

UK Health Data Research Alliance

Outputs from breakout sessions and world café discussions held on 4 February 2020



Earning, building and sustaining public trust

Discussion: practical steps and methodologies for how we earn trust

1. Practical examples that already exist

- Importance of engaging with diverse communities – some good examples in the charity sector and from public health campaigns, e.g. going to football matches to talk to men about testicular cancer
- Use storytelling that is compelling for different audiences
- Rethink engagement – be proactive, encourage face-to-face not just digital
- Engage with practitioners
- Use simple language
- Engagement works well when it's done bottom up – going out to the communities – and combining this with a top down approach

2. Who can we learn from?

- Charity sector, academia, council/local authorities
- Northern Health Science Alliance
- Examples and materials from Understanding Patient Data
- UseMyData (citizen/patient driven)
- Public Health England (use of Citizen Juries)
- New partners (e.g. universities, Biomedical Research Centres, PPI groups)

3. What are the challenges?

- How do we measure impact? Impact assessment frameworks are useful, e.g. Public Involvement assessment framework (PiiAF) – Crohn's and Colitis UK and Theory of Change

- Fairness – need to demonstrate equal distribution of benefits
- Language can be a challenge in terms of we talk about data with clinicians, patients, public
- Importance of using the data citation and data provenance

4. How can we share learning and co-ordinate engagement efforts?

- Share best practice, tools and resources – connect up as a network across HDR UK and the Alliance
- HDR UK could be a convenor and provide national oversight
- Plan campaigns together and explore 'hot topics'
- Explore different communication channels
- Proactive messaging around data
- Utilise existing campaign and ensure consistency



Trusted research environments

Discussion: Trusted Research Environments (TREs), secure analytics environments for researcher to perform analysis.

1. New model proposed

- Proposed that UK HDR Alliance commits to TRE model based on the “5 Safes” to ensure ethical research and innovation use of health data
- Central principle is no distribution of individual level data with all processing carried out within the TRE
- Multiple examples of TREs operating successfully in this way

2. TRE architecture

- De-identified datasets provided within a TRE
- Request for access to carry out analysis received, reviewed, approved
- User logs into TRE to carry out analysis
- User can only request export of summary data, e.g. for publication

3. How to manage safe outcomes

- Data custodian maintains control of data; retains public trust
- Data privacy no longer depends on de-identification process
- Less costly holding and securing a single copy of the data
- Privacy controls shifts from individual to summary level

4. The role of public Cloud

- Virtualisation/cloud computing make being algorithms to data and provision of scalable compute environment practical
- Security design implemented by public cloud vendors ensures that as service providers they have no access to the data

5. Considerations

- Clear commitment to operating in this way provides a unique chance to “reset” public confidence
- To adopt this approach, there are some issues to consider (e.g. certification of TREs; operation across different cloud providers; provision and licensing of analytical tools; federation analysis across multiple TREs).



Ethics and information governance

Discussion: Health Data Research Hubs developing best practice for access to data assets.

1. Different perspectives

- There are differences between organisations depending on start point and legal basis for sharing
- We need to differentiate between existing data sharing agreements and new ones

2. Legal frameworks and lessons learned

- Identifying the set of relationships that need to be covered by legal agreements
- Defining activities under the control of the “Hub Data Controller(s)” and flow of data to/from other Data Controllers and Data Processors (e.g. NHS Trusts)
- Identifying Legal Basis and exemptions
- Assurances around information security
- Transparency to data subjects

3. Joint data controllers

- Co-owned/authors of way data is processed
- Depending on research purpose linkage might be available
- Negotiation between entities
- Data Protection Impact Assessment (DPIA) could be written together
- Possible challenges: lack of common data sharing agreement and common risk tolerance, achieving joint approvals for linkage requests, issues of cost-recovery model as barrier to sustainability

4. Standardise the way we work

- We need to develop a consistent approach to the ethical and information governance scrutiny of data access requests
- We aim to agree on best practices and standardised the way we work

5. Build a community of best practice

- We can look at good examples and learn from others
- We can provide a way to bring together relevant individuals from each of the hubs and invite other Alliance members to join.



Data quality and standards

Discussion: How organisations could improve data quality and identify the needs of data users.

1. Definitions

- Initial scoping definitions and agreement of the language used
- What is data? What is a dataset? When do standards apply? What do we mean by 'curation'?

