

HDRUK
Health Data Research UK



**INDUSTRIAL
STRATEGY**

UK Research
and Innovation

Supporting development of the Innovation Gateway – Metadata Onboarding

November 2019

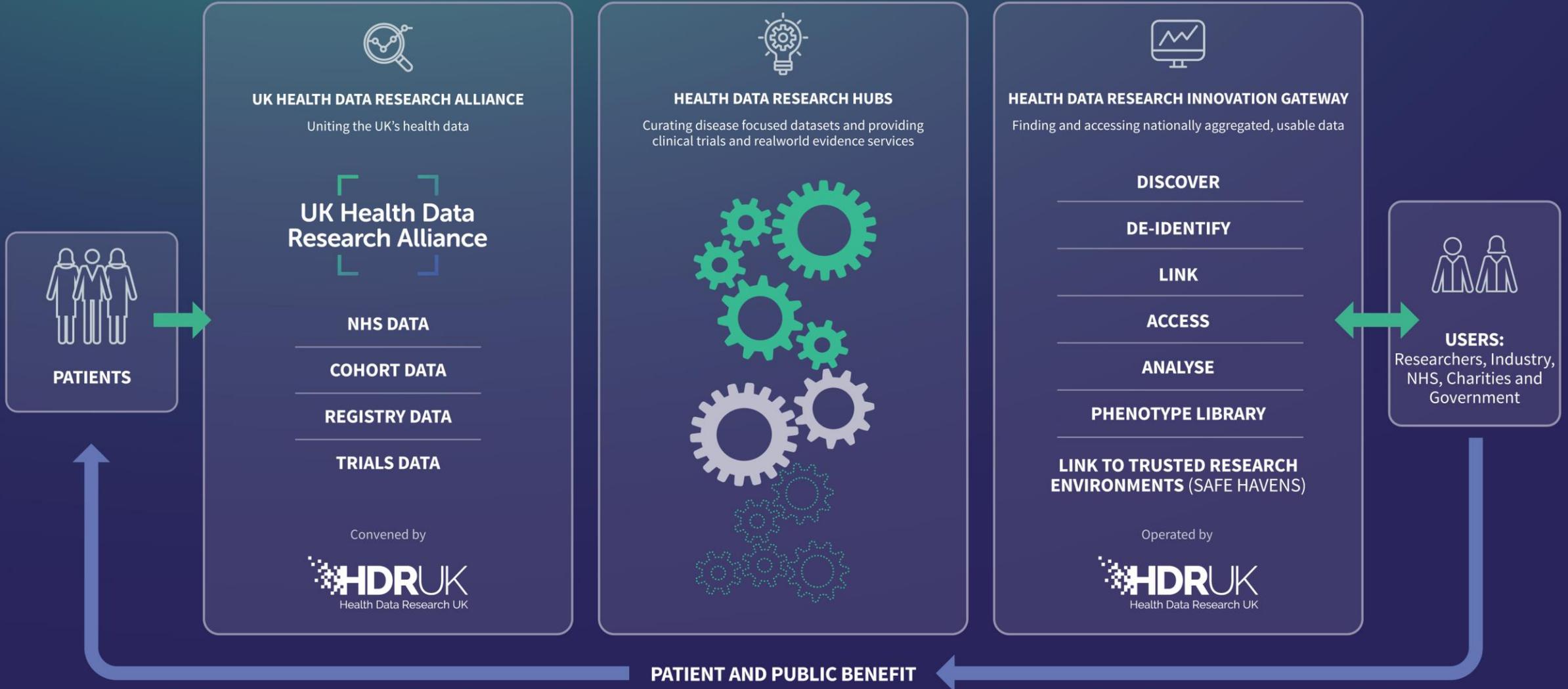


Agenda

- Welcome & Introductions
- Overview of the UK's robust health data research infrastructure
- Metadata onboarding – how it fits it
- Two phase investment in Alliance members
- Overview of Metadata requirements
- Questions

