

Topic 1: Metadata

Information is only valuable... If we care about metadata!

Diversity of Biological Data

Being one of the most important parts of a database, metadata is basically anything that describes everything else. "Data about data", literally. All information objects have a **content**, a **context** and a **structure**, and the metadata refers to these three dimensions. The content refers to what the object is about; the context includes the who, how, where, when, etc. was the object created. The structure refers to the hierarchical or structural relations between different objects or within each object. It is crucial to **clarify everything related to the data** (temporal and spatial dimensions, but also the methodology used to access the data, units, software, etc.), but it is also relevant to know **how is this information stored** (structure of the data).

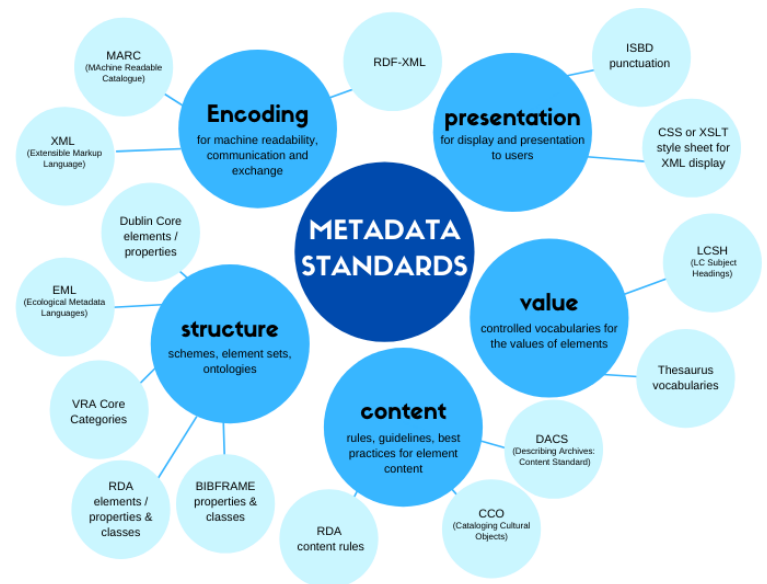
Information professionals have developed community-fostered best practices to ensure consistency and interoperability between information sources, the **Data Standards**. Those are only **voluntary consensus** that can refer to the structure of the metadata (Data structure standards), to the value (Data value standards), to the

content (Data content standards), or to the format (Data format standards) (**Figure 1**). Choosing the proper standard for the metadata will facilitate further sharing of the datasets and basic understanding of the data. (e.g. when integrating databases). Whilst there is a technical requirement for standardized metadata organization, following a standard is also appropriate to maximize the user understanding of the information. Furthermore, the chosen standard has to model the most important attributes of the type of object that the metadata describes.

WHAT CAN THE METADATA DO FOR US?

1. **Control of information objects:** It is the function of the metadata to keep track of which version of the data is being described, together with the changes and recontextualization of the specific version respect the original one.
2. **Clarity of the data:** The metadata helps to understand the terms used in the data, which can be confusing when compared with other sources of data (if the data do not follow the same standard).
3. **Validation:** Users refer to the metadata in order to ascertain the trustworthiness of the data itself. The metadata contains information about the origins and the author of the dataset.
4. **Information storage and searching:** Complete and descriptive metadata can be very useful to find any information about the data, but also to organize different datasets following a specific scheme

Figure 1. Typology of metadata. Simplified high-level overview of the different categories of metadata, depending on the organisation level that they refer to. Most of the standards specifically designed for biological sciences (Table 1) refer to the way that the data is structured, and not so much on the other dimensions of the data. This figure is adapted from Steven J. Miller, 2019.



Who creates the Data Standards?

There are several organizations from many different disciplines that design and propose data standards. Many are highly detailed standards developed by individual communities which have the potential to create extremely rich metadata. However, their creation and maintenance are so complex and resource consuming that most of the time are seldom justifiable. On the contrary, other initiatives aim to create relatively simple and small metadata elements that can be used by any community without expert data knowledge.

Biological metadata standards	Developer organization
Dublin Core Schema	Dublin Core Metadata Initiative
Darwin Core	Biodiversity Information Standards -Taxonomic Database Working Group
Genome Metadata	Bacterial Bioinformatics Resource Center (PATRIC)
Ecological Metadata Language (EML)	Ecological Society of America and associated efforts
Access to Biological Collection Data (ABCD)	Biodiversity Information Standards -Taxonomic Database Working Group
Open Microscopy Environment XML	Open Microscopy Environment Consortium

Table 1. Biological data standards. Some of the most widely used standards in biological sciences and the organization and developed them.

To realize how important metadata is, think on the exclusive rights regimes in which data are most of the time embedded. Those are specified in the metadata and not considering them can lead the user into trouble!



REFERENCES

1. Baca, Murtha (editor). 2016. Introduction to Metadata. Third Edition. Los Angeles: Getty Research Institute. Online Edition: <http://www.getty.edu/publications/intrometadata>
2. Steven J. Miller. "Metadata Resources" University of Wisconsin-Milwaukee. Last update in June 1, 2011. Some revisions in February 2020.
3. National Information Standards Organization (U.S.). Understanding Metadata. Bethesda, MD: NISO Press, 2004.

ABOUT THIS POLICY BRIEF

This Policy Brief is part of a series aiming to inform policy-makers on the key results of the NeDiT research project and provide recommendations to policy-makers. The series of NeDiT Policy Briefs can be found at <http://nedit.net/downloads/>. This publication was commissioned, supervised and produced by NeDiT project partners.

DISCLAIMER

The policy recommendations made do not necessarily reflect the views of the ZMT, IMS or its partners.

IMPRINT

Authors:

Cesc Gordo¹, Hauke Kegler¹, Daudi Msangamo², Christopher Muhando² and Hauke Reuter¹
The authors work at, or are affiliated with, ¹Leibniz Center for Tropical Marine Ecology (ZMT) and ²Institute of Marine Sciences Zanzibar (IMS).

You can find more information about the project [here](#).

The NeDiT project was funded as part of the MeerWissen Initiative by GIZ, you can find more information [here](#).