

Trusted Research in the Public Eye: A toolkit for public engagement for trusted research environments (TREs).

1 Background

The UK has a leading position in health and clinical research due to its heritage of innovation and novelty. This has been underpinned by central government funding and support from wider society as the benefits of scientific and technological advances have been embraced and, in the main, well received. There have been some clear examples where things went very wrong from the thalidomide tragedy in the 1960s to genetically modified food via the “Mad Cow” disease crisis in terms of the application of scientific advances, the perception or acceptance of technological developments, or the public messaging. Substantial improvements have been made.

Scientific research is well regarded and trusted by the public in the UK. A recent government report on “Public attitudes to science (PAS)”¹ found that trust in science and its regulation is increasing. Trust is earned through transparency and trustworthiness where leaders or experts are visible and scrutable by government, the public and other third parties. Much scientific research in the UK is publicly funded, meaning that the public need to be considered and included as a partner at every stage of the scientific process: project development, funding review, hands-on research and dissemination of findings.

1.1 Health Data Research

Public inclusion is particularly true for Health Data Research where the core source materials are collected from all of us. Our patient data are collected and stored as Electronic Health Records (EHRs) which can then be used to inform future research and advance patient care and treatment for others. The NHS in the four nations of the UK has a fundamental role in capturing the data, making it available and informing the public and patients regarding its use. The data that the NHS collects and manages is an exceptional and global resource which has substantial public benefit if used appropriately for research.

Health Data Research UK (HDR UK) is the UK's national institute for health data science. It is funded by UK Research and Innovation, the Department of Health and Social Care in England and equivalents in Northern Ireland, Wales and Scotland, and leading medical research charities. Its aim is to bring together the UK health data resources for innovation and to ultimately improve people's lives.

1.2 Trusted Research Environments (TREs)

Patient data contains sensitive data which must be treated carefully and are regulated under the provisions of the General Data Protection Regulation (GDPR) and the Data Protection Act (DPA)², overseen by the Information Commissioner's Office (ICO). Failing to properly manage the data or allowing leaks of the data are potentially damaging to the individuals involved and can result in sanctions against the data controller.

In order to make EHR data useable safely and securely, several processes are put in place. Firstly, use of data must pass ethical and governance requirements to make sure the data is used appropriately and proportionately to the research needs and no risk of unnecessary release of data is requested. Secondly, once approved the data are

¹ <https://www.gov.uk/government/publications/public-attitudes-to-science-2019>

² <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/>

pseudonymised and reduced to only the necessary information. For example, names are removed and addresses may be reduced down to only the first part of a UK postcode. Only data relevant for the research are included so if, for example, a study is on heart failure then X-rays of the legs or arms will not be made available to the researchers.

Once the appropriate and approved data are prepared, they are provided to researchers or analysts within a secure and restricted computing infrastructure called Trusted Research Environments (TREs). TREs are managed facilities for enabling research on health data where no access to public resources like the internet nor can the researcher copy data into or out of the environment. In Scotland, TREs – also known as Safe Havens – need to comply with principles and standards of the Safe Haven Charter³. HDR UK have published a useful guide on TREs⁴.

What is a TRE?
A TRE is a **Trusted Research Environment**. Also known as 'Data Safe Havens', TREs are highly secure computing environments that provide remote access to health data for approved researchers to use in research that can save and improve lives.

Why are they important?

- TREs make research safer. Making data available through a TRE means that people can be **confident** that their personal health data is accessed **securely** and their **privacy protected**.
- TREs help make **research efficient, collaborative** and **cost effective**, providing rich data that enables **deep insights** which will go on to improve healthcare and **save lives**.
- TREs provide approved researchers with a **single location** to access valuable datasets. The data and analytical tools are all in **one place**, a bit like a **secure reference library**.

How is my data safeguarded?

