Data Access & Discovery
- HDR Innovation Gateway
Interactive Feedback Workshop

11:00 - 12:00
Thursday, 11 August 2022

#HealthDataGateway
1 and 2. Welcome and Introduction

3. Please share your name, organisation, job title and how you kept cool during July’s record-breaking heatwave.

4. Please share a few keywords that best describe your area of research or field of expertise.

5. What web-based portals do you use most often to support your research?

6. What is the biggest challenge you face when searching for and accessing the right data for your research?

7. What interests you the most about the Gateway as a platform to discover and request access to health datasets?

8. What would encourage you to use the Gateway more? And why?

9. Thank you for participating!

Click here to view the Jam Board template
3. Please share your name, organisation, job title and how you kept cool during July's record-breaking heatwave.

4. Please share a few keywords that best describe your area of research or field of expertise.

You represented:
- Research (academic and clinical)
- The public
- Research support and professional services
- Private companies (health research related)
- HDR UK staff

You had expertise in:
- Clinical health
- Life sciences
- Data science
- Other (cancer, imaging, mental health, AI)
- HDR UK staff (marketing, technology, project management)
The most common categories mentioned were:
- Search engines (for publications mostly)
- Professional networks and collaborative workspaces
- Best practice resources
- Databases and repositories
- Dataset specific (like the Gateway)
- Cloud and computing services

The most common challenges related to:
- Data information
- Data access (approval process)
- Data availability
- Latest research
- Data discovery platforms
- User experience
5. What web-based portals do you use most often to support your research?

- Various public dataset resources
- Combined Intelligence Population Health Action - population health management platform
- World Bank (for data?) https://datacatalog.worldbank.org/home
- NHS community currencies
- NHS Digital Data Access Request Service (DARs)
- One in Million dataset from Bristol University: http://www.bristol.ac.uk/primaryhealthcare/researchthemes/one-in-a-million/

**Funding and Project Progress Portals**
- NIHR, UKRI (for funding, grant, project applications and updates)
- Health Foundation
- Gavi, the Vaccine Alliance
- ResearchGate
- The Global Health Network Hub (professional network and knowledge sharing hub)
- REDCap for designing research databases.

**Professional Networks and Collaborative Workspaces**
- Kaggle (online community platform for data scientists and machine learning enthusiasts, Google subsidiary)
- Synapse

**Search Engines (for pubs)**
- Imperial College London Library
- Google Scholar
- PubMed
- NCBI
- BMJ

**Databases and Repositories**
- Ensembl genome database
- ISRCTN - primary clinical trial registry, BMC SN (searching for existing/ongoing research)
- BioPortal (an open repository of biomedical ontologies)
- Cochrane (collection of databases)
- Clinical Record Interactive Search (CRIS) database

**Best Practice Resources**
- Clinical trials toolkit by NIHR
- MAGIC App (platform for clinical evidenced, standards and guidelines)
- OHDSI/OMOP (data standardisation resource)
- NHS Data Model and Dictionary (reference point for approved Information Standards Notices)

**Cloud and Computing Services**
- Amazon Web Services
- GitHub
- Power BI - Microsoft data visualisation software
- Nvidia (accelerated computing and transformation AI, Inverted GPU)
6. What is the biggest challenge you face when searching for and accessing the right data for your research?

Data Discovery Platforms
- Too few sites for discovery, each with small/local coverage (hence use PubMed and Google)
- The need for a single search engine for all health data
- Always a snapshot, not end-to-end data.
- Google and Amazon are paid ones

Data Information
- Having the right information on how to access - REC, CAG, DARs and what order to approach each organisation
- Data quality - what data can be used as reliable source data, mandatory completion, quality checks etc
- Understanding Data flow - is the data anonymised, de-identified? Harmony on terminology and contractual terms
- Data sensitivity, patient data

User Experience
- Poor narrative on datasets
- Unintuitive interfaces
- Ease of use interface

Project design and proposal
- complex (meta) data models

Data Access
- Application process - delays in receiving data, often data received beyond grant timelines
- Understanding the scope of [approval] committees
- Accessing dataset is complex and time consuming
- Turn around time for data access approvals

Related datasets
- Local vs. national data
- Access to primary care data
- Related datasets [not] listed

Latest Research
- Getting current literature
- Understanding what research is underway
- Literature not open access
- Keeping up! Rate and pace of publication very high

Data Availability
- Access to primary care data
- UK wide data - trials recruiting across nations, which data matches
- Data not available for commercial research
Group Findings: Page 7 and 8

7. What interests you the most about the Gateway as a platform to discover and request access to health datasets?

This is what appeals to you most:
• Data discovery
• Trust and transparency
• The datasets!
• Service and support
• Open source and open access

8. What would encourage you to use the Gateway more? And why?

These are your suggested areas of improvement:
• Federation
• Search
• PPIE
• Development (focus)
• Visibility
7. What interests you the most about the Gateway as a platform to discover and request access to health datasets?

The most common categories mentioned were:

- Search engines (for publications mostly)
- Professional networks and collaborative workspaces
- Best practice resources
- Databases and repositories
- Dataset specific (like the Gateway)
- Cloud and computing services

The most common challenges related to:

- Data information
- Data access (approval process)
- Data availability
- Latest research
- Data discovery platforms
- User experience

8. What would encourage you to use the Gateway more? And why?
8. What would encourage you to use the Gateway more? And why?

- Improved search functionality, search options to help identify ongoing/complete datasets, whether a dataset is UK wide. Options/descriptions in plain English
- Very hard but some harmony with the regulators especially in terms of what is needed for public engagement
- Understanding of what elements of the Gateway are available to the public / can be used - seeing what reports are available and what can be used
- Avoid trying to do too much... data discovery, cohort discovery, use cataloguing, data access request management, federated analysis
- Need an app for for cross platform use
- More workshops on using API
- How is it tangible? Show me how it's being used
- I look forward to using it! :)
- Unclear on which options to pick when using search tools - what type of researcher etc, worried about removing data from search
- Search
- Visibility
- Development
- PPIE Engagement
- Search
- Federation

The most common categories mentioned were:
- Search engines (for publications mostly)
- Professional networks and collaborative workspaces
- Best practice resources
- Databases and repositories
- Dataset specific (like the Gateway)
- Cloud and computing services

The most common challenges related to:
- Data information
- Data access (approval process)
- Data availability
- Latest research
- Data discovery platforms
- User experience

The Group Findings: Page 7 and 8
Thank you for taking part in our Breakout Room sessions!