

Understanding Patient Data response to HDR-UK Green Paper on ‘Trusted Research Environments’

Understanding Patient Data (UPD) is pleased to respond to HDR-UK’s consultation and welcomes its aim to align researchers’ needs with steps to improve public confidence in health data use. We have provided a response on consultation questions related to public trust and transparency, together with additional commentary on the Green Paper’s broad approach.

The ‘Five Safes’ model provides a strong, clear structure for providing public assurance about the safeguards used to protect health data. However, it is equally important to acknowledge that developing innovative approaches to health data science means some calculated risks may be necessary to produce research that is likely (though not guaranteed) to have a public benefit. The Five Safes approach should be accompanied by an assessment of the intended positive public benefit, so that public communications and engagement are not solely about risk mitigation.

Response to Consultation Questions**1. Safe settings**

‘How is the use of public cloud best explained to data custodians, researchers and the public and not just security experts?’

Talking about data use

While there is widespread willingness amongst the public to allow use of patient data for research in the public interest, two general concerns often emerge: questions over a party’s competence in keeping data secure, and the motivation a party might have to use the data¹. Both of these aspects need to be clear for all stakeholders when explaining data use.

To overcome the challenge of explaining complex engineering and design in a way that engenders trust (p.12), we recommend further explaining what ‘the public cloud’ means in practice. This could set out how the cloud might differ from other types of data storage and the security measures in place to manage access. For example, that ‘public cloud’ does not mean openly accessible to the public. Language must be consistent and clear without assuming specialist or prior knowledge.

Visual representations can be a powerful and effective way of conveying complex information simply. However, the infographic on page 15 refers to ‘public cloud’ is difficult to understand. The current diagram assumes technical knowledge and covers several different aspects of the model in one visual representation (technical data flows, governance processes and data locations). We suggest designing a visual that is tailored to the specific audiences it is intended for, to convey the right information at an appropriate level.

Clear standards for assessing effectiveness, safety and accuracy in an ongoing way also need to be developed. We therefore welcome references to involving patient/public oversight groups in audit

¹ Stockdale J, Cassell J and 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. Available at: <https://doi.org/10.12688/wellcomeopenres.13531.2>

reports so that these can be responsive to people's questions and concerns, rather than solely focused on technical and security issues.

Finally, clinicians are crucial stakeholders who should be included in efforts to consult with patients and the public on the model being developed. Health data science relies on health professionals to collect and input quality data, to enable the right flows of data and to act as gate-keepers for public trust and support. Their perspectives, including their concerns, their insights from 'on the ground' data collection, and their potential to act as clinical champions for trustworthy data use, will be essential for engendering trust in the proposals.

Minimising risk

Our recent *Foundations of Fairness* report also underlined the importance of data security.² However, the HDR-UK approach should consider the need balance concerns about security with the need to undertake critical research in the public interest. We therefore recommend that a 'safe settings' approach include reference to the importance of *minimising* risk in its approach to security, not *eliminating* risk entirely. We know that the public expect data custodians to be honest about the risks and benefits of health data use and are able to consider complex trade-offs.

'Safe computing'

'*Safe computing – an extension of Safe setting*' (page 12) refers to private sector computing infrastructure:

"In order to build public trust, use of private sector computing infrastructure to provide a safe setting must be done in such a way that none of the hardware and software layers outsourced make it possible for the third-party provider to access any of the individual health data."

Our research suggests that the public do not necessarily distinguish between types of services (software provision, business analytics, computing infrastructure) provided by third parties when considering how health data is being used or accessed. They expect that, whatever the partnership, due regard is given to how it will bring equitable benefit across the NHS and the health system. As such, it will be important to ensure that all these partnerships are subject to rigorous public benefit assessments, with clarity over why certain third-party providers are being chosen. For example, they could be uniquely well-placed to provide technical expertise, scalable compute or some other necessary and specific resource.

Additionally, the business model of any commercial third party involved in setting up, managing or using this infrastructure will need to be clear. One of the first questions people frequently ask of these partnerships is "What's in it for them?" Many prominent tech companies profit from brokering and selling data: this should not be the commercial model for TRE partners if they are to be trusted with health data and this should be made explicit.

² Hopkins, H., Kinsella, S., van Mil, A. (2020). *Foundations of fairness: views on uses of NHS patients' data and NHS operational data* [pdf]. Available at: <https://understandingpatientdata.org.uk/sites/default/files/2020-03/Foundations%20of%20Fairness%20-%20Full%20Research%20Report.pdf>

2. Public trust

How can we engage patients and the public to demonstrate the benefits of health data research and build public trust around the use of trusted research environments for research and innovation at scale?