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

- Safe People**
Only trained and specifically accredited researchers can access the data
- Safe Projects**
Data is only used for ethical, approved research with the potential for clear public benefit
- Safe Settings**
Access to data is only possible using secure technology systems – the data never leaves the TRE
- Safe Data**
Researchers only use data that have been de-identified to protect privacy
- Safe Outputs**
All research outputs are checked to ensure they cannot be used to identify subjects

Learn more about TREs and discover examples of how TREs are being used to enable life-saving health research. [Learn more about TREs](#)

In the recently published and government commissioned Goldacre Review⁵, "Better, broader, safer: using health data for research and analysis", thirty recommendations were made on how to improve the health data research landscape in the UK. Many of them involved the increased and only use of TREs in accessing, disseminating and sharing patient data. The review also highlighted the importance of trust, transparency and public involvement.

2 Public Involvement

Public opinion of science, either positive or negative, can hugely influence government and company policy. It can change the direction of policy temporarily or indefinitely, and it can change how a policy is presented. Areas around genetic modified crops, animal testing and the environment have historically been contentious topics for scientists and policy makers to deal with. More recently the use of people's data and, relatedly, the application of "Artificial Intelligence" are at the forefront of the public's and regulators' minds.

³<https://www.gov.scot/publications/charter-safe-havens-scotland-handling-unconsented-data-national-health-service-patient-records-support-research-statistics/>

⁴ <https://www.hdr.uk/ac.uk/access-to-health-data/trusted-research-environments/>

⁵ <https://www.gov.uk/government/publications/better-broader-safer-using-health-data-for-research-and-analysis>

In the PAS report, it was found that the public are feeling more involved than in the past, but more was still needed to include the views of ordinary people into science. Overall people want to be involved and not just informed in the decision-making process about developments in science. The preferred routes for involvement are via survey, consultations or citizen science activities.

2.1 DARE

Data Analytics and Research Environments (DARE) is a 5-year programme⁶ funded by UK Research and Innovation (UKRI) which aims to advance the use of health and social data for the public benefit. This is to be achieved through the development of novel techniques and infrastructure to preserve and manage the safety of personal data whilst enabling more of its use in research. DARE has a strong programme of Public and Patient Involvement and Engagement to include and learn from ordinary individuals, as their voice is crucial in the development of the most beneficial science whilst taking on people's ideas and concerns. In doing this we avoid the situation where science or health policy is imposed on the public with poor messaging and little information such as was seen with the care.data initiative and gain the public's trust⁷.

As part of the DARE strategy, the Trusted Research Environment and Enclave for Hosting Open Original Science Exploration ([TREEHOOSE](#)) project aimed to involve the public to develop a resource to aid data controllers and TRE managers to better communicate what and how patient data are used in clinical research.

2.2 Workshop inclusion and Content

An online workshop was run on 14th May 2022 to discuss the use of health data with TREs, explain the current methods used and then listen to the public participants to see what their current understanding was, and learn how to improve information sharing. The workshop was run with a short initial presentation ([slides](#)) of the current process and then a series of open questions for comment and discussion. Participation recruitment was from across the UK with members coming from England, Northern Ireland and Scotland.

The bulk of this report is based upon the findings of this workshop and the resource developed to address the comments raised.

A second, follow-up, workshop was run on 23rd June 2022 to verify whether this report accurately and fairly presented the workshop and to receive feedback on the proposed resource.

Both workshops were led by public members of the TREEHOOSE team. The TREEHOOSE team used multiple tools to support the engagement of the PPIE group including presentation slides, interactive Menti content such as word clouds, and the online game [Up Goer Five Text Editor](#).

2.2.1 Topics Covered

Six questions were developed and divided into two parts: 1) to probe current understanding and knowledge of the use of health data in research, 2) to identify how to improve the sharing of information regarding health data research in the UK. The use of "Cloud" within health data research was also queried as this is a core focus of the TREEHOOSE project and because the use of personal data by commercial entities – often in the cloud – is an active topic of concern by the public and policymakers.

