Health Data Research Innovation Gateway

Gateway Portal – MVP Specification   July 2019

Contents

Purpose of the Document

Health Data Research UK is requesting proposals for an online portal (Gateway Portal – MVP) as a component of the UK Health Data Research Innovation Gateway (the Gateway). This document provides further detail for suppliers who are responding to the opportunity posted on the Government Digital Marketplace. The Gateway is part of the UK Research and Innovation (UKRI) Industrial Strategy Challenge Fund Digital Innovation Hubs Programme.

Health Data Research UK
Health Data Research UK (HDR UK) is the national institute for health data science. Its mission is to unite the UK's health data to make discoveries that improve people's lives. By working in partnership with the NHS, industry, academia and patients, and providing safe and secure access to rich health data, it aims to better understand diseases and discover new ways to prevent, treat and cure them. HDR UK is leading the development and delivery of the UK-wide Digital Innovation Hub Programme. For more information visit www.hdruk.ac.uk

Industrial Strategy Challenge Fund
The Industrial Strategy Challenge Fund, part of the Government's modern Industrial Strategy. The fund is delivered by UKRI. The Industrial Strategy Challenge Fund brings together the UK's world-leading research with business to meet the major industrial and societal challenges of our time. It provides funding and support to UK businesses and researchers, part of the government's £4.7 billion increase in research and development over the next 4 years.

UK Research and Innovation
UK Research and Innovation (UKRI) brings together the UK Research Councils, Innovate UK and Research England into a single organisation to create the best environment for research and innovation to flourish. The vision is to ensure the UK maintains its world-leading position in research and innovation. For more information visit www.ukri.org

Introduction

The major barriers to the use of data for research and innovation are the inability to quickly access data (reported by 70% of respondents across direct industry engagement[1]) and the inability to identify the location of data and understand data quality (reported by 55% of respondents).

The Health Data Research Innovation Gateway is being established as part of the Digital Innovation Hubs (DIH) programme which is being run by HDR UK. The vision of the DIH programme is to make the UK home to data-driven research, scientific advances and innovation in healthcare to improve patient outcomes. This will be achieved through:

i.   The UK Health Data Research Alliance: a partnership of data custodians, uniting the UK's health data, facilitating partnership working and standard setting across NHS organisations and other data controllers
ii. Digital Innovation Hubs: centres of excellence which will improve the quality of the data and provide expert services to support its use
iii. The UK Health Data Research Innovation Gateway: a UK-wide platform which will support health data science. This is described in more detail below

More detail on the Digital Innovation Hubs programme is given in the Prospectus, [available here](#).

The Gateway
The Gateway will address the challenges of access to health data. Its function is to provide a common access point to UK health research data for industry, researchers and innovators. The Gateway is designed to support the whole user journey, from discovering the data that is available, coordinating an access request across multiple data custodians, requesting a linkage of multiple datasets and facilitating access to these through Safe Havens. The Gateway will be available to industry users from all sectors, as well as to researchers and innovators across health and social care and academia. While the focus of the Gateway is to support research and development in industry, academia and the NHS, HDR UK will work in partnership with NHSX and other NHS bodies to align this work with mainstream NHS endeavours, including the development of clear standards for the format and use of in the NHS.

The Gateway will support the use of health data, acquired by members of the UK Health Data Research Alliance (the 'Alliance') and Digital Innovation Hubs (the 'Hubs'), which will be held in Safe Havens to provide a safe location for data storage and access, facilitate interoperability, and provide analytical capability[2]. **Only the rich metadata (i.e. no personal identifiable information) that describes the detailed structure, content, permitted usage and community feedback will be stored locally in the Gateway.**

Gateway Design Approach
The Gateway will build on tools and best practices that already exist in the UK and make use of realised assets and existing 'best of breed' solutions which will be sought from the market. Custom development will only be required where a suitable product or application does not already exist. Harnessing capability

---

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

[2] As part of the UK Health Data Research Alliance, a workstream has been established to define the functions and best practice of Safe Havens across the UK, with a view to agreeing on a standard/accreditation process.

across the UK, the majority of the elements of the Gateway will be developed by industry experts in partnership with NHS and academia. Once developed, HDR UK will work with partners to evaluate the appropriate long-term approach to running a sustainable, reliable and scalable service.

