DARE UK Draft report, v1, July 2022

"Towards a coordinated national infrastructure for sensitive data research: A summary of findings to date from Phase 1 of the UK Research and Innovation DARE UK programme"

https://dareuk.org.uk/wp-content/uploads/2022/07/DARE_UK_DRAFT_Recommendations_July2022.pdf

26 July 2022

Comments from use MY data

The recommendations in the draft report are organised according to seven core themes:

1. demonstrating trustworthiness;
2. access and accreditation of researchers;
3. accreditation of research environments;
4. data and discovery;
5. core federation services;
6. capability and capacity; and
7. funding and incentives.

We are asked the same questions for each of the seven themes:

- To what extent do you feel the recommendations accurately reflect the current challenges in this area? (Please explain your answer, addressing any challenges within this area that you feel are missing from or not sufficiently addressed in the recommendations.)
- Are you aware of any initiatives not already mentioned in the report that are currently working on solving some of the issues covered in this area?
- Are there any recommendations you feel should be prioritised in this area? (Please explain your answer)

The text below shows the theme heading and summary recommendations for each of the seven themes extracted from the DARE report (shown in black text), together with some draft responses from use MY data, shown in blue text.

Demonstrating trustworthiness

1. Proactive transparency should be consistently practiced by all those handling and using sensitive data for research, particularly data collectors, data custodians and data guardians.

2. A public information campaign should be conducted to raise general awareness of how and why sensitive data is made accessible for research.

3. Data use registers should be published and maintained by the custodians of all types of sensitive data.

4. A culture shift is needed to recognise the crucial importance of public involvement and engagement in data research.

5. A central, independent coordinating function for public involvement and engagement in data research should be set up, either as a new entity or as an off shoot of a relevant existing body.

6. Where feasible, processes enabling access to sensitive data for research should be standardised and centralised across the UK.

7. Researchers hoping to access sensitive data should be stringently and ongoingly vetted and monitored to ensure public benefit is the principal motivation for accessing sensitive data.

Trustworthiness: Clarity of wording

The term "proactive transparency" as used seems to relate to proactive communications. That is an important element of transparency, but it is only one part. Transparency is a culture which needs to be embedded, rather than an action to be undertaken. Similarly, raising awareness is a positive step, but we would much rather see a culture of involvement being the norm, not just engagement.

Why does page 11 talk about "user personae"? This is not a term that will be understood, and usage of terms such as this is not a way to engender trust.

Trustworthiness: Involvement, not just engagement

Page 14 "emphasised the need for those handling and using sensitive data for research to actively reach out to the public with information about how and why their data is being used". We would argue that the evidence also supports the need for "involvement" of patients/public in the actual decision-making processes. This must become mandatory.

We agree that a culture shift in involvement is key. That must include involvement in decision making processes about data uses. Consideration should also be given to embedding the power of patient-veto.

In relation to engagement and involvement, we note the document identifies that "necessary resources should be dedicated" and included in research grant applications. Whilst we agree with the intention, there are fundamental differences in how you would engage or involve the public and patients. This needs to be recognised. Full public engagement is very different to patient engagement.

Establishing a central, independent coordinating function for public involvement and engagement in data research would be a significant task and would require a substantial infrastructure. To a degree, this was part of the role of Understanding Patient Data, though that organisation is now ceasing.

We note the point that lots of existing work exists in this area, and strongly encourage the intention to do some further scoping work on how this theme might best be developed. We hope that this scoping is done in an inclusive manner, bringing a range of views to the table.

Trustworthiness: Data release and usage registers should be mandated

The paper says that data use registers "could be mandated". We disagree – these **should** be mandated. Data release (and usage) registers are an essential element of increased transparency about the uses of health data. They are also an important starting point to be able to demonstrate the benefits that data can bring. There should also be consistency of style and content.

Trustworthiness: Ensuring both patients and industry are partners in conversations

We have always highlighted that where policy or decision-making groups exist, it is crucial that patients have a strong voice. But we would equally argue that industry should also be included in developing solutions for SDE/TRE usage. All discussions and solutions need to include the breadth of interests and views.

We think that potentially difficult areas would be far better debated out in the open than thinking it can only be discussed with selected interests.

Trustworthiness: Industry/commercial uses of TREs

It is very unclear how industry is going to be able to use a TRE. The question just seems to be side-stepped, even though if you could solve that in a way which the public would see as ok, everything else becomes much easier…!

In particular the tensions about the access rules by which industry users gain approvals for access, and then secondarily about the ways that industry users can operate in an environment where code and algorithms are shared. It is unclear whether industry users will have an exemption from code-sharing, and if so, how that will be undertaken in a transparent manner, and audited. This will be an essential element of operating in a trustworthy manner and being transparent.

We can understand concerns from industry about protecting their commercial interests, but this point isn't discussed in the draft, nor any solution to the problem.

Whilst there is greater concern in the public about industry uses of data, in principle they are no different from anyone other user, in that data uses must be for clear public benefit. Rigorous checking of the intended purpose would be essential, together with checking of any analysis code that they submit to run against the TRE. But that would apply for all users/uses.

