, linking and use of administrative data for research, International Journal of Population Data Science, 2020.5(3). https://doi.org/10.23889/ijpds.v5i3.1368

21. Stockdale J, Cassell J, Ford E. "Giving something back": A systematic review and ethical enquiry into public views on the use of patient data for research in the United Kingdom and the Republic of Ireland [version 2; peer review: 2 approved]. Wellcome Open Res 2019; 3:6 [https://doi.org/10.12688/wellcomeopenres.13531.2]

22. New research finds data trust deficit with lessons for policymakers. Ipsos. 2014. Available here: https://www.ipsos.com/en-uk/new-research-finds-data-trust-deficit-lessons-policymakers.

23. MORI, Ipsos: The One-Way Mirror: Public attitudes to commercial access to health data. Wellcome Trust. Journal contribution. 2017. https://doi.org/10.6084/m9.figshare.5616448.v1

24. National Data Guardian. What do we mean by public benefit? Evaluating public benefit when health and adult social care data is used for purposes beyond individual care. 2022;Dec. Available here: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1124013/NDG_public_benefit_guidance_v1.0_-_14.12.22.pdf [Accessed 04.05.2023].

25. Principles for analysis and use of health data by ABPI members. Available here: https://www.abpi.org.uk/media/fwclh3nu/data-governance-principles-abpi-members.pdf [Accessed 12.05.23].

26. Graham M. Data for sale: trust, confidence and sharing health data with commercial companies. J Med Ethics. 2021; Jul 30:medethics-2021-107464. https://doi.org/10.1136/medethics-2021-107464

27. Building trust in data access through public involvement in governance. Survey findings and recommendations from HDR UK's Public Advisory Board. 2021; June. Available here: https://www.hdruk.ac.uk/wp-content/uploads/2021/07/280621-PAB-Data-Access-procedures-paper-Building-trust-in-data-access-through-public-involvement-in-governance.pdf [Accessed 04.05.2023].

28. Our Advisory Groups. HDR UK. https://www.hdruk.ac.uk/about-us/who-we-are/our-advisory-groups/ [Accessed 04.05.2023].

29. UK Standards for Public Involvement. https://sites.google.com/nihr.ac.uk/pi-standards/standards?authuser=0 [Accessed 04.05.2023]

30. NIHR public contributor payment policy. https://www.nihr.ac.uk/documents/nihr-public-contributor-payment-policy/31626#:...:text=confirmed%20in%20writing.-,Expenses,public%20contributor%20and%20NIHR%20staff [Accessed 04.05.2023].

31. L. Arbuckle, F. Ritchie. The Five Safes of Risk-Based Anonymization. IEEE Security & Privacy, 2019 Sept 17(5):84-89. https://doi.org/10.1109/MSEC.2019.2929282

32. van Staa T, Goldacre B, Buchan I, Smeeth L. Big health data: the need to earn public trust BMJ 2016; 354 :i3636. https://doi.org/10.1136/bmj.i3636

33. Mello MM, et al. Waiting for data: Barriers to executing data use agreements. Science 2020. 367, 150–152 (2020). [https://www.science.org/doi/abs/10.1126/science.aaz7028]

34. Saulnier KM, Bujold D, Dyke SOM, *et al.* Benefits and barriers in the design of harmonized access agreements for international data sharing. Sci Data 2019, **6,** 297. https://doi.org/10.1038/s41597-019-0310-4 [https://www.nature.com/articles/s41597-019-0310-4]

35. Al-Rodhan N. The Social Contract 2.0: Big Data and the Need to Guarantee Privacy and Civil Liberties. 2018;Jun. Available here: https://www.bbvaopenmind.com/en/humanities/beliefs/the-social-contract-2-0-big-data-and-the-need-to-guarantee-privacy-and-civil-liberties/.

36. Carter P, Laurie GT, Dixon-Woods M The social licence for research: why care.data ran into troubleJournal of Medical Ethics 2015;41:404-409 https://doi.org/10.1136/medethics-2014-102374

37. NHS Digital. About the GDPDR programme. Available here: https://digital.nhs.uk/data-and-information/data-collections-and-data-sets/data-collections/general-practice-data-for-planning-and-research/about-the-gpdpr-programme [Accessed 04.05.2023].

38. Jones K, Heys S, Thompson R., Cross L. and Ford D. Public Involvement & Engagement in the work of a data safe haven: a case study of the SAIL Databank. International Journal of Population Data Science. 2020 5(3). https://doi.org/10.23889/ijpds.v5i3.1371. Available here: https://ijpds.org/article/view/1371/2815.

39. Office for National Statistics. Become an accredited researcher. https://www.ons.gov.uk/aboutus/whatwedo/statistics/requestingstatistics/secureresearchservice/becomeanaccreditedresearcher [Accessed 04.05.2023].

