closer

The home of longitudinal research

Building a consensus on both technical & conceptual interoperability

Jon Johnson

16 January 2020

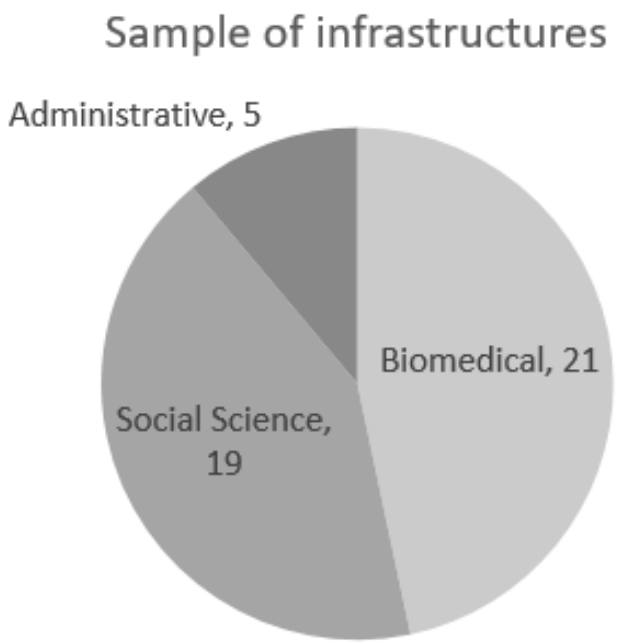
CLOSER, UCL Institute of Education

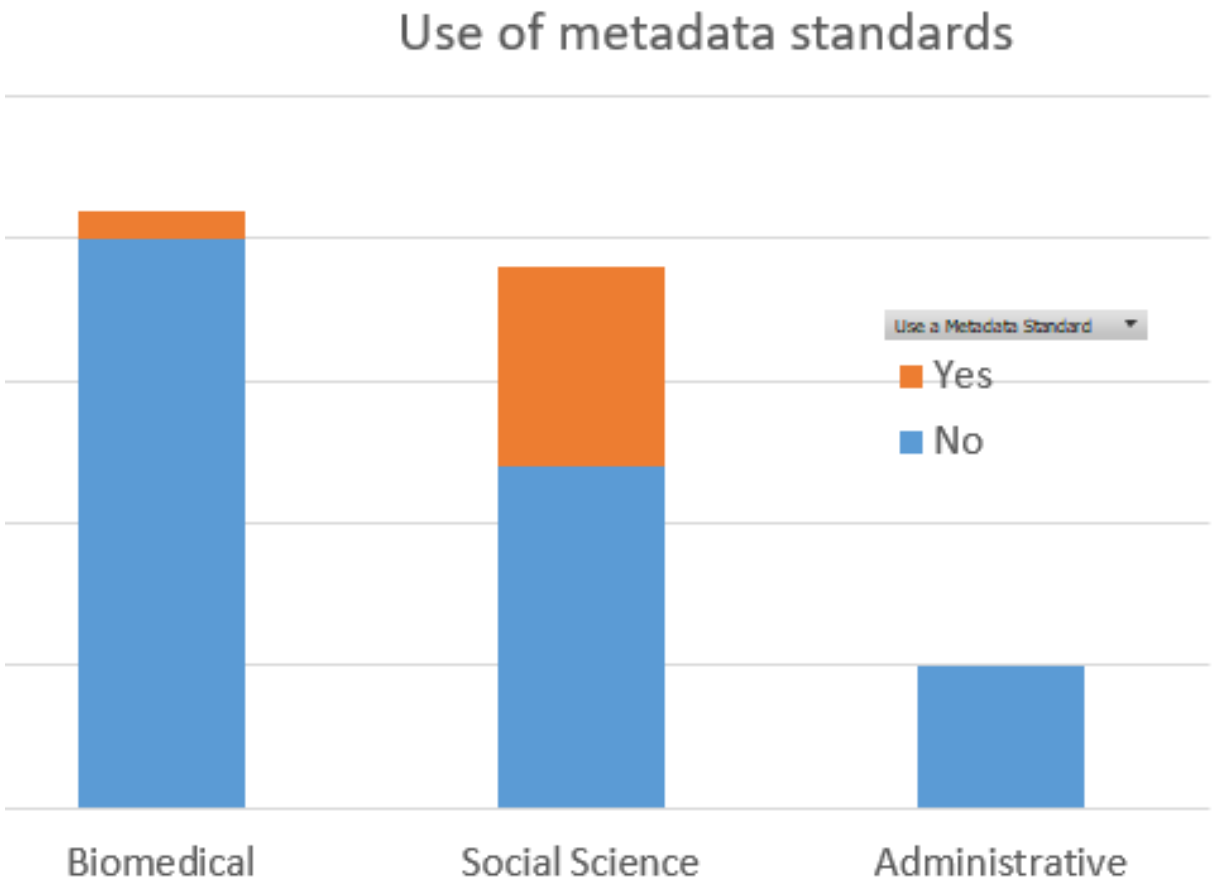
UKRI Economic and Social Research Council

