



Topic 1: Metadata

Sharing data the right way - with standards!

Diversity of Biological Data

When entering a library for the first time, it is quite probable that a citizen manages to find out how to look for a specific book. In a similar way, filling up a document for registering to a social club is easy and predictable even when you do it in a language that is not your native one. This is actually possible thanks to the fact that those procedures are standardized to a certain degree. This means that there are common patterns that apply for each specific case but also for all other similar cases.

Librarians normally order the books in topic categories following author surname and name, and registering sheets normally ask for your name, ID number, address, date of birth, etc. If every library would set up its own classification system, finding a book every time that we enter a new library would be an arduous task!

In a similar way, organizing data can also follow particular standards, although data organization can be addressed through many dimensions. A simplified classification of the most widespread standards, with some examples:

Type of standards	Description	Examples
Structure Standards	Determine the categories of the data object. In the registering sheet example, this standard would determine the different containers to fill in (i.e. name, birth date, address).	Dublin Core Metadata Element Set (DCMES), Encoded Archival Description (EAD), Creative Commons Rights Expression Language (ccREL)
Value Standards	Vocabulary, values, units, etc. that an information object can contain (i.e. a birth date cannot be a name, it has to be a date between now and, let's say 100 years before present).	Thesaurus (specific for each discipline), Union List of Artist Names (ULAN), Integrated Taxonomic Information System (ITIS)
Content Standards	These standards define the rules for how data has to be entered in a database, for example syntax conventions or cataloguing rules (i.e. Birth date has to be entered as dd/mm/yyyy and not dd.mm.yyyy).	Cataloging Cultural Objects (CCO), International Standard Bibliographic Description (ISBD), Anglo-American Cataloguing Rules (AACR)
Encoding standards	These standards define the technical framework for information exchange (i.e. the method through which a library communicates between the other libraries of the city, and send or receive specific books).	Machine Readable Cataloging (MARC), Extensible Markup Language (XML)

How to decide for an appropriate data standard?

There are hundreds of standards out there. Some refer to specific type of data, some refer to metadata, some even define how to reference this data. How can a user decide to use one type of standard? Usually each specialized knowledge domain tends to use their own data standards, in a similar way as the different groups of a team tend to use similar ways to communicate. Only when the data is shared with a broader community, interoperability challenges and problems arise.

In this fact sheet, we discuss the two most widely used metadata standards across all the biological sciences: The Dublin Core Metadata Element Set (DCME), and the Ecological Metadata Language (EML).

Dublin Core Metadata Element Set (DCME)



The DCME is a list of 15 metadata properties used in resource description and it is part of a larger set of standards maintained by the Dublin Core Metadata Initiative (DCMI). Although this metadata standard is not restricted to biological data, it is widely used in many biological fields due to its simplicity.

Dublin Core properties:

- | | | |
|---------------|--------------|-----------|
| - Contributor | - Format | - Rights |
| - Coverage | - Identifier | - Source |
| - Creator | - Language | - Subject |
| - Date | - Publisher | - Title |
| - Description | - Relation | - Type |

Ecological Metadata Language (EML)



This is a metadata standard created by and for the ecology discipline. It is normally the metadata standard associated with data following Darwin Core, which is a standard for biological diversity.

This standard is maintained by “The Knowledge Network for Biocomplexity”, which is a project from ecoinformatics.org, a group of voluntary developers and researchers.

The strength of this metadata standard is its modularity and flexibility. It is written in XML and includes modules for describing spatial, temporal, taxonomic and thematic extent of data, and for annotating data with semantic vocabularies.

REFERENCES

1. More information at: <https://www.dublincore.org>
2. More information at: <https://knb.ecoinformatics.org>

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 Msangameno², 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](#).