⁶ <https://dareuk.org.uk/>

⁷ <https://doi.org/10.1136/bmj.i3907>

The questions in the first workshop were:

1. How do you feel about your anonymous health data (GP, hospital records etc) being shared for research and planning purposes?
2. Where would you go to get trusted information or advice on how your health record data is managed and stored?
3. If this anonymous health data was put in the Amazon cloud (AWS) for a time-limited research project, what concerns would you have?
4. What would we need to do to reassure you that your health record is secure and used appropriately?
5. How could we present the data security information to you better and where would you expect to see it?
6. What other questions regarding the storage of health data either on or off the cloud do you have?

2.2.2 Commentary

Participants had no concerns with their EHR data being used for health research or other public health benefit reasons provided the data was anonymised or pseudonymised. It was assumed that the data was already being used for those purposes. This provided a positive basis for the use of health data in research. The rest of the discussion, although led through specific questions, resulted in several themes occurring throughout which will be discussed individually.

Information Sources

In terms of getting information regarding how health data is used in research, the picture was much more confused. It is apparent with the four UK nations and the disparate NHS boards or trusts being somewhat independent, that different systems are in place for use by patients. For example, “Patient knows best” is not something that was known beyond one or two people in the group. Other sources suggested were “Health Improvement Scotland”, but the overall experience of having this kind of information was limited. Suggestions where the data should be presented varied from gov.uk – the UK government website – to healthcare settings like GP websites or surgeries themselves or NHS websites or pharmacies.

Asking people in clinical roles such as GPs or hospital specialists or even personal contacts considered to have specialist knowledge in the area was also a commonly mentioned source of information. There was a consensus that information on access to health data, for research or otherwise, should be improved.

Given the variety of responses, it seems clear that there is no clear definitive information point for health data, at least not in the minds of the public. It was also clear that no single source would be sufficient as being able to ask people, or refer to pamphlets/posters and search online were mentioned. To reach people a variety of resources are needed. There is an opportunity here to establish an information resource to fill the void. A comment was made that there are too many NHS websites. HDR UK would be a natural leader to develop a reliable and definitive presence such as the National Cyber Security Centre⁸ does for personal, business and national resources to protect from cyber-attacks.

Knowledge gathering/dissemination

Many of the processes and concepts mentioned in the workshop were unfamiliar to participants or were different to certain assumptions they held. Using patient EHR data was almost an expectation by participants, and it seemed to come as a surprise that access was not automatic. A desire for more information on what data is used and how

⁸ www.ncsc.gov.uk

was expressed more than once. The explanation of how data are pseudonymised, managed and access secured was well received and as such where the data are stored raised few additional concerns beyond general hacking or security of servers or IT systems.

It does appear that the whole health data research space is unfamiliar to the public and poorly explained. A better understanding of the gaps in knowledge is an important aspect.

Patients access to their own data

A request was made to be able to have access to patient's own NHS records and perhaps have a more nuanced level of control on what data is available or not for access. An opt-out system for patients to temporarily or selectively remove their clinical data for research was mentioned. Currently, the system is very slow and manual. It was noted that the systems in place across the four nations are at very different levels of maturity with Northern Ireland having very limited systems for patient access to data whereas patients within certain health boards in England can view their GP practice results and prescriptions information via the [Patient Knows Best](#) system.

This topic is not really within the scope of this study but does raise wider questions regarding patient engagement with their own data and provisions within GDPR and DPA frameworks. EHR data is very fragmented and stored in inconsistent formats or systems in the UK. An ability for all patients to access their EHR data would require a substantial, nationwide data rationalisation or standardisation process to occur.

Security

The question of where data is kept and/or made available for research mattered in terms of who had control and how were they secured. There are some concerns that the cloud could be less secure or prone to hacking or there was a risk of "giving the data to AWS". It was also noted that NHS servers are not infallible in these regards. A better explanation of what "the cloud" means would be beneficial.

