quality, interoperability, and accountability

Jim Davies and James Welch

NIHR | Oxford Biomedical Research Centre

BIG DATA INSTITUTE

UNIVERSITY OF OXFORD

big data

- ‘big data’ presents challenges not only in terms of size and complexity, but also in terms of quality
- quality is a relative notion: is the data good enough for this purpose? can we use the data to answer this question? do we have sufficient confidence in our answer? can we provide sufficient assurance?

big data

- ‘big data’ is collected in one context – or many contexts – and analysed in another
- those who do the analysis may not have a sufficient understanding of the context of collection, or of any subsequent processing
  - and if they achieve an understanding, they may not record or share that understanding

problem

- to establish data quality and guarantee interoperability we need to capture this understanding: we need to describe and compare contexts
- to do this at scale, we need a high degree of automation
- as human input is required, we need also effective mechanisms for collaboration: for sharing, re-using, extending, and updating descriptions

solution

- technology for generating, annotating, sharing, re-using, and comparing descriptions of context
  - ‘a data dictionary that works at scale’
- used for defining and maintaining data standards, for capturing information to support analysis and re-use, and for determining suitability and interoperability
- originally the ‘Oxford metadata catalogue’, now an open source project: Mauro Data Mapper

existing applications

- NHS data dictionary and model service
  - health and social care data standards
- 100,000 Genomes Project
  - data collection and quality assurance
- NIHR Health Informatics Collaborative
  - semantic interoperability of health data
- SAIL Databank, HDR UK, DataLoch
  - data documentation and discovery

NHS Digital

epcc

HDRUK Health Data Research UK

CONNECTED HEALTH CITIES

SAIL DATABANK

INDUSTRIAL STRATEGY | UK Research and Innovation

NIHR | Health Informatics Collaborative

NIHR | Oxford Biomedical Research Centre

BIG DATA INSTITUTE

UNIVERSITY OF OXFORD

current developments

- extending the technology to support federated data description and data analysis across multiple organisations and trusted research environments
- implementing support for policy-based management of health data, including medical imaging
- adding digital signatures to descriptions, including descriptions of workflows and data transformations, to support data validation and accountability