Paper D: Data Access Management

Aim

Health Data Research UK is working in partnership with NHS organisations and other data controllers across the UK to lead a £37.5 million government investment on behalf of UK Research and Innovation to improve the safe and responsible use of health-related data at scale for research and innovation. To achieve this, we must address the difficulties in accessing data quickly (mentioned as a major barrier by 70% of respondents across researcher and direct industry engagement [1]), and identifying the location of data and understanding data quality (mentioned as a major barrier by 55% of respondents).

Through the DIH programme, seven Health Data Research Hubs and over 25 members of the Health Data Research Alliance have made their datasets discoverable through the Innovation Gateway MVP.

Our ambition is for the Gateway to support a streamlined, proportionate approach to access requests based on the five safes model for research and innovation uses with a clear public benefit, in line with the Principles for Participation[2]. We aim to make life easier for both requestors and decision makers through a combination of automation, built-in validation, transparency of progress and the capability to host virtual data access request panels. The intention is to build on existing cross-sector best practice both nationally across the UK and internationally.

The project

In the next development phase of the Gateway, HDR UK will work with the Technology Partner to provide an Access Management module that enables users to request access to datasets, submit required information and track the progress of applications directly through the Gateway. For data custodians with existing 'in house' solutions, the Access Management module would provide validated inputs to their approvals processes, whilst for data custodians that do not currently have an automated, web-based workflow it would provide a 'best of breed' web-based access management request solution.
In order to inform next steps for the development of this phase of the Gateway, we intend to work with the health data research community including data users, data custodians, health data research hubs, regulators and public and patients.

---

[1] Industry Engagement conducted as part of the DIH Design and Dialogue phase, which involved 32 in-depth interviews with industry representatives

[2] https://www.hdruk.ac.uk/wp-content/uploads/2019/07/Digital-Innovation-Hub-Programme-Prospectus-Appendix-Principles-for-Participation.pdf

We would like to hear from this community about current practices, any challenges associated with accessing and/or sharing data and explore any opportunities for us to improve future use of health data for research and innovation.

Planned activities and proposed timeline

Input from key players, including data custodians, users and patients will directly inform the Phase 2 of the Gateway development with end goal of delivering a Minimum Viable Product at the end of October which includes a functional Data Access management module. Below is our proposed timeline for activities.

Date	Activity
March-April	Engage Hubs and custodians via individual visits (including Scotland, Wales and Northern Ireland). *NB. These plans have been impacted by COVID-19*
2 April	1st workshop RAG sub-group – understand current practice, identify commonalities, pre-validation and standard checks
April-May	Individual interviews with Hubs representatives
29 April	2nd workshop RAG sub-group – draft wireframe and potential workflow for Gateway Alliance Board meeting – will update on what we are doing and who is involved and ask for names of others to contribute
1 May	1st Milestone to inform data access request module development
Mid-May	Data access wider workshop (Alliance + others) – or questionnaire
20 May	2nd Milestone to inform data access request module development (strawman produced)
12 June	3rd Milestone to inform data access request module development (focus on dashboards and reporting)

Progress so far

In addition to individual conversations with representatives from data custodians and the Research Hubs plus a desk-top review of existing data access request forms and processes, we held the first Innovation Gateway Data Access workshop held on 2 April. HDR UK Public Advisory Board members and representatives from HQIP joined a scheduled meeting of the NHS Digital Research Advisory Group (RAG) Streamlining & Ethics workstream, which includes NHS Digital, Public Health England, CPRD, Health Research Authority, the MRC Regulatory Support Centre.

Below are the main areas highlighted at the workshop.

Transparency and the importance of public trust

From a patient and public perspective, it is crucial that the processes are explained well and clearly so that key principles and steps are clear from the start. It is important to communicate clearly who can access the data, what data and what level of data can be shared, for instance if aggregated data or not, and where the responsibilities lie. Making these concepts clear to a lay audience is critical to building public trust. HDR UK could clearly play a key role in communicating key messages in an efficient manner.

It is also important for public to know how data custodians provide access to health data and if they involve patient and the public in decision making. Ideally, all data access committees should have PPIE (Patient and Public Involvement and Engagement) through direct representation. If not, more work could be done in finding people willing to join these committees and provide training to those willing to participate. Involving patients and public in decision making is crucial to demonstrating transparency and gain public trust. As an outcome of this discussion, HDR UK will be producing a document collating information about decision-making processes for each Alliance data custodian.