2. Reflection on the matrix proposed

- Having a national position in this is hard and a lot of work needs to be done, but we need to consider the existing work and not duplicate efforts

3. The value of metadata

- Metadata should use standards methodologies where possible
- Metadata should be based on a user need and user story
- Metadata quality is not binary. We should talk about 'qualities'

4. Health data standards and interoperability issues

- Define what to collect
- Define data elements
- Define format of data
- How to collect, store, represent, model etc.
- Define terminologies / classifications / ontologies used
- Define semantics / phenomics

5. Use cases

- Data quality and preparation of data for re-use will vary depending on the user needs
- For different type of data (Clinical, Audit, AI) depending on the users and the applications, the data formats and specs will be different

6. Next steps

- We need to have a pragmatic focus on achievable and interoperable standards and avoiding the creation of specific additional standards
- Analysis of data quality and standards in the collection of data depend on who completes. Understanding bias would be interesting.



Commercial sustainability

Discussion: Commercial framework to support sustainability of the Health Data Research Hubs

1. Commercial framework

- Provides flexibility to allow a variety of models
- With any monetary arrangement, the value of non-financial collaboration needs to be recognised

2. The value of NHS data

- Many potential industry and academic users of data do not fully recognise the value of NHS data and costs of gathering and curating it
- Need to demonstrate ways in which data can be used, and listen to users to understand their needs/provide datasets that they will want to use

3. Transparency

- Clear language: how we talk about data with clinicians, patients, public
- Public and patient involvement/engagement is essential. Use of data has to be legal but also publicly understood and supported.



4. Commercial models

- Outcome-based commercial arrangements impose risk and complexity, e.g. if the arrangement involves an equity share

5. Equity

- There are some issues that need to be considered in commercial arrangements for data:
 - ✓ How does the NHS obtain fair value?
 - ✓ How will any value be shared so that it benefits the entire NHS and not a small group of already well-funded institutions?
 - ✓ How will value be shared across all organisations that contribute to the availability of user-ready datasets, and how will they all be involved in decision-making?

6. Main learnings

- We need to raise awareness in industry of what the value of NHS data is
- We need to understand the user needs
- We need to gain support and trust from the public; transparency is crucial
- We need to ensure shared value across the systems, institutions involved and meaningful say in decision making
- There is appetite to learn from other organisations.

World café table discussions

<p>1. Data infrastructures: How will we know when we've got it right? (Paul Jackson & Elizabeth Waind, ADR UK)</p>	<p>We can measure progress of an infrastructure's in a qualitative and progressive way, rather than in a quantitative way, using a framework of levels of maturity for defined primary goals.</p>
<p>2. Methods and evidence standards for data (Adrian Jonas, NICE)</p>	<p>The data and analytics transformation programme was generally supported. Hubs and other academics were keen to collaborate directly (or be kept updated on this).</p>
<p>3. HDRA terminology: Are we speaking the same language? (Macey Murray, UCL)</p>	<p>Understanding of key terms used in health data research can cause confusion when we communicate. Discussions led to the idea of developing a glossary page on the HDR UK website, with potential to include explanations in lay-friendly language.</p>
<p>4. What does a fair partnership for accessing NHS patient data look like? (Grace Annan-Callcott, Understanding Patient Data)</p>	<p>Participants discussed how the findings from recent research by Understanding Patient Data related to the work they do. Key is that benefits to patients are prioritised and shared across systems.</p>
<p>5. How can we become more efficient providers of health data for researchers? A regulator's challenge (Amanda Evans & Stevan Cirkovic, HFEA)</p>	<p>Data should be provided at various levels of identifiability to researchers, ensuring panels approving projects have the right expertise to review the research proposals, and options for providing data to researchers to ensure security.</p>
<p>6. How do we evaluate the impact of using healthcare data to support safer prescribing? (Arlene Gallagher, CPRD)</p>	<p>The reasons why we measure impact, qualitative vs quantitative measures, current metrics used and how much impact we expect to see are important considerations when evaluating use of healthcare data.</p>
<p>7. Data + Human Tissue = Impact (Katy Glazer & Amir Gander, UKCRC Tissue Directory & Coordination Centre)</p>	<p>Consent discussions focused on interpretations of consent between different organisations. Impact discussions focused ensuring data holders and tissue banks are acknowledged in outputs of the research process</p>
<p>8. Health Data and Industry – opportunities for health data to drive research and innovation in the SME sector (Joanne Hartley, MDC)</p>	<p>There are perceived barrier of costs or data access and the need for industry to recognise the value of quality health data. The Hubs and the Gateway can help industry understand what can be achieved.</p>
<p>9. Open science, GitHub, and open approaches to health data research (Caroline Morton & Ben Goldacre, EBM Datalab, University of Oxford)</p>	<p>There are barriers and opportunities around open analytic methods. Need for easy access to training on methods (e.g. Jupyter notebooks, Docker, and GitHub), and reassurance that sharing code and methods does not put people at a competitive disadvantage.</p>
<p>10. Standardising national data access for health tech product development e.g. machine learning components (Deborah Wake, MyWay Digital Health)</p>	<p>The voice of the business and innovation community should be included in privacy and public benefits/ethics panels/decisions. An open free model for all could be considered.</p>