40. Karrar N., Khan SK, Manohar S, Quattroni P, Seymour D, Varma S, & The UK Health Data Research Alliance. Improving transparency in the use of health data for research: Recommendations for a data use register standard. 2022. Zenodo. https://doi.org/10.5281/zenodo.5902743

41. TRE Data access agreement template. https://zenodo.org/record/8256235.

Abbreviations

AR:	Approved Researcher
CPRD:	Clinical Practice Research Datalink
DAA:	Data Access Agreement
DDA:	Data Depositing Agreement
DPIA:	Data Protection Impact Assessment
EU:	European Union
GPDPR:	General Practice Data for Planning and Research Direction
HDR:	UK Health Data Research UK
HO:	Host Organisation
HRA:	NHS Health Research Authority
HSC:	Health and Social Care (Northern Ireland)
IP:	Intellectual Property
NHS:	National Health Service
ONS:	Office for National Statistics
PETs:	Privacy Enhancing Technologies
PIE:	Public Involvement and Engagement
SAIL:	Secure Anonymised Information Linkage Databank
TRE:	Trusted Research Environment
UK:	GDPR UK General Data Protection Regulation
UK:	LLC UK Longitudinal Linkage Collaboration
UO:	User Organisation
VPN:	Virtual Private Network

Supplementary appendix 1

TRE Legal Toolkit
Patient and Public Involvement

The aim:

There are various contracts involved in health data research. These come in different forms and vary between different data custodians, resulting in a complicated process that can be confusing for researchers. The differences between contracts can also cause caution amongst contracts teams within organisations, resulting in delays and effectively acting as blocker to research projects starting. This is an area that could really benefit from streamlining and collaborative working across the four nations. With the move towards the use of Trusted Research Environments (TREs), this offers a chance to introduce best practice at a relatively early stage. TREs by nature provide data protection and security assurances that were not possible with traditional data sharing, and this should be reflected in the associated contracts. While the TRE may contain pseudonymised individual-level data, only anonymised aggregate data may be downloaded. The processing will occur within the highly secure TRE and any outputs will be subject to checks and approval prior to release from there.

The TRE Legal Toolkit group are developing the following templates:

1. A data access agreement (DAA), a contract between those that hold the data and those that want to access the data
2. A data depositing agreement (DDA), a contract between those that will have their data within the TRE and those that are responsible for the TRE
3. A data protection impact assessment (DPIA), a form that is completed to identify and mitigate potential data protection risks to an acceptable level before processing data that identifies individuals (personal data)

Guidance and a glossary will be developed alongside these documents to aid ease of use. We will also produce a tool to help with defining roles and responsibilities against General Data Protection Regulation (GDPR) for all parties involved.

The TRE Legal Toolkit Action Force includes various stakeholders, such as experts in health data research contracting across the four nations and patients and public representatives. This diverse collaboration will ensure our work is clear, accessible, and responsive to current priorities.

The overall aim is to produce templates that are trusted, fit for purpose and that ultimately result in the speeding up of contracting to facilitate research that improves and saves lives.

For further info see here: Data Access and Governance UKHDRA (ukhealthdata.org)

STEP 1: Developing a data access agreement (DAA) template

We have reviewed DAAs currently in place for established TREs as a benchmarking exercise to define key shared principles. We will use these principles as building blocks for our DAA. The principles have been broken down into sections in the numbered list below. We have included questions that we would like you to answer but would also appreciate your reflections on any of the points listed. All comments, amendments and suggestions are welcomed, as we believe it is essential to include patient and public representatives' contributions.

1. Contractual Parties:
    - The DAA should be between the 'Host Organisation' (an organisation which is accredited to host and control a TRE) and 'User Organisation' (an organisation which is responsible for individuals using a TRE service),
    - with 'Approved Researchers' (people that access the data within a TRE, that have a contract with a verified User Organisation) agreeing that they have read and understood terms and conditions in the DAA around security and user requirements. For example, that they must only use data for the approved purposes.
    - The User Organisation will sign a warranty that makes them responsible for the approved researcher meeting the requirements to access the TRE. The User Organisation is liable for the compliance of the researcher.

[We have found that some DAAs follow this arrangement whereas some ask a researcher to sign the full DAA and accept responsibility, where there may be little possibility of legal action against an individual unless there is criminal intent]

Q: Do you agree with this arrangement and the distribution of responsibility?

1. **Access:**
    - The User Organisation shall only permit access to and use of the Data in the TRE by Approved Researcher(s) for the Approved Project and not by any other persons, and not for any other purpose.
    - Approved Researcher should be affiliated with the User Organisation. [we still need to define affiliation – this could include those with employment contracts, those with honorary contracts and students]
    - The User Organisation shall and shall procure that the Approved Researcher(s) keep confidential (i) the Data, and (ii) any access credentials to the Data.
    - Restrictions on access to a defined 'safe setting': approved area of User Organisation covered by the NHS Data Security and Protection Toolkit (or equivalent)
    - OR Approved researchers can access the TRE using a Virtual Private Network (VPN) from any location in the UK
    - User Organisation/ Researcher access restrictions if subject to an investigation/breach.