Better knowledge and information around several aspects of health data were often mentioned as the questions posed raised concepts or processes which are not familiar or widely known. However, more knowledge, as mentioned in the other sections, was a desirable outcome.

Commercialisation

The privatisation of the NHS was also raised as a concern. What implications regarding ownership of patient data would that have? There were strong opinions that people would not be happy to have their data commercialised even if they were paid for it directly. Direct payments could target vulnerable people or encourage or coerce people to reveal information that they might later regret. The consensus, which reiterated the baseline position of the participants, was that if EHR data could be used to benefit others then it should be and that was enough incentive, or benefit, for them. There was also a concern that commercialisation would make research more expensive.

3 Recommendations

3.1 Top Level

Using the responses and feedback from this cohort of workshop participants alone, it is possible to draw some broad conclusions. Patients and the public are interested in health data research, however, the level of information available to them is lacking and/or poorly signposted. More efforts should be made to disseminate information via different media (print and online) and different venues (clinical or non-clinical). The use of health data for research and other public benefit uses raises few concerns provided the data are managed securely, safely and pseudonymised or anonymised, and better access to their own data was requested. It appears that the four UK nations are advancing data access at very different speeds. Finally, being paid for access to data or

commercialisation of data drew a strong reaction and was not considered beneficial in any way. It was understood that access to health data had cost implications, and that research was not a free service.

A larger exercise with different participants with wider representation would be beneficial to identify additional points and to support the findings identified here.

Below is a list of recommended statements or wording for health data researchers to use to support them when engaging with the public or patients or other non-experts when presenting their work or the work of Trusted Research Environments.

3.2 Suggested Wording

Health Data Research:

Knowledge generation on personal and patient health data to improve wellbeing and benefit society.

Trusted Research Environment:

A safe and secure place which can be reliably used to access patient data for research.

Data governance:

Legal and regulatory standards required to comply with the data privacy act and general data protection regulation governing personal data. It covers traceable audit trails, secure management and, importantly, apply best practice to ensure any mistakes are fixed.

Electronic Health Records:

Electronic health records are the shareable personal medical information that is stored in a safe computer database. Its primary aim is to provide centralised access to medical professionals (e.g. GP) to the medical history of the individual.

Consented/unconsented data:

Consented/unconsented data is about the permission received from the individual on how their information is allowed to be used. Unconsented data usage is highly regulated. It can be used for research that is for the public good or the common good.

Data Pseudonymisation:

Removing or changing identifying details within data so they can no longer be linked back to a specific individual, while still being useful for research. For example, removing names, replacing patient ID numbers with project-specific numbers, 'rounding' a date of birth to the start of that year.

Cloud computing:

A global and centralised, yet with an undefined location, computing environment. Information and data are safely stored and accessed from anywhere via the internet, often via a password.

Security of infrastructure:

Protecting computer systems from unwanted access, both legitimate users trying to do inappropriate things and other people trying to gain access they shouldn't have.

Encryption:

A way to obscure information or data so that it is not visible to others. A password is usually required to view encrypted data.

Research projects:

An investigation to explore, learn and understand. Usually performed at a University which creates new knowledge and explanations. Good research often includes patient and public involvement.

On-premise TRE:

A computing server system housed in a local building which can be safely accessed via a networked computer.

4 Acknowledgements

We thank all the participants in the workshops and the members of the TREEHOOSE project. This work was funded by UK Research & Innovation Grant Number MC_PC_21032 as part of Phase 1 of the DARE UK (Data and Analytics Research Environments UK) programme, delivered in partnership with HDR UK and ADRUK.

5 Authors

Dr Christian Cole – Senior Lecturer in Health Informatics and TREEHOOSE project lead. University of Dundee.

Jillian Beggs – PPIE co-lead.

Antony Chuter – PPIE co-lead.

Dr Chuang Gao – Health Data Researcher. University of Dundee.

Dr Simon Li – Senior Computing Infrastructure Engineer. University of Dundee.

Dr James Sutherland – Health Informatics Researcher. University of Dundee.