UNITING THE UK'S HEALTH DATA TO MAKE DISCOVERIES THAT IMPROVE PEOPLE'S LIVES

← TRUSTWORTHY USE OF DATA →



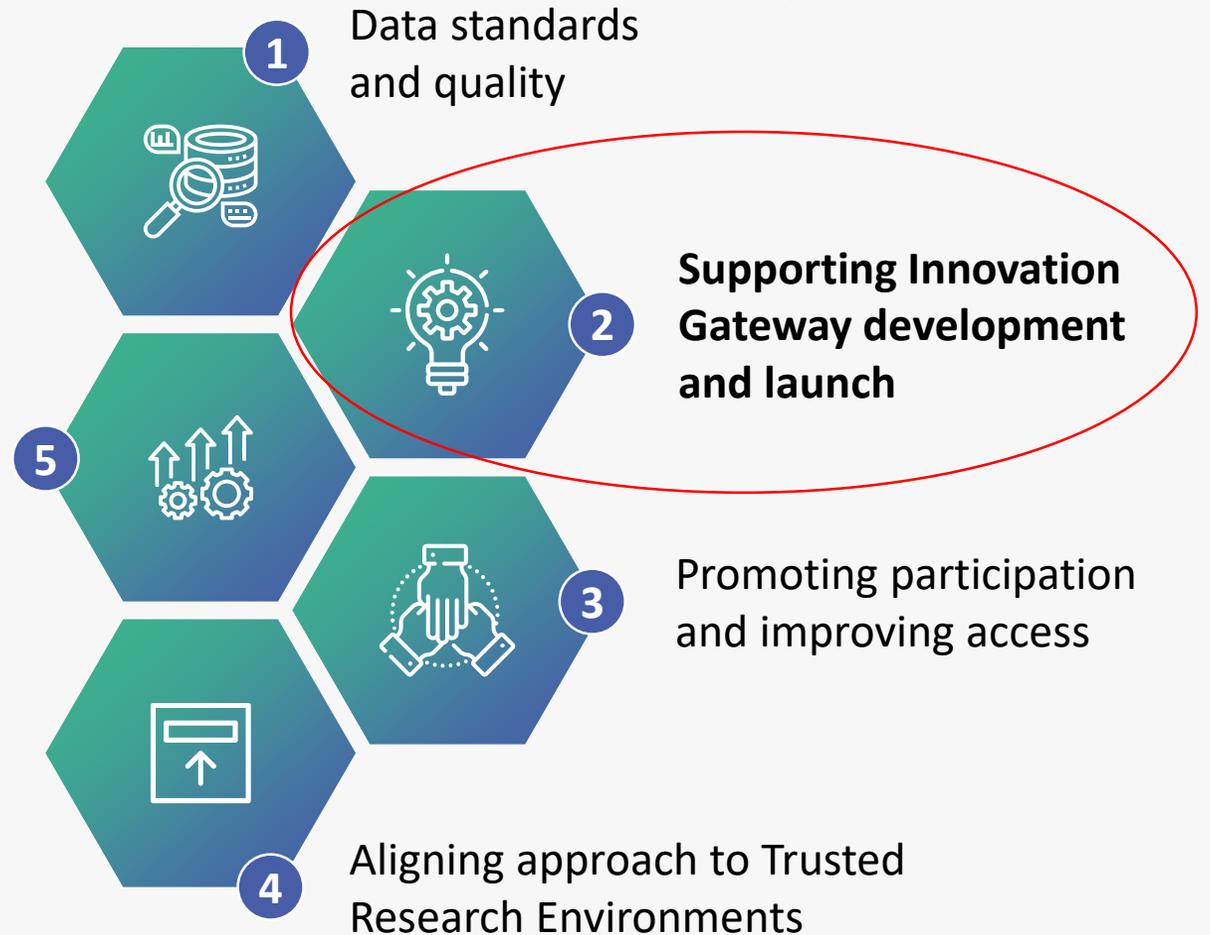
Building a UK-wide capability, where we are today