In all cases it is essential that the TRE operator can vet the code with an understanding of the precise nature of the research that the company is carrying out. Clearly, this may concern the company (or academic user), but we consider it to be a fundamental requirement that all users of the TRE will have to accept. Without that, transparency and trust will be seriously compromised.

We do not think that restricting any vetting just to the outputs (such as for small number suppression) is transparent. Unless code is vetted, it would be impossible to verify any outputs or conclusions.

We would suggest that scrutiny of code could be undertaken under some form of contractual agreement or NDA to protect commercial assets and learning. But we think it would be wrong to exclude any code from scrutiny by the TRE operator.

Trustworthiness: The opportunity for synthetic data

We also highlight our Position Statement on the use of synthetic data which would allow code to be developed in a risk-free environment and vetted before it is run against real data.

Access and accreditation of researchers

1. Provide a unified user authentication capability to enable researchers to access services more easily across the entire sensitive data research ecosystem.

2. Provide a streamlined user accreditation framework to enable trustworthy researchers to access sensitive data for research in the public benefit in a timelier fashion.

3. Develop a standardised and streamlined – yet extensible – process to request access to sensitive data from TREs whilst maintaining appropriate levels of data privacy and security.

We agree with the description of the problems which researchers face in gaining access to data to undertake their research.

However, the report says that “Data Access Committee processes, policies, and standards are outside the scope of this report, as they are often subjective and subject to the combinatorial requirements of the data guardian, data custodian (e.g., TRE provider), and research project”. Surely this is a key part of the problem? Focusing just on ensuring consistent “baseline standards and support” only addresses part of the problem, with the risk that the required improvements will not be realised?

The focus on user authentication, though beneficial, is at risk of addressing the technical problems without addressing areas of culture and behaviour in data-access decisions.

The paper says that “international and industrial researchers would also need to be considered, as currently accredited researchers need a link to a UK institution”. But we could not find any details on what this means. Given the low public trust in commercial users, this needs to be much clearer.

We agree that “harmonising existing data access request processes into a single baseline procedure around the Five Safes that can be instituted and maintained by a centralised service would be a substantial step forward”. We also agree that cross-links should be used to drive the publication of data release registers.

Accreditation of research environments

1. Review and extend the existing standard, accreditation, and audit framework under the Digital Economy Act (DEA) to establish a nationally recognised trusted research environment (TRE) standard, accreditation, and audit framework.

“We note the comment that “heterogeneous and often in-compatible TREs are being created almost like a cottage industry in response to the need to manage secure access to sensitive data for research”. We highlighted similar concerns in our response to the recent Goldacre review.

In relation to the audit of a TRE, the paper notes “in the context of TREs, an independent authority and process should be established to effectively accredit and audit TREs against the relevant standard”. How would the patient/public voice be embedded in this audit process? And would the results (positive and negative) be published?

Data and discovery

1. Enhance the data lifecycle for cross-domain sensitive data research.

2. Explore implications of new data types, models for sharing and velocity of delivery.

3. Develop guidelines on privacy enhancing technologies (PETs).

4. Establish a UKRI-wide metadata standard working group.

5. Leverage existing digital object identifier (DOI) minting services to provide persistent identifiers for all UKRI council resources at UKRI-wide and council levels.

No specific comments

Core federation services

1. Develop reference architectures for trusted research environments (TREs).

2. Assemble an API (application programming interface) library to support core federation services.

3. Run a competitive call for driver projects to utilise the new infrastructure services and to validate that they are fit for purpose.

> No specific comments

Capability and capacity

1. Establish clear technical career pathways that can be adopted across the UKRI research domains

2. Improve recruitment pathways for technical roles.

3. Improve the availability of career development resources and training.

4. Use automation to reduce the dependency on shortage skills.

> We note that whereas the Goldacre Report recommended limited numbers of TREs, that the draft seems to be passively accepting the proliferation of TREs, instead relying on federation. Some clarity on this would be helpful to avoid confusion.
>
> The draft doesn't highlight enough the need for cooperation or standardisation with other Government bodies such as ONS, ESRC and others. It would be helpful to see more included about this principle.

Funding and incentives

1. Develop a new type of grant tailored for addressing the costs for maintaining cross-domain, national sensitive data research infrastructure.

2. Determine the funding requirements to establish the first phase of federated infrastructure for sensitive data research, with a focus on enabling federation across existing national data infrastructure and complimenting existing investments.

3. Investigate, test and prototype the operational model(s) for a federated ecosystem of national sensitive data research infrastructure. Critically, ensure federation lessons and insights from those outside of the sensitive data space are considered.

4. Investigate the cost implications for appropriate business continuity and disaster recovery requirements for a national, federated infrastructure for sensitive data research.

5. Investigate the scope and funding requirements for the integration of large-scale compute availability in a federated infrastructure for sensitive data research.

6. Building upon existing best practice, improve the availability of all data produced through publicly funded grants for reuse and investigate the funding requirements for provisioning such archival capability.

7. Raise awareness amongst data guardians regarding the legal framework around the secondary use of data for research.

8. Dedicate greater resource to incentivising data guardians to routinely make their data accessible for research in the public benefit.

> We agree that there is “a lack of sustained, dedicated funding allocation for public engagement and involvement activities”.