Recommendation S1.0

Recommendation to enrich data with rich metadata

Description

Status: Under development, Date: 2025/07/08 10:18, Version: 001

Motivation for this Recommendation

As data can only be embedded in semantic frameworks when it is described with rich metadata, the first step toward a standardized approach for implementing semantic resources is for data producers to enrich data with metadata - even if this may seem self-evident. Only the standardized use of metadata enables the annotation with identifiable terms from recognized controlled vocabularies, which allows machines to interpret and connect data across disciplinary and institutional boundaries. Since the needs for these metadata can vary greatly between research communities, data infrastructures, and use cases, we recommend using existing metadata schemas commonly used in each of these. In addition, there are generally applicable schemas that can be recommended.

Recommendation summary

All data producers in Helmholtz Earth & Environment should enrich their datasets with rich, standardized metadata at the time of dataset creation, submission to repositories or publication. Repositories, sensor registries and other data infrastructures should ensure that metadata is requested and archived in a standardized and structured manner. This should be done following metadata categories specified in established general or discipline-specific metadata schemas or workflows (see also I2.0 Define exchange format).

Binding Convention:

	mandatory	conditional	optional
Helmholtz FAIR Principle	x		

Precondition for Implementation:

Related Recommendations

Parent: S0

Dependent: S3.0

Other: related to I2.0

Contributors

Content

1. Explanation of the Background and Benefits of the Recommendation

A metadata schema defines the structure, content, and semantics of metadata elements used to describe a dataset. It specifies what metadata should be captured, how it should be named, and in which format it should be stored. Schemas often include controlled vocabularies and formal structures, allowing metadata to be understood both by humans and machines. Widely used schemas include DataCite Metadata Schema for citation metadata, ISO 19115 for geospatial data (hat Platz für Details, viele Abhängigkeiten, in Ausformulierungen überlegen), DCAT for XXX (behörden, flach) and Dublin Core for general resource description (bibliothekarische Metadaten/Attribute?(manche dieser Attribute Pflichtfelder bei Datacite und ISO=, aus den 70er Jahren).

Across Helmholtz Earth and Environmental sciences, metadata standards are applied in diverse repositories and infrastructures, to name just a few of them:

Providing sufficient enrichment of data with metadata forms the basis for the implementation of standardized semantic concepts.

2. Possible alternative solutions

3. Consideration of the advantages and disadvantages of implementing the recommendation

(quality of content, limitations, interoperability, sustainability: expected future dissemination / technical availability / funding)

4. The Recommendation

Bibliographic Metadata/(Administrative metadata?)

It is recommended that for the accurate and consistent identification of a resource for citation and retrieval purposes, each published dataset should be provided with the core metadata elements defined in the most up-to-date DataCite Metadata Schema (see https://schema.datacite.org/).

Community and Repository Alignment

When selecting metadata schemas, data producers should always consider the intended purpose of the metadata:

1. If datasets are to be published in repositories such as PANGAEA or GFZ Data Services, metadata must follow repository-specific workflows and schemas.
2. If datasets need to be interoperable within a scientific community (e.g., oceanography, climate science), community standards like CF Conventions, NetCDF Climate & Forecast metadata, or GCMD keywords should be adopted.
3. If datasets must be integrated into international portals and infrastructures, metadata must be aligned with globally recognized schemas such as ISO 19115 or DataCite.

5. Naming of communities that have already implemented the recommendation

6. Documentation of the test to validate correct implementation

7. Examples of Instances

8. Further Information

References

Relevant Community Recommendations

9. History of this document