FAIR DH data need ontology standards

Ludger Jansen
University of Rostock
ludger.jansen@uni-rostock.de

FAIR Principles

- Findable
- Accessible
- Interoperable
- Reusable

Spell out the rules …

- Findable
  The first step in (re)using data is to find them. Metadata and data should be easy to find for both humans and computers. Machine-readable metadata are essential for automatic discovery of datasets and services, so this is an essential component of the FAIRification process.
- F1. (Meta)data are assigned a globally unique and persistent identifier
- F2. Data are described with rich metadata (defined by R1 below)
- F3. Metadata clearly and explicitly include the identifier of the data they describe
- F4. (Meta)data are registered or indexed in a searchable resource
- Accessible
  Once the user finds the required data, she/he needs to know how can they be accessed, possibly including authentication and authorisation.
- A1. (Meta)data are retrievable by their identifier using a standardised communications protocol
- A1.1 The protocol is open, free, and universally implementable
- A1.2 The protocol allows for an authentication and authorisation procedure, where necessary
- A2. Metadata are accessible, even when the data are no longer available
- Interoperable
  The 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.
- I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (Meta)data use vocabularies that follow FAIR principles
- I3. (Meta)data include qualified references to other (meta)data
- Reusable
  The ultimate goal of FAIR is to optimise the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings.
- R1. Meta(data) are richly described with a plurality of accurate and relevant attributes
- R1.1. (Meta)data are released with a clear and accessible data usage license
- R1.2. (Meta)data are associated with detailed provenance
- R1.3. (Meta)data meet domain-relevant community standards

… or capture the spirit

- Data need persistent storage
- Researchers need rights and permissions to use data
- Researchers need to understand the data
- Data should allow for interoperability

- IT
- Licences
- Ontologies

Ontologies

What are ontologies good for?

- Organisation of massive amounts of data
- Integration of data from diverse sources
- Coping with a highly complex domain
- Support for automatic reasoning
- Increase of searchability

Not all ontologies are equal

Reference ontologies

- Generic top-level ontologies OR taylored for a certain domain
- Benchmarked by reality
- Community standards

Application ontologies

- Taylored for a certain application
- Often not semantically adequate: „it works“ > „it is true“
- Idiosyncratic

Only reference ontologies solve the data silo problem and sustain interoperability!

- Shared community-based standards
- Strict user-understandable semantics
- Formal computer-processable characterisation

What are reference ontologies good for?

- Organisation of massive amounts of data
- Integration of data from diverse sources
- Coping with a highly complex domain
- Support for automatic reasoning
- Increase of searchability
- By providing a rigid data semantics
  - shared across domains and applications
  - allowing for interoperability

Ontologies in DH: The beginnings

- “Big” standards:
  - CIDOC Conceptual Reference Model (CIDOC CRM; ISO 21127)
  - RDA Resource Description and Access (Lazarinis 2015)

- Small domain ontologies (Philosophy):
  - InPhO Indiana Philosophy Ontology (Buckner, Niepert & Allen 2010)
  - PhilOnto (Grenon & Smith 2011)

Ontologies in DH textbooks

- No mention: Kurz 2015
- Mention in passing only: Fiormonte, Numerico & Tomasi 2015, 154.
- Dealt with, but not up to current standards from the lifescience perspective: Rehbein et al. 2017, 162-178.

Ontologies in the Life Sciences

- Open Biological and Biomedical Ontologies (OBO) Foundry
  - www.obofoundry.org
  - Repository for ontologies
  - Community tools for devolopers
  - Peer review
  - Gold practise rules
- Top-domain ontology BioTop
- Good Ontology Design (GoodOD)

About ▾ Principles ▾ Ontologies ▾ Citation ▾ Participate ▾ FAQ ▾ Legacy ▾ Search Ontobee Submit

The OBO Foundry

The Open Biological and Biomedical Ontology (OBO) Foundry is a collective of ontology developers that are committed to collaboration and adherence to shared principles. The mission of the OBO Foundry is to develop a family of interoperable ontologies that are both logically well-formed and scientifically accurate. To achieve this, OBO Foundry participants voluntarily adhere to and contribute to the development of an evolving set of principles including open use, collaborative development, non-overlapping and strictly-scoped content, and common syntax and relations, based on ontology models that work well, such as the Gene Ontology (GO).

The OBO Foundry is overseen by an Operations Committee with Editorial, Technical and Outreach working groups. The processes of the Editorial working group are modelled on the journal refereeing process. A complete treatment of the OBO Foundry is given in "The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration".

On this site you will find a table of ontologies, available in several formats, with details for each, and documentation on OBO Principles.

You can contribute to this site using GitHub OBOFoundry/OBOFoundry.github.io or get in touch with us at obo-discuss@sourceforge.net.

Download table as: [YAML | JSON-LD | RDF/Turtle]

bfo	Basic Formal Ontology CC BY	The upper level ontology upon which OBO Foundry ontologies are built. Detail	
chebi	Chemical Entities of Biological Interest CC BY	A structured classification of molecular entities of biological interest focusing on 'small' chemical compounds. Detail	
doid	Human Disease Ontology PUBLICDOMAIN	An ontology for describing the classification of human diseases organized by etiology. Detail	
go	Gene Ontology CC BY	An ontology for describing the function of genes and gene products Detail	
obi	Ontology for Biomedical Investigations CC BY	An integrated ontology for the description of life-science and clinical investigations Detail	

Gold practise rules

- Community-based development
- Rigorous semantics
- Orthogonality
- Hierarchical organisation
- Common top-level ontology (BFO = Basic Formal Ontology)
- Common set of formal relations (RO = Relation Ontology)

Towards Application: How to Re-Use

- Find ontologies
  - Repositories like the OBO Foundry
  - Ontology Lookup Service (OLS; maintained by EBI)
  - OntoBee
- Find terms
  - Ontology Lookup Service
  - OntoBee
- Integrate terms:
  - MIREOT: the Minimum Information to Reference an External Ontology Term
  - OntoFox

The Future for DH Ontologies?

- Open Platform for Socio-Cultural Ontologies
- Mid-level ontologies for the socio-cultural domain („SocioTop“)
- Integration of existing standards