Pre-submission stage

It was noted that pre-submission stages are a very important step that may be harder to integrate into the Gateway.

Before applications are submitted, data custodians might offer a pre-submission service that includes assessing the feasibility of a project and availability and utility of data, understanding the methodologies of anonymisation, understanding application requirements and identifying potential issues.

Many pre-submission discussions with applicants are about understanding what is needed before submitting an application and it is crucial that applicants know how to articulate the reasons for using the data, the level/format of data required and how the use of that data would demonstrate health and care benefit.

Recognition of Data Security and Protection Toolkit (DSPT) as an accreditation system

All the data custodians involved currently use a 'data release' model primarily. As such assessing 'safe setting' for the data is specific to each application. The use and recognition of Data Security and Protection Toolkit (DSPT) as an accreditation system was highlighted. It was suggested that more work should be done on recognising DSPT as one of the main accreditation systems across data custodians, while also being open to other accreditations in some instances. Of note, if DSPT and further accreditations are required, it will be important to have a mechanism in place to perform these policy checks (ideally in an automated manner).

Validation checks could be integrated into the Gateway

In addition to DSPT, other checks that data custodians apply at various stages of the application process are similar between data custodians and might happen at the same time in the journey. HDR UK could take note of those checks and integrate them in the Innovation Gateway in an automated fashion so that burden on those custodians would be reduced.

Training, guidance and accreditation for researchers need to be customised

The group discussed the importance of training and the provision of clear guidance about how to write good applications. Key points raised were:

- Need to consider different audiences e.g., if applicants are from commercial or academic organisations, if the requestors are applying for data access on a one-off basis or are regular requestors. It might not be efficient to train people who request access on every 2-3 years or less (which is the majority of cases for academic applicants). Should funders or universities invest in training professional data mangers rather than researchers to help with applications?
- Guidance about UK-wide research would be welcome, acknowledging differences in the devolved nations.
- Information about researchers, including CVs, could be recorded in a central place. While not all data custodians have these types of checks in place, it might be helpful to be able to check the list of people mentioned on a project automatically, pulling through data from a database.
- The concept of research passport and research visas were mentioned. The Global Alliance for Genomics and Health (GA4GH) already uses these specifications to authenticate a researcher's digital identity and automate their access to requested genomic datasets. A similar approach could be used in the context of health data.

Common Project identifier to help join things up

Participants also talked about the utility of connecting information through different systems (including IRAS). For instance, it would be helpful to pull through the relevant approvals obtained before application, simply by having a project identifier connecting all the relevant documentation digitally. This would be particularly useful to link up ethical approvals and protocols to data provider systems.

Standard Data Sharing Agreements and common legal definitions

Participants highlighted that it is important to think about the type of legislation data custodians are complying with, as consent definitions under GDPR and common law duty of confidence differ. GDPR requirements are at the organisational level, common law duty of confidence applies at individual level.

The level of the data people need is different and data are not always re-identifiable. We need to share clear definitions and provide guidance about what anonymised/de-identified means.

It was suggested that given that all data controllers hold data obtained on NHS patients, a uniform NHS Data Sharing Agreement (DSA) would be very helpful to speed up the legal processes and negotiations.

One suggestion was also to consider what happens from a legal perspective when international researchers or organisations apply for UK data. International users might use different accreditation systems and might need to comply with different policies so careful considerations should be given.

Transparency of progress across systems

One of the potential benefits that the Innovation Gateway could deliver is greater transparency to both the data custodians and data requestors about the application processes. With sufficient commonalities across systems and the main process steps common across the pieces agreed, there is scope for the Gateway to go back and forth with individual systems possibly in an automated manner and to report that progress.

Essential process steps and areas of improvement

It was acknowledged that data sharing processes are not always linear, instead it is common to go back a step and seek clarifications at different stages of the process (for instance if ethical issues are raised). In the last session of the workshop we briefly discussed what the main steps of an ideal data access process should be and started to explore whether we could identify some essential steps and some non-essential steps. We also discussed areas of improvement and ideas that could be potentially taken forward when developing the Innovation Gateway access module.

It was proposed that some of the essential steps could include:

1. Initial assessment (includes pre-submission engagement)
2. Application submitted
3. Seeking independent advice and considering approval
4. Data access request approved and DSA signed
5. Data access granted (via data released or access to TRE)
6. Data destruction/safe outputs checks

Participants also provided specific suggestions to improve current processes or for HDR UK to develop a new efficient data access management process through the Gateway.