The Gateway will be developed in two phases. First will be a Minimum Viable Product including the metadata catalogue. The second phase will deliver the full end-to-end Gateway, which will be delivered through a technology partnership. Each component will be required to integrate with the other components through open application programming interfaces (APIs). The work completed should follow HDR UK's and NHSX Development Principles, given in the Appendix. All modules should be developed such that they can be replaced and extended as requirements evolve.

The Gateway will not take on the role of data storage. Data security is paramount. Data will remain under the control of existing data custodians in certified Safe Havens. HDR UK will partner with Save Havens (e.g. SAIL, Scottish National Safe Haven, NHS Digital and other academic, NHS and commercial providers) to ensure integration with the Innovation Gateway and this integration will be a key deliverable in Phase 2. The Gateway will operate under principles of federated access, with existing data custodians retaining robust governance procedures to ensure that any access to data remains appropriate, proportionate and necessary. The Gateway will provide the ability to operate at scale, as further health data research activity takes place.

Delivery Phases

Phase 1 of Gateway development will commence in July 2019. Its components will operate independently and integrate through fully documented open APIs. Components of the Gateway will be developed in the open, using open-source principles and with a focus on providing services that incorporate best practice in security, reliability, availability and scalability such that it can meet the business critical requirements of industry as well as researchers from academia and the NHS.

As part of the wider Digital Innovation Hubs programme, the development of the Gateway is being timed to coincide with the priorities of the Alliance, and the data which will be made available through the Digital Innovation Hubs. The hubs will be expected to make initial metadata available by December 2019, which will be included in the Gateway to facilitate discovery.

Phase 1

Phase 1 has two primary functions:

- A **Gateway Portal** – MVP (Minimum Viable Product) (this document). The portal will be a web application hosted on a public cloud that will provide the front-end for users to discover and request access to datasets, and to providing supporting capability such as user profile management and logging of activity.
- A **Metadata Catalogue** which will provide the underlying functionality to support the Discovery, Access Management and User Experience Portal, and will be developed through a separate process. The metadata catalogue will be a web hosted backend application providing all the functionality to manage the metadata for the datasets discoverable through the portal. The request for proposals document for this component is available here.

This specification should be read as a supplement to the opportunity posted in the Government Digital Marketplace (GDM). The metadata catalogue is being tendered through a separate process.

This specification outlines the request for the **delivery of an online portal (the Gateway Portal – MVP)** which will provide a straightforward and good user experience and the following functionality:

- Allow users to **discover data** through a common point of access, which provides a view of available datasets, enabled by federated metadata management and search tools to provide improved visibility, comparability and navigation of existing datasets.
- Provide a method of requesting **access to data** that retains data controllers' requirements whilst harmonising processes, reducing transaction costs and improving access to more isolated, important datasets.
- Embed analytics capability to **assess use** in order to understand how to meet the needs of users and support further access to data.

Phase 2

The future development of the Gateway will be delivered through a technology partnership. This is out of scope for this contract, but an outline is provided here as additional context on our ambitions for the Gateway.

HDR UK will be requesting proposals for potential technology partnerships in Q4 2019. The selected technology partnership will be responsible for ongoing development of the Gateway over at least three years. It is our expectation that the partnership will consist of a lead partner with an established international track record of deploying complex and secure web applications working with other partners with domain specific leadership in areas such as data de-identification.