- The overarching programme - Building a UK-wide capability that will enable the safe and responsible use of health-related data at scale for research and innovation
- 25 members of the UK Health Data Research Alliance (including cohort to be announced shortly)
- 7 Health Data Research Hubs
- Innovation Gateway – the metadata catalogue, metadata onboarding, Minimum Viable Product – Phase 2 – Technology Partnership
- 7 members of the User Group – the Sandbox
- Our roadmap and how we are going to be working together



UK Health Data Research Alliance workstreams

Engaging and involving practitioners, patients and the public



Health Data Research Innovation Gateway



The Gateway is a common access point to UK health research data for accredited researchers and innovators

Developed in 3 phases:

Phase 1 (Minimum Viable Product) – Sep 2019 to Jan 2020

- Portal (Discovery and Access) – developed by IBM
- Metadata catalogue – delivered by NHS Digital and University of Oxford
- Metadata Onboarding - delivered by Parity & Metadatavorks

Phase 2 – Mar 2020 to Aug 2022

- Technology Partnership



HEALTH DATA RESEARCH INNOVATION GATEWAY

Finding and accessing nationally
aggregated, usable data

DISCOVER

DE-IDENTIFY

LINK

ACCESS

ANALYSE

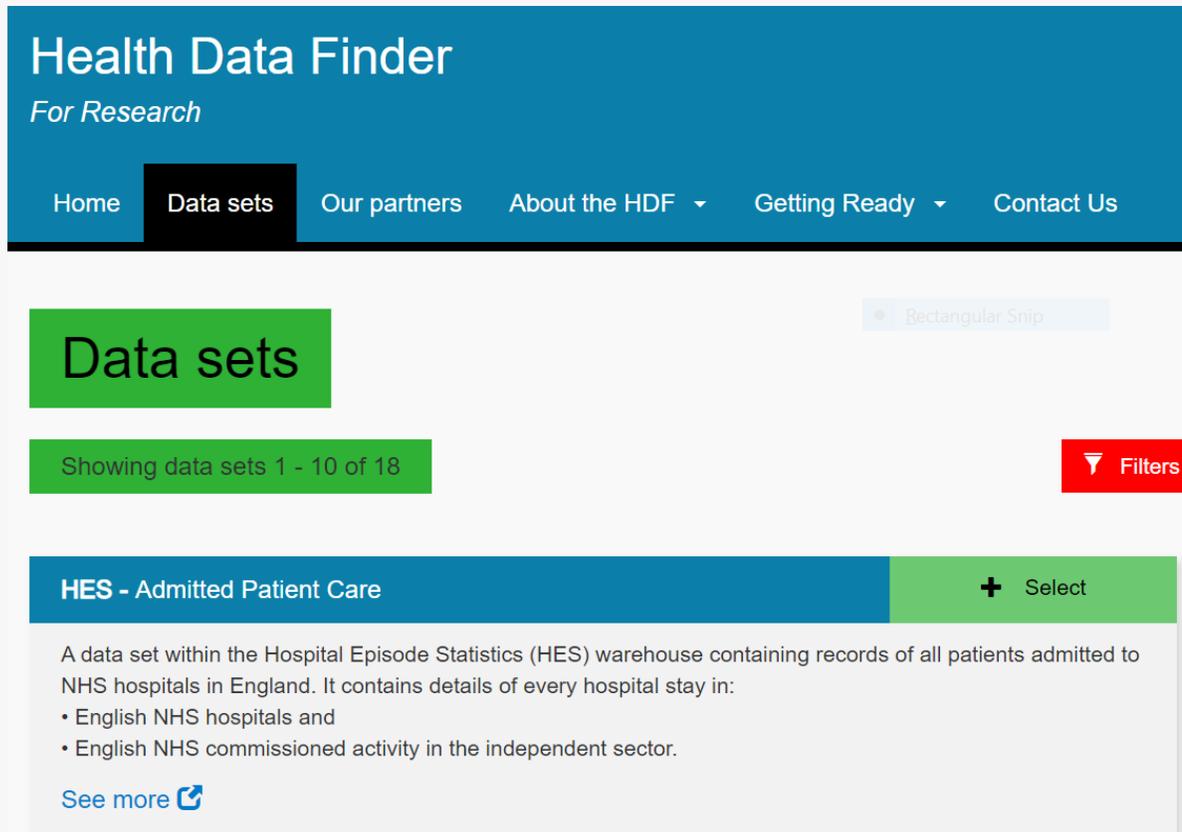
PHENOTYPE LIBRARY

LINK TO TRUSTED
RESEARCH
ENVIRONMENTS
(SAFE HAVENS)

Operated by



Building on the NIHR Health Data Finder



The screenshot shows the 'Health Data Finder' website. At the top, there is a blue navigation bar with the title 'Health Data Finder For Research' and a menu with items: Home, Data sets (highlighted), Our partners, About the HDF, Getting Ready, and Contact Us. Below the navigation bar, there is a green box with the text 'Data sets'. Underneath, a green box indicates 'Showing data sets 1 - 10 of 18'. To the right of this is a red 'Filters' button. The main content area features a card for 'HES - Admitted Patient Care' with a '+ Select' button. The card description reads: 'A data set within the Hospital Episode Statistics (HES) warehouse containing records of all patients admitted to NHS hospitals in England. It contains details of every hospital stay in: English NHS hospitals and English NHS commissioned activity in the independent sector.' A 'See more' link with an external icon is at the bottom of the card.

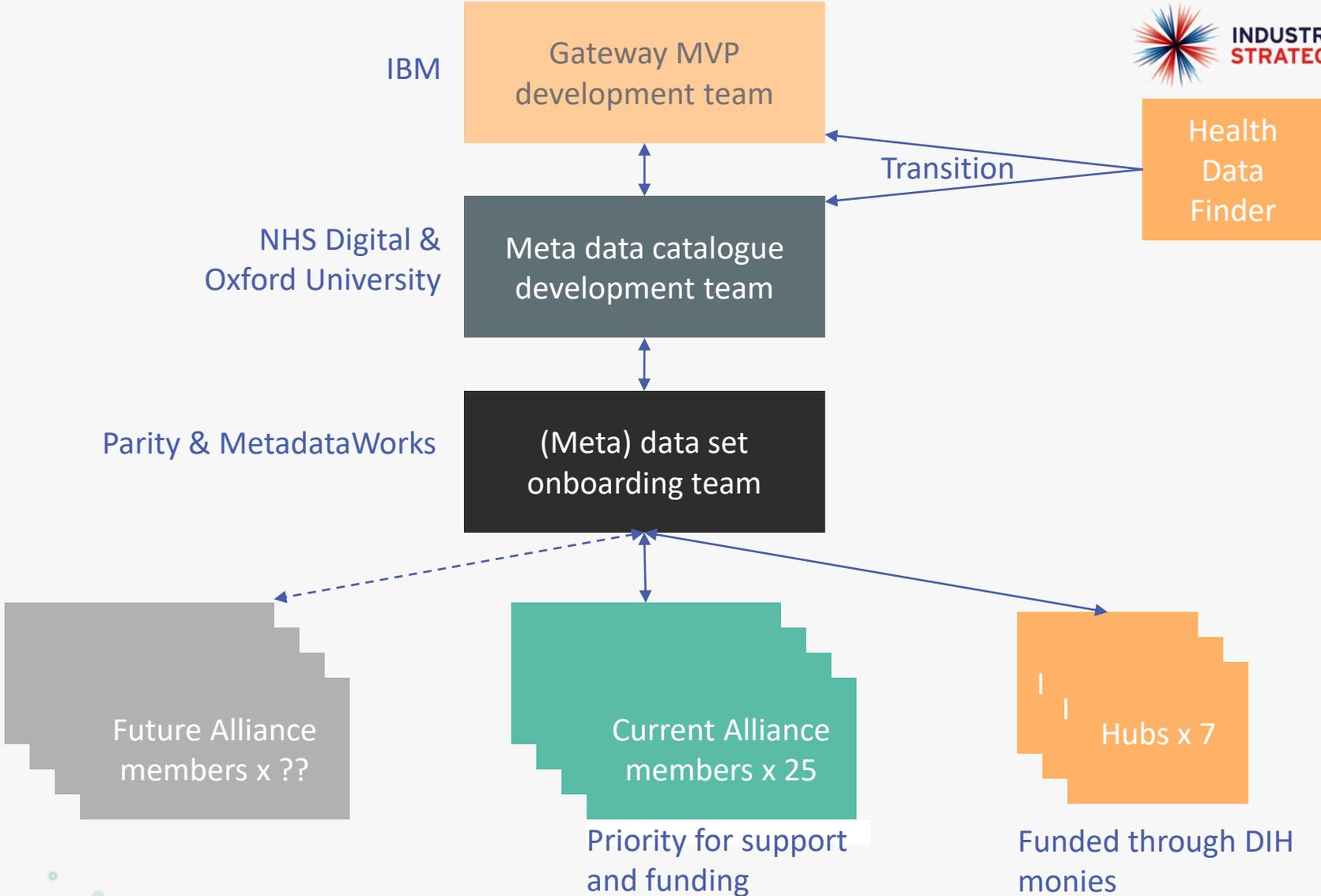
Improved User experience: Search and browse functionality, first steps to data access requests

