DARE UK – Early Thinking

Funding and Incentives

*Stakeholder workshop, Thursday 3 March 2022*

DARE UK aims to design and deliver a national data research infrastructure that is **joined-up**, **demonstrates trustworthiness** and supports **research at scale for public good**.

Draft recommendations for how to fund and incentivise a more cohesive sensitive data research infrastructure landscape

Based on:

- Our initial Landscape Review (October 2021) – involved interviews and workshops with stakeholders from across the sector

- Other conversations and discussions with stakeholders across the landscape
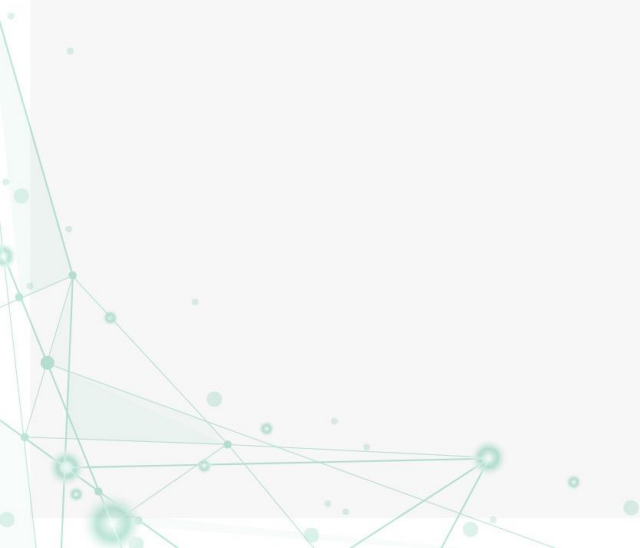
What is the context around funding and incentives for sensitive data digital research infrastructure (DRI)?

- **Sensitive data digital research infrastructure (DRI) today is generally funded through the traditional investigator-based grant model, supplemented with sporadic capital grants**

- **A federated ecosystem does not yet exist and needs to be established (and funded) in a coordinated way**

- **Access to large-scale compute for sensitive data research, both ad-hoc and over longer periods**

- **Business continuity, high availability and disaster recovery critical to protect national data research assets**

- **Incentives, service levels and research culture**

Recommendation 1: Develop a new type of grant tailored for addressing the costs for maintaining cross-domain, national, sensitive data DRI

- Tailored for the operational funding of national, sensitive data DRI and core federation services

- Appropriate time horizons to ensure stable, predictable infrastructure planning

- Address both capital (e.g. hardware refresh) and operational expenses

- Investigate and define minimum service levels for providers of national, sensitive data DRI and core federation services

- Transparent funding projections for national, sensitive data DRI and core federation services including key performance indicators thereof

Recommendation 1: Develop a new type of grant tailored for addressing the costs for maintaining cross-domain, national, sensitive data DRI

Broadly agreed that recommendation would serve to address long-term, sustainable funding challenges for research DRI and the current fragmentation of the landscape, if effectively and thoughtfully implemented

1. Less than 100% full economic recovery (FEC) does not apply to DRI facilities and should be carefully considered within the context of such a grant.

2. Minimum service levels are acceptable so long as funding allocations are aligned with the expectations placed upon DRI environments – there is no one service level for all types of DRI environments. Will be favourable for researchers in providing a degree of planning certainty and incentivising data asset use.

3. Appropriate funding timeframes are difficult to define for varying DRI environments and should be investigated comprehensively.

4. Grant needs to consider a broad set of complimentary funding requirements across a matrix of both functional (e.g. data management, inter-disciplinarity/cross domain, reuse of software assets etc.) and structural (e.g. human resource, hardware resources etc.) requirements.

5. There are advantages/disadvantages to funding flowing through 'host' organisations (i.e. HEIs) – these should be considered within grant development and, if applicable, dovetail with funding for DRIs at 'host' organisations.

6. Will need to consider what is working well today in terms of DRI environments and sustain that work rather than purely drive 'new' DRI environments.