Through the technology partnership, further functionality will be developed in the Gateway to:
- **Assess** and pre-validate data requests based on workflow management to improve response times for decisions, whilst allowing data controllers to stay in control
- Extend the existing **access** functionality to provide a business process engine, collecting information on requests, turnaround times and use, allowing for seamless and rapid processing of requests
- Link with best in class de-identification and encryption software across the Alliance in **privacy** and interoperability, in line with information governance requirements, to provide secure and novel data linkages across a wide range of datasets to meet researcher needs while ensuring privacy of sensitive health data
- Harness **interoperability** to facilitate access to linked datasets across the data custodians in the Alliance. The interoperability support will also extend to support mapping between datasets. This will focus on enabling re-use of mapping carried out as part of planned research projects, as well as codifying interoperability issues between different datasets, for example through the calculation of an interoperability score which would be available to other researchers considering working with the same datasets.
- Provide a **library** to act as a single storage area supporting reusable or shareable analytics, linkage maps and phenotypes.
- Support close integration with certified **Safe Havens / Trusted Research Environments** provided through partner organisations in academia, NHS and industry to allow for direct provisioning of flexible analytical workspaces.

High Level General Requirements for the Gateway Portal - MVP

The specification for this work is to create a usable web portal which will provide the core functions outlined below, and integrate the metadata catalogue, being developed in parallel, to form the Phase 1 Gateway. The code for the portal will provide an open and extensible fabric for development of community driven tools built upon the underlying code.

The portal is required to provide the following functions, and these are covered in more detail in the below section:

- **Search and discovery**: provide a standardised way to locate network-accessible material (including but not limited to datasets, data transformation and models) which are integrated with the metadata catalogue through open APIs
- **Access Management Workflow:** provide the management of a data access request approval workflow for dataset collections available
- **Identity and User Profiles Management:** provide a common component to allow the identification, validation and authentication of users of the Gateway
- **Audit and Logging:** provide a common component to allow the logging of all user-generated activity and to generate metrics for monitoring the use of the portal
- **Integrated Web Portal User Experience:** provide a common component to allow the efficient, consistent, available, dependable, reliable web portal that integrates the above components in an intuitive, easy to use, understandable, navigable and learnable web portal that is also flexible and adaptable to cater to future user needs

The development should be consistent with the HDR UK Development principles (see Appendix 1). In this project, all operations and features must be exposed via an open and documented RESTful API and data model.

During development, the project team will work under the guidance the HDR UK Chief Technical Officer (CTO) and Solutions Architect, who will support the development of stories as part of the agile process, with the CTO acting as the design authority. The specific requirements outlined below provide an overview of the elements required, but these may be superseded by requirements from co-development with users, in line with agile working practices. HDR UK will provide access to a wide stakeholder community for the development, testing and improvement of the tool, including members of the Alliance, hubs, industry users and researchers. The project team will run a formal design workshop with the stakeholder community to identify user journeys, personae and priority user stories. These user cases will form the basis for the continued application of design thinking and user engagement through the project delivery.

Members of the project team will be required to demonstrate working iterations of the tool in fortnightly sessions at HDR UK's headquarters in London, to a panel of HDR UK senior team members.

Specific Technical Requirements

The following items detail specific requirements, categorised by type:

Search/Discovery

Description: Provide a standardised way to locate network accessible material (including but not limited to datasets, data transformation and models) which are integrated through open APIs

Components and functions:

Using deep integration with the metadata catalogue, support the frontend capability for the following capabilities.

- Detailed summary of available datasets, which should include facets for granular searching of metadata results.
- Integration with an ontology server to support queries based on common coding standards and terminologies
- Ability to support exploratory browsing with simple general queries
- Ability to search entities by type, classification, attribute value or free text
- Rich RESTful APIs to search by complex criteria (property facets, Boolean and comparison operators)
- SQL-like query language to search entities - Domain Specific Language (DSL)
- Provision of an API for searching, tagging and management of Metadata
- Ability to use templates, viewpoints and perspectives to drive the presentation of metadata that we hold on each of the datasets, offering different views of that metadata to support different presentation modalities

Data Access Management Workflow

Description: Allow the management of a data access request approval workflow for dataset collections available. In Phase 1 the committed scope will be to support the submission of a harmonised access request which will later be extended to full workflow with rules based pre-validation. This additional capability is targeted for Phase 2 but will be a stretch objective for Phase 1. This work will need to be designed in collaboration with both research users and data controllers.