More datasets for discovery: UK wide, greater scale and diversity through Alliance

Tiered functionality: Registered user login

Greater insight: Improved analytics of user behaviour and needs

MetaData onboarding – who and what is involved?



Metadata Onboarding team – scope of work

- **Rapid evaluation of the current landscape of Alliance and Health Data Research Hubs datasets**, the metadata status, the data custodian’s readiness for onboarding, and potential investment required to develop a sustainable process for onboarding and updating metadata in a timely fashion.
- **Co-ordinate the metadata onboarding process for Alliance members and Health Data Research Hubs**, working closely with the metadata catalogue supplier and providing advice and targeted support to Alliance members to complete the onboarding process.
- **Demonstrate ‘proof of concept’ for onboarding the metadata** from at least two Alliance members that have not previously been involved with Health Data Finder.
- Identify and, where possible, address **opportunities to improve the efficiency and sustainability of the process** for the benefit of future datasets and ensuring metadata available through the Gateway remains current.

Two phase approach to investment in Alliance members

- 1. Rapid evaluation** of 'within scope datasets' across all members of the UK Health Data Research Alliance.
- 2. Targeted support** for preparation and onboarding of metadata based on prioritisation following the rapid evaluation work.



- Both phases will be co-ordinated by Parity & MetadataWorks and form part of the development of the Innovation Gateway Minimum Viable Product (MVP).
- The MVP is scheduled to be launched on 10 January 2020
- It is expected that funded onboarding of metadata will continue until the end of March 2020 and demonstrate how metadata can continue to be added and updated

Phase 1: Rapid evaluation work

- **Organisational engagement fee.** Same across all Alliance members who agree to engage. It is proposed this should be £5,000. Payment of this engagement fee would be the return within 2 weeks of an information request covering:
 - “Within scope dataset return”. List of datasets/ data assets that are in scope for the rapid assessment and potential loading onto Innovation Gateway Metadata Catalogue.
 - Provide basic access request process or documentation (or links to it) – this could be achieved by confirming that the information provided on ukhealthdata.org is correct
 - Provide their data access register and any other transparency info (or links to them)
- **Within Scope dataset fee.** This would cover the costs associated with providing the basic dataset level information requested above for each dataset in scope for the rapid assessment and supporting the Parity/MetadataWorks team to carry out an assessment of effort required to fulfil the metadata requirements of the MVP.
 - Differential fee between distinct datasets and derived or related datasets (where the business metadata is essentially the same as another in scope dataset).