What we know about trust

We welcome HDR-UK's questions on this topic. Public understanding about health data use is generally low, especially when it comes to purposes beyond care.^{3 4} However, simply seeking to educate the public about health data is insufficient for engendering trust – and this is extremely difficult at a population scale. We suggest the following are important factors for HDR-UK's approach.

Informing people

Informing people about when, why, how and what data is used is still important for building trust, and can be complex. A TRE model has significant advantages over governance models for data that allow the data to flow to researchers, because several of the key concerns people will typically have about data can be readily and comprehensively addressed. For example, a TRE can guarantee that:

- data cannot be passed on to third parties and beyond the custodian's control
- all analyses and uses of data can be tracked and audited
- data is only used for permitted purposes
- statistical disclosure control of outputs can mitigate risks of re-identification through jigsaw linkage.

HDR-UK should look to develop information materials based on pre-empting these key public concerns about data use. These assurances can be provided by security and computational measures and are not contingent on data users, who may not be trusted by sceptical members of the public, abiding by contractual terms. The model of 'users must come to the data, the data doesn't go to them' is intuitive and understandable. It is therefore more straightforward to demonstrate a TRE is fulfilling conditions for trustworthiness.

Our *Foundations of Fairness* report found that people want consistency in the rules applied to data access and use. The TRE model lends itself well to creating and scaling a consistent set of rules and parameters for data use, even if the type of data and the purposes of use will vary across different TREs, data types and disciplines.

Meaningful transparency

An important component of trustworthiness is transparency. This means not only providing accessible, meaningful information (the 'what' you are doing) but also conveying it in a way that is easy to find and follow up (the 'how' of explaining). People should not need to understand the intricacies of HDR-UK's structure to navigate to the right information. Identifying the right touch points for people to

³ *Foundations of Fairness* (ref 2): 63% of people are unaware that the NHS gives third parties access to data.

⁴ Healthwatch England (2018) *How do people feel about their data being shared by the NHS?* [online]. Available at: <https://www.healthwatch.co.uk/news/2018-05-17/how-do-people-feel-about-their-data-being-shared-nhs>.

learn about data use at a time and context appropriate for them will be valuable – a design project we are currently running may be informative here⁵. As HDR-UK develops its approach to transparency from high-level principle to the practicalities of a TRE model, it should recognise that what matters to a patient might be different than a cyber security or data governance perspective on what transparency means⁶.

Making data relevant

Studies by medical research organisations demonstrate that when health data research is focused on specific conditions people are more likely to approve of health data use. In a 2018 study, Asthma UK found that 83% of people with asthma are comfortable with their confidential health data being shared for research to develop new asthma treatments.⁷ The Brain Tumour Charity conducted a survey of people living with brain tumours and found that 97% of respondents would be happy to share their data, not only to accelerate the development of new treatments and diagnostics, but also to help inform other patients in a similar position.⁸

Such studies demonstrate that when data use is connected to issues that resonate with people and feel meaningful for their lives, they tend to be more comfortable and positive about its potential. A patchwork of engagement efforts on specific topics that matter to different communities may be more successful than a nationally-led, single exercise to inform and engage people about data. A TRE model has the advantage of being able to see and pull together the wide range of research conducted on a specific topic or condition. This means a clear, compelling narrative can be developed about why data matters, rather than the benefits from disparate research efforts by different teams being conveyed in a piecemeal way.

Engaging people

A trustworthy environment must engage with a wide range of people about how it operates, on an ongoing basis, rather than only communicating information about decisions already made. Complex technical requirements do not prevent engagement with members of the public, especially for evaluating the benefits of health data research.

People feel a sense of ownership over health data and want to be involved in decisions about its use. In a recent national survey co-commissioned with NHS England we found that 74% of people believe the public should be involved in decisions about how NHS data is used.⁹ There was a clear interest

⁵ Annan-Callcott, G. (2020) 'How to talk about data without talking about data' 28 April [blog]. Available at: <https://understandingpatientdata.org.uk/news/how-talk-about-data-without-talking-about-data>

⁶ Additional information is available at: <https://understandingpatientdata.org.uk/what-does-transparency-mean>

⁷ West, B. and Cumella, A. (2018) *Data sharing and technology: Exploring the attitudes of people with asthma* [pdf]. Available at: <https://www.asthma.org.uk/8195030f/globalassets/get-involved/external-affairs-campaigns/publications/data-report/data-sharing-and-technology---exploring-the-attitudes-of-people-with-asthma.pdf>

⁸ Cheema, T. (2018). 'How can charities support a transformation of patient care through data?' 2 October [news story]. Available at: <https://www.thebraintumourcharity.org/media-centre/news/brian-news/transforming-patient-care-through-data/>

⁹ Foundations of Fairness (ref 2)

from participants in being involved in setting the direction and priorities of health data research rather than in individual case decisions.