Recommendation 2: Determine the funding requirements to establish the first phase of federated DRI for sensitive data research

- Based on the technical feasibility and prototypes – determine the initial investments needed to establish the first stage of a federated DRI for sensitive data research

- Investigate whether there are core, standard federation services required across the ecosystem and options for resourcing these

Recommendation 3: Investigate, test and prototype the operational model(s) for a federated ecosystem of national, sensitive data DRI

- Map the operating model(s) or framework(s) for a federated ecosystem

- Determine the options for cost recovery across a federated network of DRIs and if/how they differ across different sectors

- Determine if, based on the operating model(s), there are scenarios where cost recovery is not technically feasible and if so what other approaches could there be to manage federated DRI costs

Recommendation 2: Determine the funding requirements to establish the first phase of federated DRI for sensitive data research

General agreement that this is needed, however it was stressed that these requirements should be complimentary to existing national efforts and balance budget constraints with strategic value.

1. Should be mindful that a trade off exercise to balance strategic value of funded federation components versus budget constraints will be a likely reality – must consider existing, complimentary national efforts to ensure maximum impact.

2. Requirements gathering should proceed apace so that piloting and proof-of-concept can start as early as possible.

3. Should make clear whether federation will deliver cost efficiencies, additional capability or both in the longer term.

4. Possibility to investigate a 'playground' concept for exploratory research work within a federated DRI ecosystem.

Recommendation 3: Investigate, test and prototype the operational model(s) for a federated ecosystem of national, sensitive data DRI

Investigation and testing of different operational models supported, encouraged to be pragmatic and conscious of the risks of 'over centralisation'.

1. Cost recovery should be pragmatic and as simple as possible in order not to obstruct good research and should consider that there will be different models for different DRI environments.

2. Ensure wide consultation across councils – even those without sensitive data requirements will have federation insights. Investigate practical, demonstrable examples of where federation has worked to deliver efficiency or deliver additional value.

3. Operational model(s) should be appropriately balanced between a centralised versus federated approach.

4. As part of this there should be a comparative analysis of different cost recovery models (e.g. data access versus outcomes fees), including investigating opportunities for cost recovery through commercial partner charging model(s).

5. Facilitating and supporting integration across DRI environments is a valuable service that could incentivise cost recovery from institutes - it is critical that researchers know where to find these specialised skills, capabilities, and services.

DRAFT – not official policy

Recommendation 4: Investigate the cost implications for appropriate business continuity and disaster recovery requirements for a national federated DRI ecosystem

- Strategically important to ensure failsafe's in place for critical, national data assets and underlying DRI

- Investigate whether federation can serve to mitigate the costs for high availability and disaster recovery

- Determine funding requirements to pilot scenarios between DRI sites

Recommendation 5: Investigate the scope and funding requirements for the integration of large-scale compute availability in a federated sensitive data DRI ecosystem

- Investigate and define the use cases for large-scale compute requirements for sensitive data research (e.g. short term versus long term access requirements)

- Investigate and develop a cost model to identify when it is optimal to utilise commercial cloud providers versus provisioning through existing national large-scale compute capabilities

- Validate the feasibility and initial investment(s) needed to integrate large-scale compute capabilities into a federated ecosystem of sensitive data DRI

Recommendation 4: Investigate the cost implications for appropriate business continuity and disaster recovery requirements for a national federated DRI ecosystem

Broadly agreed that appropriate fail-safes should be in place to protect national, critical sensitive data research assets.

1. It should be demonstrated that this is more effectively addressed at a national level rather than institutionally - to some extent the inherent characteristic of a federated system addresses this (i.e. distributing the risk across multiple sites)

2. Investigate the offering from public cloud providers in this regard, both from a cost perspective and the risk of vendor 'lock-in' (data ingress is easy, data egress is cost prohibitive).

3. Should determine the relative priority level for this piece of work and define what is meant by "national, critical DRI".

4. Feasibility work should be conducted to explore this further to understand to what extent the current landscape addresses this.

5. Is there a separation between the data assets themselves and the DRI on which these assets are stored and analysed?

6. Should be built into the operational and cost model(s) as per recommendation 3.

Recommendation 5: Investigate the scope and funding requirements for the integration of large-scale compute availability in a federated sensitive data DRI ecosystem

Strong view that the priority use cases for large-scale compute on sensitive data should be identified as a foundation for scoping the requirements and subsequent funding needs.

1. Important to ensure there is appropriate support to facilitate the gathering of requirements between the researchers and the DRI facilities, often the interfacing of these two domains needs a better 'translator'.