Components and functions:

- Ability to import and create sectioned data access forms with a rich set of configurable form fields: multi- file uploads, text boxes, dates, multi- and single select fields.
- Provide a minimal data access request form that will cover at least the first request stage on any data request held within the Gateway. (e.g. name of dataset, name of user, research purpose, etc.)
- Ability to extend and customise minimal data access forms for specific organisational and dataset-specific needs.
- Provide dataset access request management workflow templates that will minimally define three data access request approval processes:
    - Received: Data Access Request
    - In Progress: Data Access Request
    - Result: Data Access Request
- The data access request management workflow stages must be extensible to allow data custodians to add additional stages (e.g. Review and Authorisation stages)
- The data access management workflow module will provide a simple process editor that will allow users (e.g. data custodians) to quickly modify and extend the approval process. This should also support rules-based pre-validation of request to minimise errors in submissions

- The tool will be integrated with the user profiles management module to ensure all users of the system are identified.
- Ability for the data custodian to delegate/assign parts of the approval process to other users or groups of users
- The tool will provide a process dashboard to data requestor and data requestee that will allow both parties to view data access requests in one place.
- The tool will also allow both parties to communicate to request additional information to help clarify the access request
- The module must be able to support access requests where the linked datasets may be owned by multiple custodians and may need to be combined by a trusted third-party organisation
- All activity is logged so that each data access request has an accompanying audit trail of who did what and when.
- The logging and audit facilities will allow for data custodians and approvals committees to identify bottlenecks in their approval management process and support continuous improvement to access management.

User Profiles Management

Description: provide a common component to allow the identification, validation and authentication of users of the Gateway

Components and functions:

- Self-registration using supported Identity Providers and validation emails
- Password management and password management policies, including two-part authentication
- Framework to support integration / syndication with external user management systems
- Administration UI to allow organisational (group) management of users
- Support of user-managed profiles to allow previous data request information to be reused in future data requests
- Support for central administration of users
- Integration with Audit and Logging module for User activity logging

Audit and Logging

Description: Provide a common component to allow the logging of all user-generated activity and to generate metrics for monitoring the use of the portal

Components and functions:

- The auditing and logging module will provide the following system behaviours: Monitoring, Alerting and Analysis
- Logging of:
  - All user access and activity (User Profiles), including the ability to monitor activity by organisation and sector
  - All queries executed through the discovery modules (Discovery)
  - All dataset access requests initiated, on-going and completed (Data Access Request)
  - All datasets explored and viewed
- Integration with all modules: Discovery, Access Request Management and User Profiles Management

- Support for the development of metrics to be visualised through customisable dashboards

Integrated Web Portal User Experience

Description: provide a common component to allow the efficient, consistent, available, dependable, reliable web portal that integrates the above components in an intuitive, easy to use, understandable, navigable and learnable web portal that is also flexible and adaptable to cater to future user needs. It is expected that the user experience will be designed using a formal design process working with the key user communities including innovators from industry, academic research and data controllers. This should build on the user cases referenced above.

Components and functions:

- Design and align the development of web UI components with the HDR UK Design Thinking approach
- Design and implement a seamless web user experience integrating all modules: discovery modules, data access request, user profile management and audit & logging modules
- Integration with common UI web components using HTML + JavaScript, CSS using standard web frameworks
- Provide the unified user self-registration, login and profile management web assets
- Provide web assets to support the registration and update of new datasets including the import of descriptive minimal metadata and data access request forms
- Provide the metadata browsing and search query interface web assets needed for the discovery module
- Provide the data forms and approval process editor web assets to support the data access request management workflow
- Provide the web assets for administrators to view and analyse access logs from the user profiles module, web assets for view and analyse metadata search queries, web assets for data access requests
- Provide common web UI components that could be reused as UI interfaces for other web assets within the Gateway.
- In addition to the functionality the solution must also support following quality attributes:
  - Accuracy
  - Ease of interpretation
  - Usability and Utility
  - Interoperability and Integration
  - Automation
  - Performance
  - Consistency
  - Reliability
  - Availability
  - Appearance and Navigation

Documentation and Training

- All code and API's to be fully documented using appropriate self-documenting tools, e.g. using Swagger.