Overview

- Metadata landscape and adoption of standards
- Drivers for and what interoperability means
- Technical and conceptual interoperability
- What models are in use
- Thoughts on moving forward

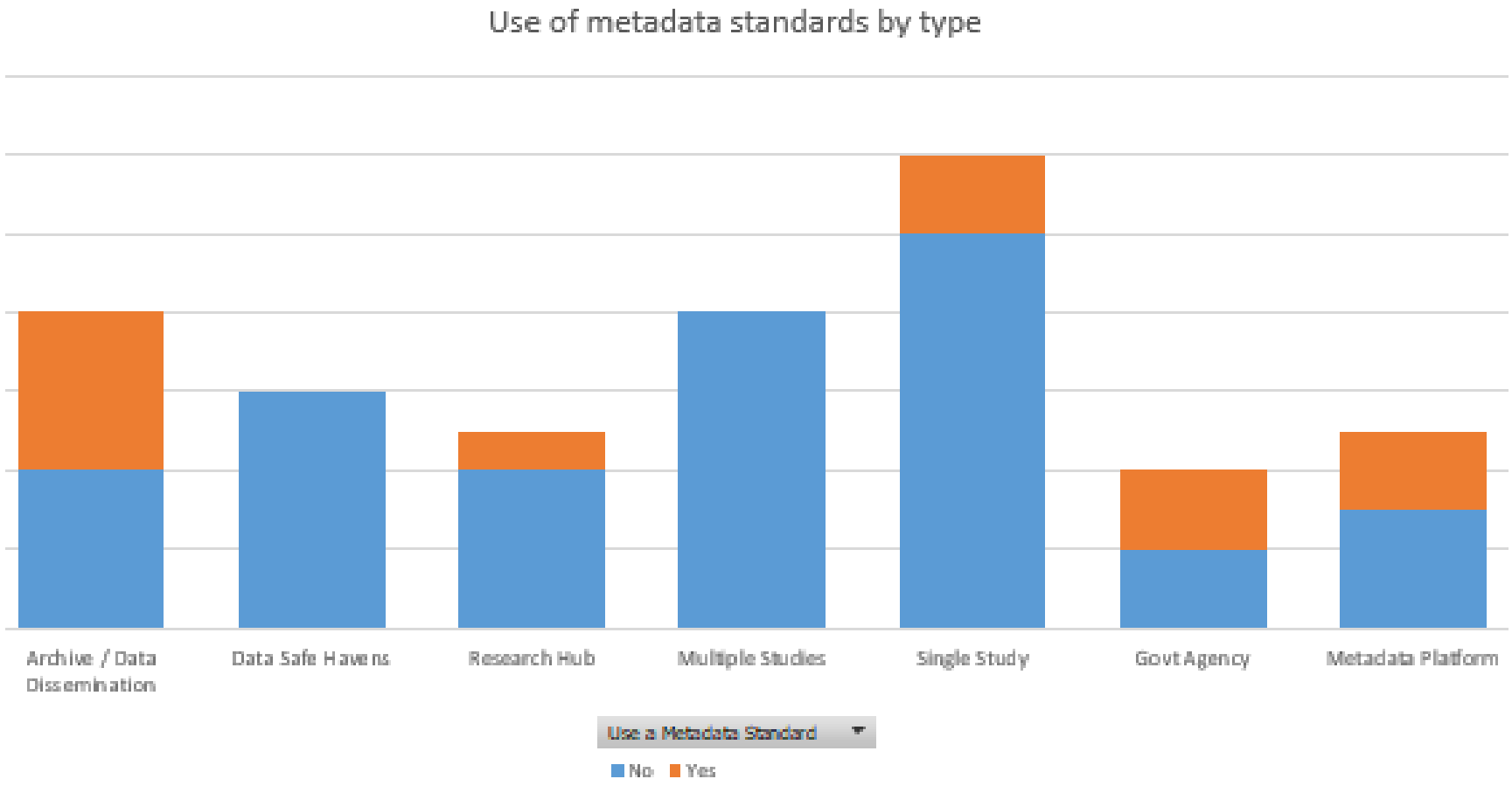
Longitudinal Metadata Landscape (1)





