Data Context Model for FAIR Health Data

The who, what, why, where, when and how

Esmond Urwin[1], Tim Beck[1], Antonella Demestri[2], Armando Mendez-Villalon[1], Gordon Milligan[3], Andrew Burton[4], Alexander Davydov[5], Bob Laramee[6], Andy Rae[1], Stefanie Thust[7,8,9], Guruprasad Aithal[10], Philip Quinlan[1]

[1] Centre for Health Informatics, NIHR Nottingham Biomedical Research Centre, Faculty of Medicine and Health Sciences, University of Nottingham, UK.
[2] Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, UK.
[3] Health Informatics Centre, School of Medicine, University of Dundee, UK.
[4] School of Science and Technology, Nottingham Trent University, UK
[5] Thesaurus Health.
[6] School of Computer Science, University of Nottingham, UK.
[7] Nottingham University Hospitals NHS Trust, Nottingham, UK.
[8] Magnetic Resonance and Precision Imaging, NIHR Nottingham Biomedical Research Centre, University of Nottingham, UK.
[9] Department of Translational Neuroscience and Stroke, UCL Institute of Neurology, London, UK.
[10] Nottingham Digestive Disease Centre, NIHR Nottingham Biomedical Research Centre, Faculty of Medicine and Health Sciences, University of Nottingham, UK.

Background

The need for rapid analysis and assessment of health data at scale is growing, highlighting the importance of and challenges associated with Findable, Accessible, Interoperable and Reuseable (FAIR) data:

- Health data is heterogeneous.
- Data is stored using different systems, formats and standards.
- Data is encoded with variable levels of precision.
- Variable encoding can lead to imprecise meaning.

The Observation Medical Outcomes Partnership (OMOP) Common Data Model is being used to surmount these issues. However, data conversion can be lossy, imprecise and data granularity may not be translated fully. Hence, data could lose its meaning and value.

Method - Workshop

A workshop entitled 'How to be FAIR with data standards' was run in November 2024. The proposition posed was '**All OMOP datasets are unreliable**' - currently there is no way to prove whether or not they are reliable, to which:

- Set activities elicited perspectives, thoughts and comments.
- Debates developed understanding.
- Grounded theory coding was used to model the collected data.
- Workshop was repeated at the ELIXIR All-Hands Meeting 2025.
- Workshop questions were asked again at the Public Precision Health Asia Conference 2025.

Results

The workshop produced a rich set of perspectives focused upon current data transformation and annotation challenges. From these, a Data Context Model has been developed to represent attributes that can affect decision making and the provenance of data conversion.

The Data Context Model (see Figure 1):

- Consists of 10 concepts, representing -
  - ***Explicit Factors*** – Source Dataset, Target OMOP Dataset, Standard.
  - ***Derived Factors*** – Transformation, Transformation Rule, Quality and Provenance.
  - ***Implicit Factors*** - Context, Actor, and Training.

The results from the repeated workshop and re-asked questions support the initial workshop findings and derived Data Context Model. These initial findings and model illustrate the common themes and challenges, shared across international communities and between differing contexts, those of healthcare and the life sciences.

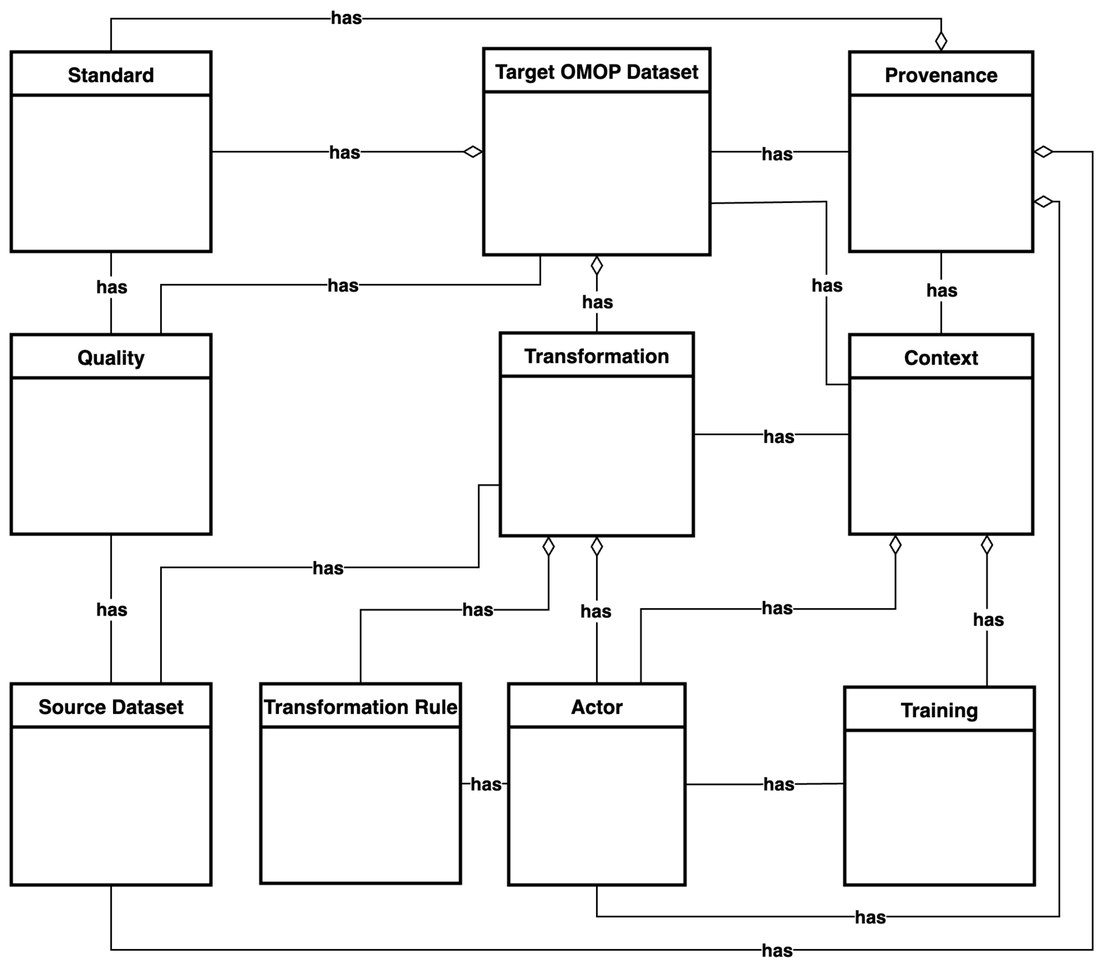

Figure 1: Data Context Model – Entity Relationship Diagram.

Conclusion

Currently it is difficult to assess the quality and reliability of data transformations. Such work is fraught with nuance, ambiguity and subjectivity. Pair this with controlled vocabularies that can change over time and what ensues is a matrix of interacting factors, each of which suffers from differing levels of variability.

The Data Context Model seeks to make explicit the factors surrounding time consuming complex data transformations. There is a need to understand how better to transform data in an ever-efficient manner by providing methods and systems to support this. The growth in federated data and analysis approaches underlines the need to create reliable OMOP datasets.

Further work will seek to:

- setup a Data Context Model study to explore context, its potential value and uses;
- create use cases to further develop, test and validate the Data Context Model;
- seek to develop methods to represent data transformation context to aid dataset reliability.

Partners & Sponsors

Many thanks to our sponsors and partners for their help and support.

HDRUK Health Data Research UK

NHS Nottingham University Hospitals NHS Trust

NIHR | Nottingham Biomedical Research Centre