- Contextual help to be provided through the Gateway user interface. This will integrate with the contextual help from the Metadata catalogue.
- Written documentation for all components.
- Three days of deep dive technical training to HDR UK technical staff focused on support requirements.
- Three days of deep dive administrative training to HDR UK staff on Gateway function, administration including supporting components.

Project Delivery Timetable

- Supplier commences work using agile approach – 23rd September 2019
- Milestone 1 – (Tuesday 15th October 2019) 2 agile iterations of the portal, fully operational and accessible by users, providing basic discovery function across the datasets in the metadata catalogue, this will include the alliance datasets that are available in the metadata catalogue by this milestone
- Milestone 2 - (Tuesday 12th November 2019) 2 further agile iterations of fully operational and accessible by users' versions of the Gateway Portal - MVP
- Milestone 3 - (Tuesday 10th December 2019) 2 further agile iterations of fully operational and accessible by users' versions of the Gateway Portal – MVP
- Milestone 4 – (Friday 10th January 2020) Fully operational Gateway portal delivered, including discovery function across the hub data available through the metadata catalogue

Appendix 1: Health Data Research UK Development Principles

HDR UK follows the principles detailed below which align with those of NHSX, and any technical work by and for HDR UK will be required to follow these processes and principles:

Process Principles

- All software developed by or on behalf of HDR UK will be shared as open source under the HDR UK GitHub organisation.
- All code developed will be under the MIT licence.
- All work is design thinking led (https://dschool.stanford.edu/resources/getting-started-with-design-thinking) – all elements should be tested with users in order to ensure it meets users' needs
- Development should take place in a modular fashion as technologies and tools will regularly advance and therefore need to be able to be replaced quickly and easily. All aspects to be supported through interoperable open APIs
- Development should take place in an agile fashion, with multiple iterations and improvements throughout the process, using the concept of a Minimum Viable Product (MVP) to launch improvements quickly. Within the fortnightly sprints, work should be prioritised around delivering working features and software after each iteration
- All front-end development should adhere to W3C standards as documented at https://www.w3.org/standards/webdesign/ include those standards associated with accessibility such as WCAG 2.1.
- Continuous integration and deployment will be used to allow for rapid feedback on all new features (code talks)
- Use open source tools, technologies and frameworks wherever possible - custom development should be by exception only and when discussed with the Technical Solutions Architect and approved by the Chief Technology Officer
- All software developed by or on behalf of HDR UK will make use of open standards. For example, existing open standards for data and metadata (e.g. DCAT, DataCite, Schema.org)

Architecture Principles

- The Gateway should provide the user community with an exceptional user experience
- The Gateway elements should not have fixed dependencies or be tightly coupled to other components, and should be able to integrate and scale as needed
- The Gateway activity data must be recoverable to a specific point in time after any technical failure
- The Gateway will be built with GDPR and other regulatory frameworks in mind, ensuring the features and documentation to remain compliant
- The Gateway should integrate and support the development of external tools and services
- The Gateway will seamlessly scale in response to high and varied loads and the software must be developed to be inherently scalable
- Trust in data and data safe keeping is paramount, User Interface, End Points and APIs should be secured to enterprise levels with input validation, encryption, using https and storing one-way password hashing and string authentication.
- Platform / Data Access, User Actions and Transactions together should be audited
- Changes to a Metadata record should generate a history of change, preserving the old record as a previous version.

- The Gateway should have a rapid deployment infrastructure using virtualisation and continuous integration i.e. Kubernetes, Docker, Virtual Box or amazon Cloud Formation
- To provide for interoperability, the FHIR (http://hl7.org/fhir/) standard will be used to model all resources.

Principles for Participation

As development of the Innovation Gateway forms part of the Digital Innovation Hubs programme, all suppliers will be required to sign up to the Programme Principles of Participation, available here.