Documenting your data

In order to make your data findable, accessible, interoperable, and reusable, in accordance with FAIR principles, you should document it properly.

Data documentation is formalised by generating **metadata,** that is, data which describes data. There are various metadata standards depending on the type of data processed. The main purpose of using standardised metadata is to provide people and machines with context for data produced.

The choice of the standard depends on four factors

- the type of resources described
- the scientific domain of the data
- the community concerned by data production
- the warehouse chosen to store data

The most widely used metadata standards in the HASS include :

- **Dublin Core**, a metadata standard which provides a shared base of descriptive elements which may be adapted to different types of data. It is composed of fifteen base properties (title, subject, description, source, language, relation, coverage, creator, contributor, publisher, rights, date, type, format, and identifier) which may be enriched, in which case one speaks of Qualified Dublin Core with three additional elements (audience, provenance, and rightsholder). Dublin Core is used in the Nakala data warehouse.
- **DDI** (Data Documentation Initiative) is a standard suited for data from the social, behavioural, and economic sciences. It is particularly suited to survey data and statistical information.
- **EAD** (Encoded Archival Description) is a standard used mainly in archives. It is expressed in XML language and provides a way of respecting the hierarchised organisation of research tools while conserving the principle of information inheritance at different levels.
- **TEI** (Text Encoding Initiative) is a consortium which maintains and develops a language for representing texts in digital form. It is a language for structuring textual data expressed in XML, and also provides a way of inputting metadata associated with texts.

Still, documentation should not concern solely data but also the methods employed so as to make the project methodology explicit, encourage trust in findings, and allow the data to be reused. This documentation may take the form of a simple explanatory text (a readme, txt, or markdown file) or may be included in executable statements (jupyter notebooks)

Contact
For questions about research data
guichet-ardoise@groupes.renater.fr
Fiona Edmond
Research data manager - University Library
fiona.edmond@univ-rennes2.fr
Morgane Mignon
Coordinator of the Digital Humanities platform - MSHB
morgane.mignon@mshb.fr
Further information

- Guide to writing "readme" style metadata (Cornell University)
- Fact sheet - Metadata, standards, formats