Specific suggestions and ideas are shown in the table below.

Process steps	Questions and suggestions
Pre-submission	How do users find out what applications/approvals are needed? How do they find out where datasets are available? How do they identify which datasets they need to answer the research question? Metadata/information about datasets available: one of the gaps identified is the level of information for each dataset available at the data custodian level. The Innovation Gateway with the provision of a rich metadata catalogue provides an opportunity to fill this gap. HDR UK could provide a flowchart for each data custodian member of the Alliance with PPIE in mind so that processes could be very clear for data requestors and members of the public. Infographics might be part of the guidance provided (to explain how data is kept safe/how the application process works). Can HDR UK (a central resource pointing people in the right direction) help navigate all the data custodians who have relevant datasets for a research project (e.g. to get linked datasets)? It might help facilitate complexity of requests. There might be an opportunity to use public approval release registers to help shape examples of successful requests and also provide guidance. Guidance could include information on how and when to apply for grant funding to avoid delays in starting the research due to application/approval processes (this would be relevant to academic users).
Submission	Dashboards: both data providers and data requestors need to track progress of submissions. Notifications and emails back to requestors should all be automated.
Approvals/Documents checks	Linking approval documents to data provider: e.g. HRA approvals and protocols approved could be linked to data providers systems. This would allow data providers to check that application forms matches the CAG/HRA approval. Linking CVs, information on researchers and information on organisation could be recorded on a central place and could be pulled through. Built-in validation opportunity: security. If all data custodians agreed on a minimum level of security (e.g. DSPT) or set a minimum threshold for 'safe people', it would be helpful.
Data Sharing Agreements	Standardised national data sharing agreements for the same set of data could be developed.
Data release	Data transfer or Trusted Research Environments? There is increase demand for TRE use when accessing NHS Digital datasets. How can we switch to TRE use by default?

Audits	More information sharing might help streamlining processes. Data providers could coordinate efforts to have a standardised approach to auditing – e.g. standard procurement. For instance: if most of the data controllers rely on DSPT accreditation – can the audit concern go through DSPT?
Post-data release	Tracking applications post-approval: we could have an automated system for researchers to put in a renewal and/or have a certificate destruction triggered automatically. It would be useful to see if researchers have applied through other organisations – and link applications. There might be scope for HDR UK to look at safe outputs – how can publications and outputs be tracked? There should be checks in place that safe outputs are compliant with data sharing terms.

Next steps

A series of engagement events will be carried out throughout April and May. A second workshop with data custodians is scheduled on 29th April 2020. This workshop will focus on the Innovation Gateway and we will ask participants to input more specifically on how we can build a solution that can work for most data custodians, having in mind three use cases:

1. Data custodians with existing 'in house' web-based workflow solutions (e.g. NHS Digital).
2. Data custodians with paper-based documentation and offline workflow.
3. New data custodians or research hubs who require a web-based access management request solution.

Outputs from these workshops, alongside the outputs from engagement exercises with the Health Data Research Hubs and the user community will directly inform development of the access request module in the Innovation Gateway. We are planning to develop an infrastructure that supports access request workflow management for existing data custodians, but that also manages the end-to-end access request process for new data custodians, e.g. Hubs.