2. Big challenge will be how to interface large-scale compute consistently across the DRI ecosystem for sensitive data – are there learnings from the public cloud approach?

3. Enabling the 'switching' between public cloud and national DRI environments will be a challenge and requires deeper investigation.

4. Important to map out use cases and requirements as a starting point, especially in the context of privacy requirements for sensitive data.

Recommendation 6: Raise awareness amongst data owners regarding the legal framework around the secondary use of data for research (2017 Digital Economy Act)

- Different interpretations of the legal basis can lead to wariness amongst data owners when it comes to data sharing and linkage for secondary purposes

- Clear, joined-up resources/advice should be developed outlining the legal basis for data sharing and linkage

Recommendation 7: Dedicate greater resource to incentivising data owners to routinely share and link data for research

- More dedicated funding to support the development of sustainable, research-ready data resources from across sectors

- Awareness raising regarding: the security processes in place to protect data from harm (particularly the Five Safes framework); evidence of public support for data research; and the policy benefits associated with data sharing and linkage

Recommendation 6: Raise awareness amongst data owners regarding the legal framework around the secondary use of data for research (2017 Digital Economy Act)

General agreement that awareness and expert support for data custodians around the legal frameworks is a critical challenge, however sentiment was that improvements to the existing legal frameworks is also required.

1. Critical to provide not only a consistent understanding of the legal framework for data research but also improve awareness of the DEA and its relevance to data owners so there is confidence in decision-making around making data available.

2. Should investigate to what extent legal support can address the challenge and whether this legal support capability is a gap for data owners. Is the DEA robust enough to manage the linkage of data across domains – should there be more work centrally to build on this?

3. Training and support for researchers in this area is critical to embed understanding.

Recommendation 7: Dedicate greater resource to incentivising data owners to routinely share and link data for research

Broadly agreed that there are fundamental resourcing challenges that inhibit routine sharing and linkage of data for research.

1. Generally the development of sustainable, research-ready data resources is not normally part of the data custodians core functions – thus the budget for resourcing simply isn't available. Base line funded resources needed to make data available – this is already appropriately resourced in some domains but not in others.

2. Does a central resource for dedicated data management skills and support, privacy professionals, and legal support make sense? Engagement needs differ according to data owners, important to build on the support from UKRI councils already being provided to grant holders.

3. Should consider how to incentivise the feedback loop from researchers to data custodians in order to improve and enhance data quality, for example embedding data owners in the shaping and prioritisation of research questions.

4. Specific funding calls for the secondary use of data do exist (e.g. ESRC Secondary Data Initiative) and could be used as a basis for similar broader funding calls, where sharing data for secondary use is a condition for receiving research grants.

5. Should think very broadly about the sharing of best practices – often communities without natural affiliations have valuable experience to share.

DRAFT – not official policy

To be discussed at a later stage…

- Formal recognition of the contribution DRI makes in research outputs

- Balancing innovation and collaboration within the research culture through intelligent incentives

- Leveraging international opportunities for funding

- Industry integration within a federated ecosystem and how this sector fits within the operating model(s)
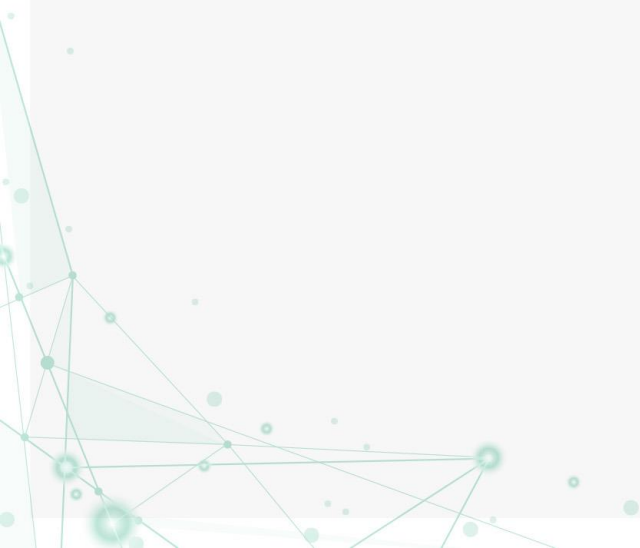
Thank you

Find out more about DARE UK: www.dareuk.org.uk