

FAIR Principles

16052019: 5-7 bullet points to make data conform to FAIR principles

Gist of Paper

- **Creation of guidelines to enhance the reusability of data across variety of stakeholders**
- **Emphasis on ability of machines to find and use data**
- Public Data Archiving (i.e. data behind results) is a step in the right direction but completeness is questionable
- Should extend to algorithms, tools and workflows used to generate data
- Some standardisation already happens – think PDB identifiers
- Some collation sites already exist but don't require same format/no harmonisation
- Where is complementary data? Can your data be combined safely with past data? Automatically? Permission? License? Citation?
- Main issue will be convincing people to submit their data in the best format
- Machines biggest challenge is selecting appropriate data (alongside types formats and access mechanisms) and citing

- **To be Findable:**
- 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 it describes
- F4. (meta)data are registered or indexed in a searchable resource
- **To be Accessible:**
- A1. (meta)data are retrievable by their identifier using a standardized communications protocol
- A1.1 the protocol is open, free, and universally implementable
- A1.2 the protocol allows for an authentication and authorization procedure, where necessary
- A2. metadata are accessible, even when the data are no longer available
- **To be Interoperable:**
- 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
- **To be Reusable:**
- 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

Examples:

- Dataverse: DOI/Handles. Categories are: metadata, data files, dataset terms, waivers and licenses, version info. Metadata always public. Token to grant access for machines when data files are restricted.
- FAIRDOM: HTTP URLs can be subsequently registered for DOI's. Rich metadata annotation
- OpenPHACTS: multiple epresntations that are human and machine readable. Each call provides URL. VOID standard?
- PDB: Machine readable formats. Cross references e.eg to PubMed
- UniProt: Human and machine readable.Cross referencing.

FAIR and PyOpenWorm

- ~~• Create a submission quiz to make it easier for researchers to submit data in required format i.e. include protocols~~
- Lead by example = key point
- Follow the lead taken by DataVerse: metadata, data files, dataset terms, waivers and licenses, version info for each data source
- Create machine readable versions of all data – this is already happening?
- Dataverse token system for restricted data (are we wanting to include restricted data)

A curated, informative and educational resource on data and metadata *standards*, inter-related to *databases* and data *policies*.

- Good resource for the expected standards