Appendix: Background information

Gateway User journey

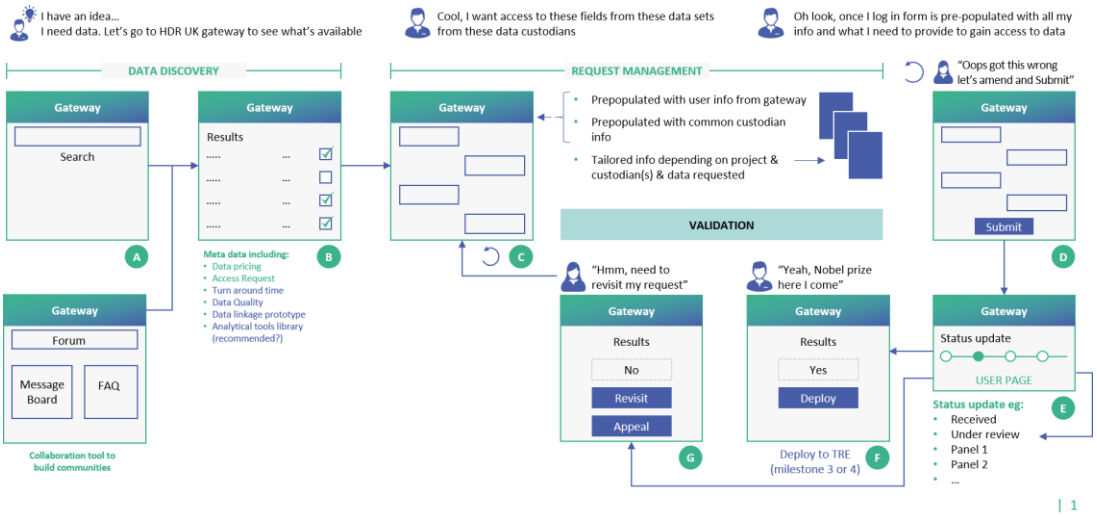

***Figure 1**. High level user journey proposed for the Health Data Research Innovation Gateway.*

Existing tools

Health data access toolkit has been developed by the **MRC Regulatory Support Centre** to help researchers navigate the processes required to gain access to routinely collected NHS health data, signposting the approvals required and exploring the issues that need to be considered during the application process.

Health Research Agency provides a number of decision support tools:

- Ethics: http://www.hra-decisiontools.org.uk/ethics/
- Confidential Advisory Group (CAG): http://www.hra-decisiontools.org.uk/cag/
- What is research: http://www.hra-decisiontools.org.uk/research/
- Managing consent: http://www.hra-decisiontools.org.uk/consent/

NHS Digital has produced guidance on the information expected in the various sections of a DARS application, named 'Standards'. For each 'Standard' there is written guidance covering important items to consider when writing your application and, where available, an accompanying video.
https://digital.nhs.uk/services/data-access-request-service-dars/dars-guidance

HDRUK
Health Data Research UK

Outline of the end to end data access and use process

In scope

1. **Study proposal development:** requiring identifying likely data sources, their applicability to the study topic, how best to manage/analyse, developing project plan and indicative budget.
2. **Study funding proposal:** usually undertaken prior to seeking approvals, but needs to reflect likely data acquisition costs/fees as well as realistic timescales.
3. **Application development & submission** (for each data source and possible CAG/REC as well): based on study proposal, but detail may need to be tailored to each data provider/approver; providing supporting materials (Data Flow Diagrams, legal approvals/bases, security assurance/documentation, etc.). May include a registration process for new organisations / users and may involve more than one data custodian.
4. **Pre-vetting:** a process carried out by the approval secretariat (or equivalent) that forms have been correctly completed, necessary documentation provided, legal & governance requirements met – this may require the application to be amended/corrected and re-submitted
5. **Actual Approval review:** this may require scheduling for next available committee meeting (and hence depend on overall workload); some processes allow for 'precedent' consideration by the secretariat and/or Chair; decision may be to reject/approve/require amendment usually with reasons and/or conditions to be met, so application may loop back to any of the earlier steps

Out of scope for initial Data Access Management Module development

6. Contract development and signing – standard set of terms for each data provider, potential for legal negotiations. This may give rise to further questions about legal bases, controllers/processors, and even require a return to the very beginning of re-designing the study protocol.
7. Data release: resources available at each data provider to generate the required dataset / provide access to the research environment.
8. Data Analysis: this can show up data quality problems and/or be delayed through lack of resources at the study centre.
9. Data Destruction: Usually a contractual requirement if data release model. Extensions may be allowed but may require access request process again.
10. Publication: may depend on contractual requirements around publication, audit of safe outputs may be required and acknowledgement of data sources in publication.
11. Audit/Monitoring: checking that security is maintained, data held safely and processed appropriately, including deletion as noted above

Cross-sector national approaches

The Office for National Statistics Secure Research Service (SRS) gives accredited or approved researchers secure access to de-identified, unpublished data in order to work on research projects for the public good.

https://www.ons.gov.uk/aboutus/whatwedo/statistics/requestingstatistics/approvedresearcherscheme

The METADAC (Managing Ethico-social, Technical and Administrative issues in Data ACcess) is a multi-agency multi-study data access structure that services several of the UK's major cohort studies and provides a scalable mechanism to incorporate additional cohorts in the future.
https://www.metadac.ac.uk/

International approaches

Data Access Support Hub (DASH): a one-stop shop for requesting access to multi-jurisdictional data across Canada https://dash.hdrn.ca/

Data Use Oversight System: Expediting data access for researchers, by facilitating and enhancing data access committees' workflows https://duos.broadinstitute.org/