World café table discussions

<p>11. How can NHS DigiTrial deliver the maximum impact for customers and Alliance members (Heather Pinches & Charlie Hooper, NHS Digital/IBM)</p>	<p>Discussion about outputs of user research and how this is being used to inform the design of the service. Additional contacts with expertise, use cases and good ideas for refining the features of the service were suggested.</p>
<p>12. Real-World Evidence: Current barriers and future wishlists (Chris Russell, Discover-NOW – The Health Data Research Hub for Real World Evidence)</p>	<p>Opportunities and challenges were discussed: e.g. ability to use RWE for economic analysis and measure impact; issue of reproducibility using RWE as dataset; ability to work with charities.</p>
<p>13. Knowledge based metadata to drive automation and interoperability (Luke Smith, Databiology Ltd.)</p>	<p>The myriad of data models and diverse range of apps & pipelines make it hard to define standards in biomedical data processing. Automations can help people down the right paths, but it's important to not limit their approaches or options.</p>
<p>14. How can Alliance members work with DATA-CAN? (Charlie Davie & James Peach, DATA-CAN – The Health Data Research Hub for Cancer)</p>	<p>Finding genomic data, linking cancer data to other disease data, partnering with other Hubs (Pioneer) to examine impact of rapid diagnostics centres on cancer diagnosis are among potential opportunities.</p>
<p>15. How do we assess, capture and report the non-financial value of Hubs? (Monica Fletcher, BREATHE – The Health Data Research Hub for Respiratory Health)</p>	<p>The general conclusion was that value could more easily be defined and assessed to the NHS, industry, academia, but it is more complex defining and articulating benefit and value to; patients, the public and society as a whole.</p>
<p>16. Can we, should we, use the public cloud for analysis of health data? (Neil Walker, NIHR Bioresource)</p>	<p>Main takeaways were: be transparent; find good analogies – e.g. data is secure; show some badges/kitemarks – but only if they have any recognition; clarify language used; let researchers know their research is secure too.</p>
<p>17. Use of FHIR for data science and app delivery to the clinic (David Sibbald & Rodrigo Barnes, Ardihia)</p>	<p>Participants shared experiences of organising data and exposing it using FHIR standards and how this made noticeable differences to patients' care and clinical observations.</p>
<p>18. HDR UK metadata standard: future progress Ashutosh Tripathi & Adam Milward, Parity/MetadataWorks</p>	<p>Suggestions for the platform included 'non-technical' guidance on metadata specification, default values and constraints to choose from. New features well received: automated metadata generation, DOI mining, commercial frameworks for calculating data value and quality scores.</p>
<p>19. What are the skills and credentials required to be a user of data? (George Moulton, HDR UK)</p>	<p>What does constitute an 'accredited user of data'? Some observations discussed: needs vs bureaucracy; different organisations, different access models; who is responsible for accreditation and/or training.</p>

Next steps

- The next Digital Innovation Hub programme Symposium is scheduled for November 2020, in Manchester – further details to follow.
- We are producing a draft data quality standards' [green paper](#) and a final version which will be published soon.
- We are developing a green paper on future options for UK Health Data TREs for Alliance members to consider.

Useful reading

- Breakout session pre-read: Trusted Research Environments
- Breakout session pre-read: Ethics and information governance
- Breakout session pre-read: Data quality
- Breakout session pre-read: Commercial sustainability



Thank you for your contribution!