Johnson, Park, Corti (2018) Longitudinal Data Landscape Review for ESRC unpublished

Longitudinal Metadata Landscape (1)



Johnson, Park, Corti (2018) Longitudinal Data Landscape Review for ESRC unpublished

Drivers for interoperability

- Funders are increasingly interested in projects which combine data from different sources and across disciplines to create new understanding and address policy objectives
- Data preparation and integration efforts are hindered by data quality, complexity and access issues*
- Disciplines which haven’t had a culture of data sharing are now playing catch up and current practices do not scale well
- The FAIR initiative is one mechanism that seeks to redress some of this information gap.

*https://wellcome.ac.uk/news/new-programme-explore-how-innovation-health-data-can-benefit-everyone

Interoperability (FAIR)*

Data usually need to be integrated with other data. In addition, the data need to interoperate with applications or workflows for analysis, storage, and processing.

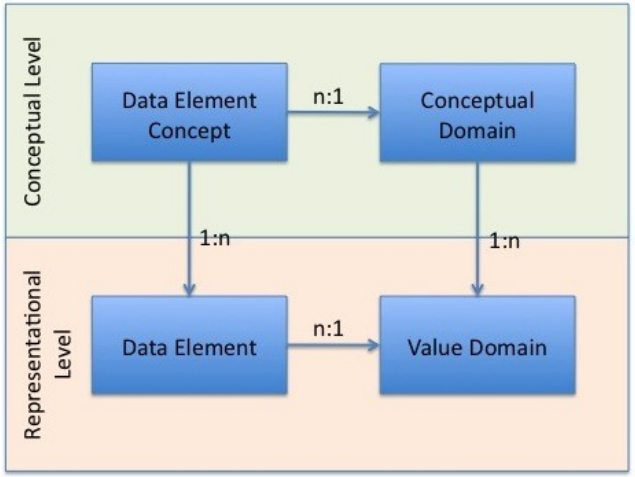
- (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation. e.g. RDF, XML etc
- (Meta)data use vocabularies that follow FAIR principles, e.g. HASSSET, MeSH
- (Meta)data include qualified references to other (meta)data, e.g. have persistent identifiers