We suggest that a ‘learning governance’ approach could be beneficial for the TRE model, as a way to broaden engagement while not overburdening governance processes. A citizen and patient panel can review examples of data projects already undertaken, assess whether the right questions were asked or what might have been missing from the scrutiny of safe people and projects, so that this learning can feed in to future governance and decisions.

Welcoming diverse views

Recent research from OneLondon consolidated other public attitudes research when it found that the public does not have a uniform view of health data issues. This research found that around a quarter of people seem to be unconcerned about the collection and use of personal information about them; another quarter are unwilling for it to be used for service enhancements; the remaining 50% of those asked would make trade-offs on a case by case basis.¹⁰ People hold a diversity of views and values about health data use that are often context-dependent and there is unlikely to be consensus in every case on what is and is not acceptable. This means it is even more important to ensure that the process by which decisions are made brings in a range of views and perspectives, so that the outcomes of that process are (and are perceived to be) legitimate and respectful of differing views. The trade-offs and balance of benefits and risks people are comfortable with may have shifted (perhaps temporarily) in light of covid-19, demonstrating a further reason why engagement on what ‘safe’ data use looks like needs to be sustained over time rather than being a one-off exercise.¹¹

Other comments

- The paper refers specifically to HDR-UK’s aim to support the use and analysis of ‘large-scale data’. While health data research naturally focuses on maximising the use of the UK’s rich health data sets for research purposes, there is also significant insight to be gained in understanding smaller-scale data sets in localised settings. The emphasis on the size and complexity of datasets should not come at the expense of recognising the contexts in which small-scale, highly granular data is valuable for research and decision-making. An emphasis on size and scale may also neglect what’s missing from data, what is inaccurate or incomplete: we are learning from covid-19 that data on ethnicity, for example, is often lacking in health data. Such limitations are vital to recognise if research on large scale health data is to serve a broad public interest.
- Section 3 ‘*How do we ensure this happens in a safe way that retains and enhances public trust?*’ refers to research conducted by Understanding Patient Data. It should be clarified that whilst there

¹⁰ OneLondon (2019). *Understanding public expectations on the use of health and care data* [pdf]. Available at: <https://understandingpatientdata.org.uk/sites/default/files/2019-07/Understanding%20public%20expectations%20of%20the%20use%20of%20health%20and%20care%20data.pdf>

¹¹ A new co-commissioned project with the National Data Guardian for Health and Care will go some way to addressing this, with the aim to develop best practice guidance for data custodians and researchers that will be relevant for HDR-UK. More information is available at <https://sciencewise.org.uk/2020/04/good-enough-assessing-public-benefit-in-data-driven-health-and-care-research-and-innovation/>

is a correlation between awareness of health data use for purposes beyond care and support for its use, there are several caveats to this. For example:

- the aim of health research should be easily understood and should intend to generate public benefit;
 - the public expect information about health data partnerships between NHS and third parties to be published;
 - the public expect to be involved in decisions about how patient data is used, especially when this involves third party use.¹²
- Page 7 rightly articulates the difficulty with referring to ‘data sharing’ when talking about data use to the public. ‘Sharing’ can have connotations of a free-for-all, or of altruism that may not be warranted. Referring to “data access” or “use of data” better conveys the activities of the TRE.
 - On page 18, the paper highlights the need for researchers’ and innovators’ requirements to be considered carefully when taking a TRE approach. We agree that getting the user experience right for this group is vital but suggest this should be construed more widely than the paper suggests: an important contributory factor in achieving a good user experience will be building a strong and trustworthy ecosystem of support and shared learning around the TRE model. This would also bring benefits to the ongoing sustainability and improvement of the TREs themselves. Such an ecosystem could include the development of open tools or example code to sit alongside a TRE, tutorials in different programming languages, reference datasets and code lists, tools and scripts for data transformation and cleaning, and validated models which can be forked, improved and used in different TREs. This kind of support for a wider community of users to flourish has been developed in a range of sectors beyond health and would be instructive to draw upon.¹³

We would welcome the opportunity to discuss this response further as HDR-UK develops the TRE model.

¹² Foundations of Fairness (ref 2)

¹³For example, the Luxembourg Income Study uses a remote execution system similar in nature to the TRE model proposed: <https://www.lisdatacenter.org/resources/other-databases/>