Phase 2: Targeted support to metadata onboarding

- One of the outputs of the rapid assessment will be an estimate of the work required to onboard the required metadata (business, dataset and field level/technical) onto the Metadata Catalogue, and hence be available for discovery via the MVP Portal.
- This long list of potential onboarding work will be prioritised, and data custodians will be eligible for further investment to support the work to supply metadata to the catalogue within agreed timescales.
- *Payment of this additional funding would only be made on successful uploading of the metadata and an agreed sustainability plan for ensuring metadata is kept up to date without additional external investment.*

Proposed Conditions of Investment

- To be eligible for funding Alliance members must provide rapid access to relevant staff and facilitate rapid assessment and work to prepare metadata onboarding.
- This proposal is about metadata to support discovery as set out in the Alliance Letter of Intent. There would be no changes to rights or control of the underlying dataset/data asset.
- The data custodian would grant HDR UK royalty-free, unrestricted and non-exclusive rights to the current and future metadata for use in MVP and any subsequent Innovation Gateway development. These rights would be perpetual and cover use of the metadata by public, academic, commercial, not-for-profit, charitable and voluntary sector entities whether based or accessing the metadata in the UK or overseas. (i.e., HDR UK would hold a metadata asset at the end of the process).
- The data custodian commits to keep the metadata subsequently uploaded to the catalogue up to date.
- The decision on what represents a distinct dataset would rest with HDR UK and any attempts to 'game' the approach would be considered contrary to the Alliance Letter of Intent and Principles for Participation.
- This is a one-time investment to support development and testing of MVP. It does not set a precedent for any future decisions or funding of metadata provision. No further investment should be expected, and organisations will be required to provide a sustainability plan to keep the metadata up to date and onboard metadata of future datasets that does not assume any further financial support from HDR UK.
- No payment would be available for datasets already in Health Data Finder unless there is a clear and agreed need for improvement.

Proposed milestones

- By Thursday 7th November – Alliance members provided with “Within Scope Dataset return” template
- Thursday 14th November [TBC] – Webinar on Metadata onboarding process for hubs and Alliance members
- By Monday 18th November - “Within Scope Dataset return” latest submission date - provision of summary spreadsheet of data assets that are within scope for the prioritisation, basic access request information and any release register data.
- By Monday 25th November – Initial Rapid assessment output – based on in scope dataset return. For ‘ready’ datasets this is expected to include metadata for uploading. For other datasets, an estimate of the work required to achieve sustainable onboarding.
- By 20th December 2019 – Metadata provided for MVP launch on 10 January – first tranche of prioritisation. Payment based on estimate of additional effort required developed as part of rapid assessment unless significant unforeseen additional costs/work encountered and documented.
- By 31st March 2020 – Metadata provided by 20 March 2020 – second tranche of prioritisation. Payment based on estimate of additional effort required developed as part of rapid assessment unless significant unforeseen additional costs/work encountered.

Metadata requirements – overview of levels

1. Summary Metadata (Mandatory) – **In scope dataset return level**
2. Business Metadata
 - Required
 - Recommended
3. Technical Metadata (variable level metadata i.e. data elements, types etc.)
 - Required
 - Recommended
4. Enhanced Metadata (not in scope for initial on boarding)
 - Profiling
 - Quality
 - Other

Metadata requirements – placeholder for Adam

1. Further information – e.g., fields for level 1



Questions?

Thank you

Find out more:

Health Data Research UK

Web: hdruk.ac.uk

UK Health Data Research Alliance

Web: ukhealthdata.org

Social: @HDR_UK

Email: enquiries@hdruk.ac.uk