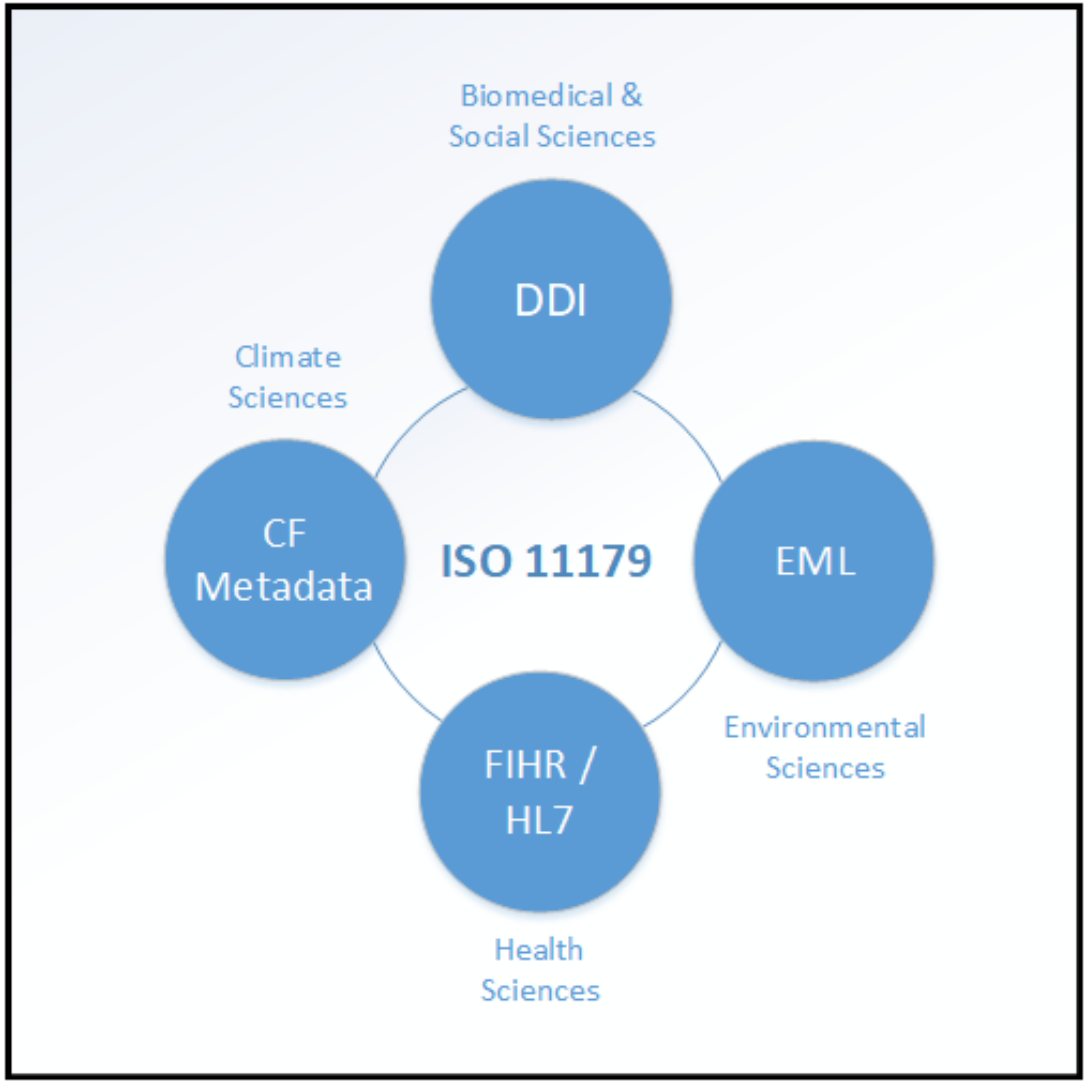
* sorry

Technical Interoperability

Much of the ground work exists.

- Metadata standards that are ISO 11179 compliant or compatible in some way cover a wide range of potential data one would be interested in combining.
- They implement (or map to) ISO 11179 and extend it for their own domain.
- There are substantial portions of these standards that are interoperable, in that they describe data fundamentally the same.