- Requirement that Researchers have appropriate accreditation (This may vary but we are seeing the ONS accreditation frequently Become an accredited researcher - Office for National Statistics (ons.gov.uk)

Q. Have we considered everything when it comes to accessing the TRE?

Q. Would you expect an approved researcher to have a full employment contract with the organisation or would an honorary contract be acceptable?

Q. Would you be happy with students accessing the TRE if their supervisor takes responsibility for them?

3. **Outputs**

- Neither data nor any research output (the analyses and any resulting write-up) must be used for any purpose contrary to the Applicable Laws
- Data must not be downloaded, extracted, transmitted, transferred, removed, copied or published from the TRE.

4. **Use of data**

- Must not attempt to identify individuals from the data or contact any research participant
- Data shall only be used for purposes defined in the approval. Further research requires new approval. Non-compliance results in termination of access.

Q. What are your views on permitting access to the datasets for exploratory research? For example, permitting approved researchers access to some or all datasets to discover where we should be directing research by looking at trends in the data. Research projects are normally subject to data minimisation, where data fields are restricted to those that are absolutely necessary to answer the research question. Here, researchers would still need approval and the defined purpose would be for exploratory research, but the data would not be limited as there would be no set project.

5. **Intellectual property**

- There will be no transfer of intellectual property (IP) ownership (ownership of the data). IP shall remain the property of the data owners for each dataset.
- The TRE host organisation owns IP for any derived data (data that has been created by combining or processing data from one or more of the original datasets. The data fields are dependent on the original data for analysis but become new data in their own right).

6. **Liability for data accuracy and availability**

- The Host organisation makes no warranty, express or implied as to accuracy or quality of the data; and
- excludes all liability for actions, claims, proceedings, demands, losses, costs, awards, damages, and payments made by the User Organisation that may arise from their use of the data or unavailability to the data for whatever reason.

Q. Do you think there are clear reasons for all of the above? Anything else we should consider?

7. **Compliance with Data Protection Legislation and Liability**

- Each party shall comply with their respective obligations under Data Protection Laws.
- The User Organisation must inform the Host Organisation without delay, and in any event within 48 hours of becoming aware of:

– any unauthorised access, disclosure, loss, damage or alteration of the Data

– any element within the data that might permit the identification of a research participant

– any complaints from an individual or supervisory authority in relation to the data; and

– any request from a research participant to exercise their rights in respect of the data.

- The User Organisation (+ Approved Researcher) agree to preserve confidentiality of information.
- The User Organisation (+ Approved Researcher) agree to application of GDPR to data.
- Must access data in controlled environment and protect from onlookers.

Q. Do you think everything has been covered in terms of protecting data?

8. **Commercial use**

- Data must be used for public benefit. Unauthorised commercial use will result in termination
- Commercial use may be permitted if stated in the approvals granted prior to the project starting

Q. Please comment on your feelings around commercial use as part of approved projects?

9. **Onward linkage**

- There should be no attempt to link or combine the data with other information or data (including any information relating to an identified or identifiable natural person) available to the User Organisation.
- With express permission as part of the approval process, data may be linked to other datasets (the other datasets would need to be ingested in to the TRE)

Q. To what extent do you think it is important that data is not linked to other datasets? (Bearing in mind that only anonymised data will ever leave the TRE)

10. **Open sharing of analysis, code, and derived data**

    - Ownership of Syntax (set of rules for analysis) and methodology remains with the User Organisation, but they shall grant a licence for use for other research and non-commercial purposes.

Q. Many feel it is important that the work that underpins research is shared so that we avoid wasting time in duplicating processes, and so that errors can be reduced along the way. Do you support the contractual obligation to share?

11. **Term**

    - Access will be terminated at the end of the term defined in the approval. Data destruction is not applicable as no personal data can be taken out of the TRE.

Annex

There will also be an Annex to the DAA which will contain Terms and Conditions that are customisable to the individual TRE

These may include further detail on:

- The data that is available within the TRE – this would be a list of names of data fields and not actual data e.g. 'date of operation', 'length of stay', 'diagnosis code'. This would never contain actual patient data. There may also be detail of restrictions to reduce the likelihood of identification e.g. Dates may only be available as month-year, only the first half of a postcode may be available, or you may only get access to time to death from an operation rather than date of death.
- Any further conditions specified for possible linkage to other datasets
- Restrictions on international access
- The protocol that will be followed in order to request output of anonymous data from the TRE
- Further detail on accreditation or training requirements for researchers before they can access the TRE
- The process for recovering costs associated with researcher access from their organisation
- How the host organisation of the TRE will go about monitoring who is accessing the TRE and how they will record and act upon this
- Acknowledgement/ copyright statements to include in any publications
- Further details if linkage of datasets occurs via a third party
- Length of term of access and the process for requesting extensions

Q. Do you support these being customisable terms and conditions?