Metadata and metadata standards

In the Data Management Plan, you are asked to indicate the metadata used to describe your research data and the metadata standard you are using. This guide should help you to understand these aspects.
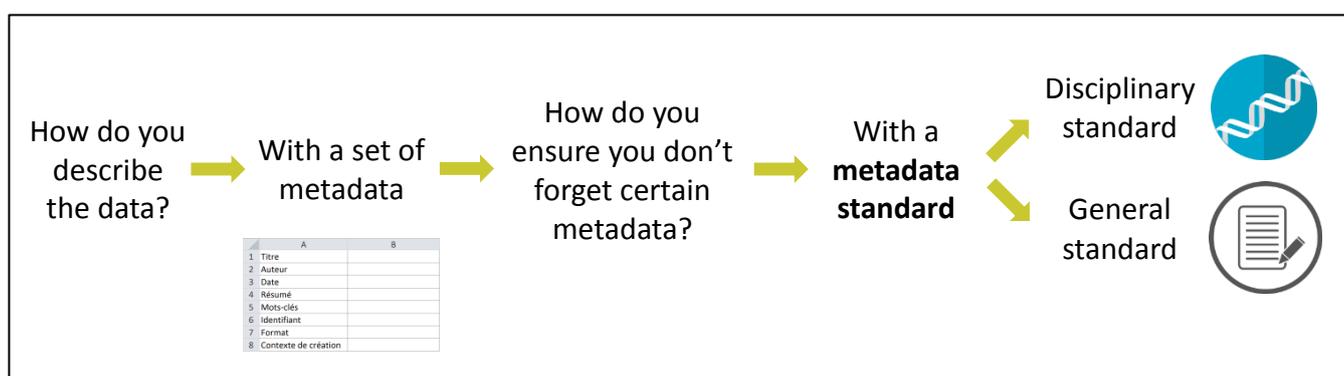
Metadata: the data describing your data

Metadata are data used to define or describe other data, regardless of the type of medium used (printed or electronic). Metadata include, for example, the title, author, abstract, keywords, identifier, format, context of creation...

Why use metadata to describe your data?

- To make your data **visible**: the richer and more accurate the description of the data, the more likely the data described are to be seen.
- To make your data **findable**: by facilitating the retrieval of the data by search engines
- To make your data **reusable**: by making it easier to understand the datasets (structure, collection method, context of creation, etc.)

Metadata standards: to help you describe your data

How do you describe the data? → With a set of metadata → How do you ensure you don't forget certain metadata? → With a **metadata standard** → Disciplinary standard / General standard

Metadata standards are templates that specify all the metadata required to describe a resource. There are two types of metadata standards:

- **General standards**: these standards often describe scientific data inadequately, but can be useful for describing administrative documents, for example.
- **Disciplinary standards**

Examples of general metadata standards:

- **Dublin Core**. This is a simple generic descriptive format, with 15 metadata elements: title, creator, subject, description, publisher, contributor, date, type, format, resource identifier, source, language, relationship, coverage, and rights management.
- **DataCite Metadata Schema.** This is a list of metadata provided by the international library consortium, DataCite. DataCite also provides a tool for generating its metadata in the XML format: DataCite Metadata Generator.

Examples of metadata standards for life sciences:

- STREGA: Strengthening the Reporting of Genetic Association Studies
- mzIdentML: mz Peptide and Protein Identification Markup Language
- CHEBI: Chemical Entities of Biological Interest
- OBI: Ontology for Biomedical Investigations
- MIRIAM: Minimal Information Required in the Annotation of Models
- OGG: Ontology of Genes and Genomes
- MCL: Microbiological Common Language
- MIATA: Minimal Information About T Cell Assays
- STROBE: Strengthening the Reporting of Observational Studies in Epidemiology
- NeuroLex: The Neuroscience Lexicon
- IDO: Infectious Disease Ontology Core

We recommend using the following directories to identify the disciplinary standard best meeting your needs:

- FAIRSharing: a directory of standards for the life sciences
- RDA Metadata Directory: a directory of standards for various disciplines

---

Why use a metadata standard?

- To describe your data **richly** and **accurately**, with the same vocabulary as the rest of your scientific community
- To make your metadata **interoperable** and to allow other systems to exploit them

---

How and when should you describe your data?

We recommend describing your data regularly during the progress of the research project, to ensure their correct management. This will save you from having to describe everything once the project is finished.

To describe your data, you can choose:

- A general standard: for administrative documents, for example
- A disciplinary standard: for scientific data
- A list of metadata in a spreadsheet: if you cannot find a standard appropriate for your needs
- The standard requested by the data repository. Indeed, repositories often suggest the standards to be used for the description of the datasets they hold. If you have already identified the data repository in which you wish to deposit your data at the time of their generation, you can simply respect the demands of the data repository concerned.

When the time comes to deposit your data in the data repository, your metadata will be ready and you can simply complete the various fields.