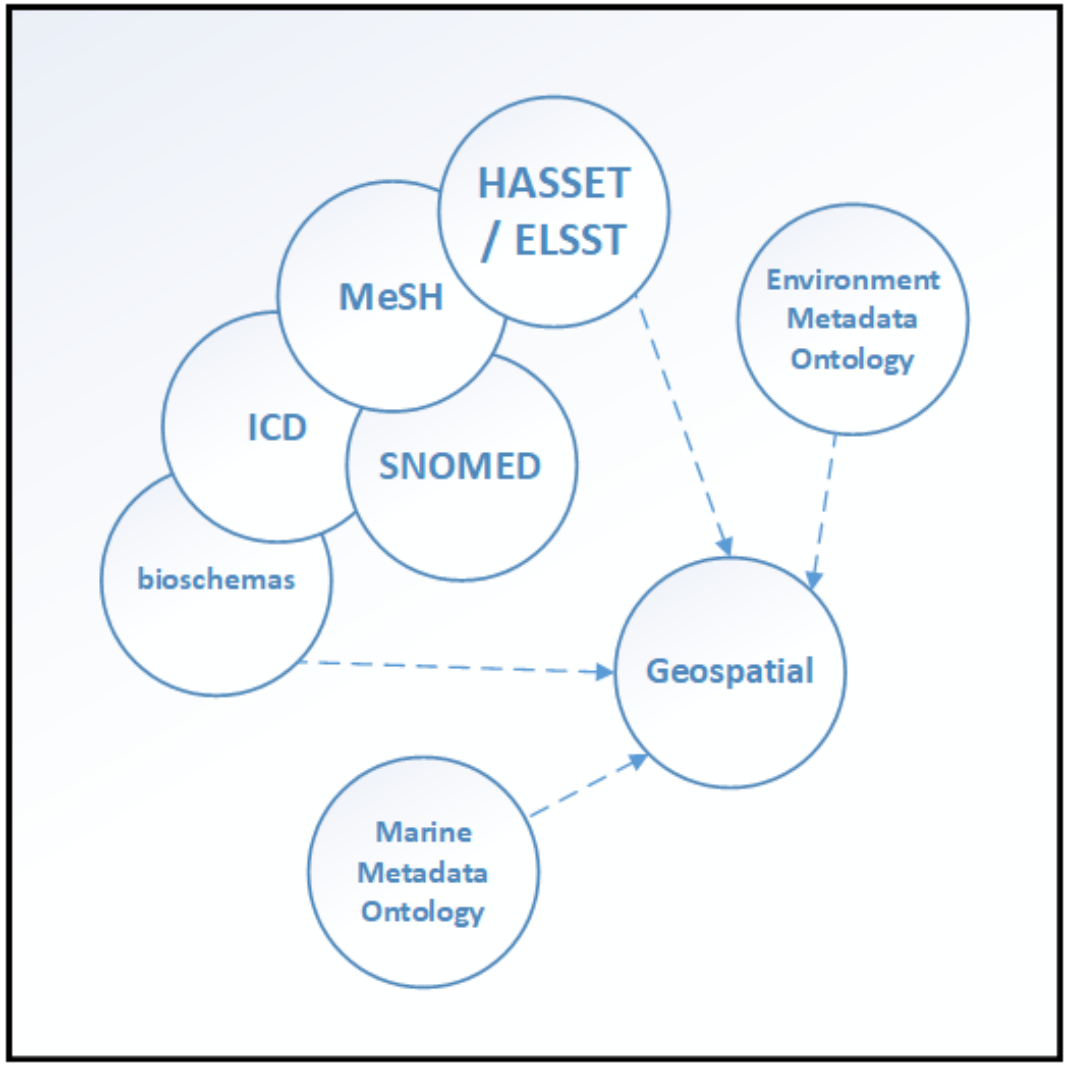




Conceptual Interoperability

Here also, much of the groundwork exists

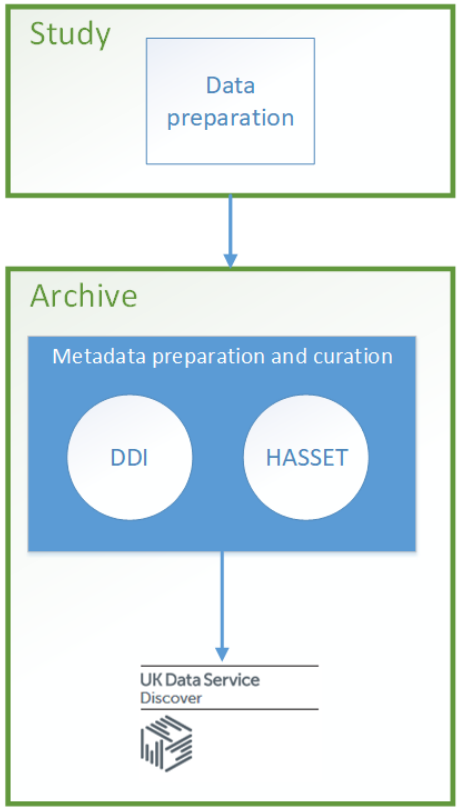
- There are already a number of standard vocabularies which exist, and have some overlap with each other.
- They are well established and mature standards which are used in the larger infrastructures and in some cases more widely adopted.
- In many cases there is good tooling to support them, but they are not always accessible to researchers to embed them as part of their day to day work.

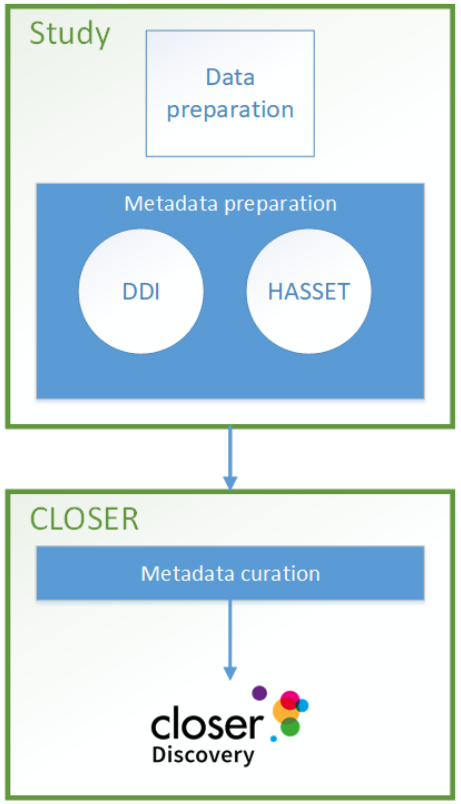


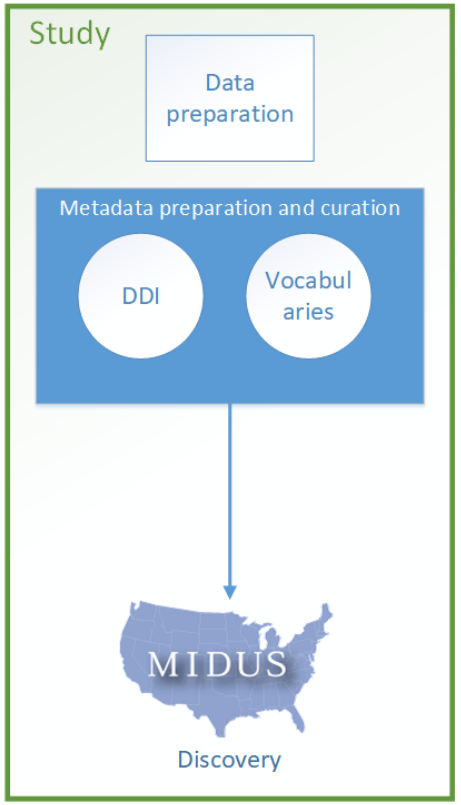
Building sustainable metadata for discovery

Creating consistent metadata across a discipline requires an agreement on both the technical and the conceptual standards for discovery to be meaningful.

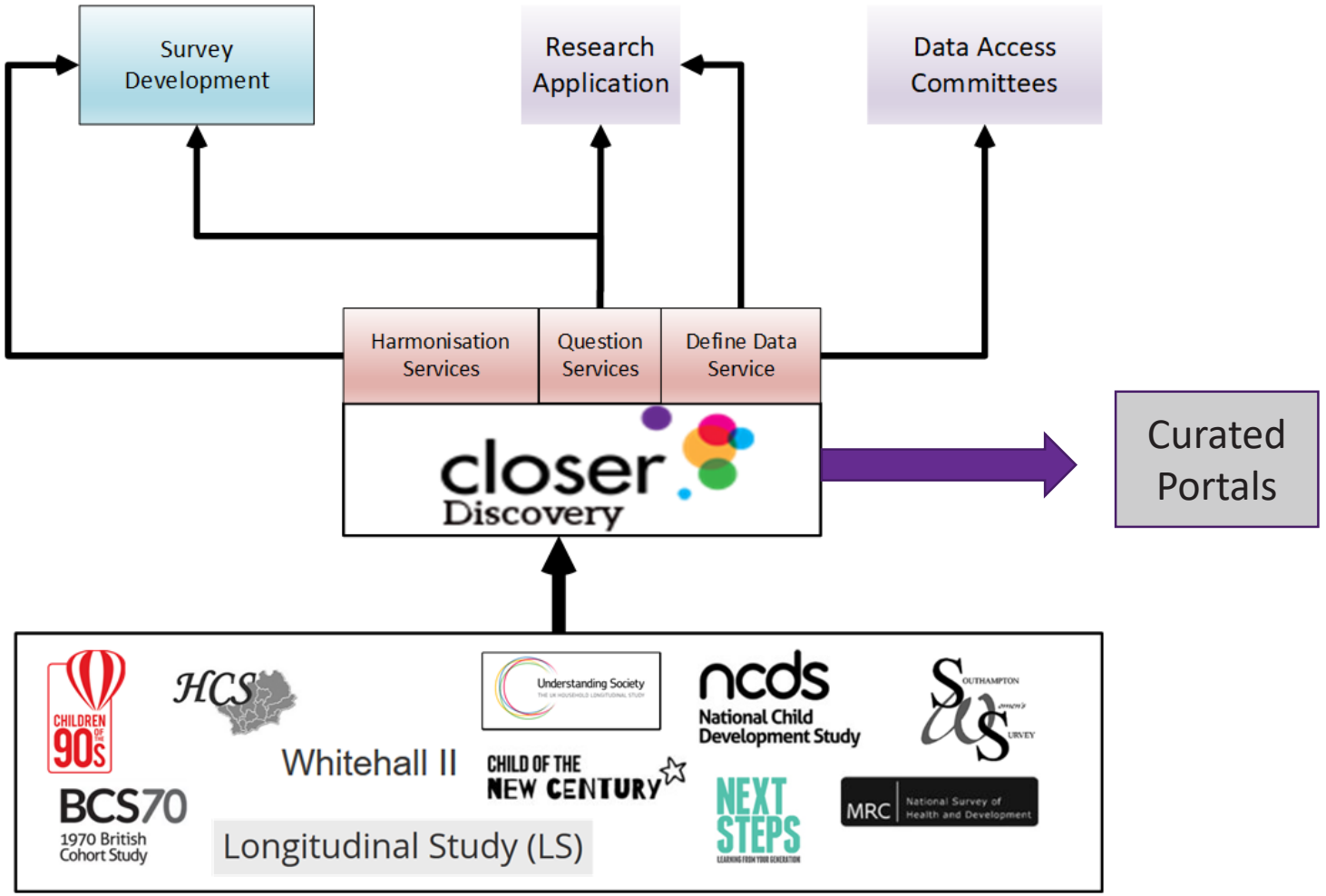
- The UKDS example, also applies to other archives, in social sciences, but also other disciplines have a similar approach such as CEDA.
- The model exemplified by the CLOSER example is likewise similar to that used by ELIXIR.
- There are some innovative studies that are large enough, (or invested enough) to be implementing standards independently.







Interoperability is not just about discovery



Moving forward

A metadata strategy for social and biomedical sciences

- Content generation
  - Tooling and cultural changes
- Content delivery
  - Infrastructure and technology to disseminate
- Aligning on interoperable standards
  - Technical
  - Conceptual

Such a strategy will need to incentivise and resource these activities

Questions

Web: closer.ac.uk

Email: closer@ucl.ac.uk

Twitter: @CLOSER_UK

CLOSER Discovery: discovery.closer.ac.uk