

Cataloguing in LTER

Context of cataloguing in LTER

Summary of LTER requirements for cataloguing

Detailed requirements

1. Do you use catalogues or require using catalogues for the following items?

- i. Observation system
- ii. Data processing system
- iii. Observation event and collected sample
- iv. Data processing event
- v. Data product
- vi. Paper or report product
- vii. Research objects or feature interest (e.g. site, taxa, ...)

For LTER Europe the use of standard metadata description is fostered within the community. Nevertheless local metadata system exist, which are under control of the local community. Within the community the following information entities are covered by metadata cataloguing:

- Research site (LTER Site, LTSE Platform, or other long term observation site – providing the thematic, environmental and organisational context of the observation. The metadata model is following the requirements of LTER Europe developed within the EnvEurope project (see <https://data.lter-europe.net/deims/research-site/documentation>). In the near future mapping of the research MD model to INSPIRE Environmental Monitoring Facility data model will be done in order to share the boundaries within the LTER Europe data infrastructure.
- Dataset (data object) is providing information on the data object itself covering information on the data collection and data access. For dataset MD a community MD profile based on EML and ISO19115 was developed in the EnvEurope and ExpeER project (see <https://data.lter-europe.net/deims/dataset/documentation>). This MD are describing the following items TITLE, IDENTIFIER, CREATOR AND CONTACT POINTS, METADATA PROVIDER, METADATA DATE, PUBLICATION DATE, LANGUAGE, ABSTRACT, KEYWORD SET, ACCESS AND USE CONSTRAINTS, INTELLECTUAL RIGHTS, ONLINE DISTRIBUTION, GEOGRAPHIC BOUNDING COORDINATES, GEOGRAPHIC BOUNDING ALTITUDES OR DEPTHS, TEMPORAL EXTENT, TAXONOMIC COVERAGE, METHODS DESCRIPTION, INSTRUMENTATION DESCRIPTION, SAMPLING DESCRIPTION and LEGAL OBLIGATION REPORTING. In the instrumentation description is done in a textual manner. Reference to a method repository would be desired. This is not present at the moment. The research features are part of the MD model for datasets.
- Persons related to the metainformation (e.g. research site or dataset) are stored in the metadatabase. They can/should be referenced by the ORCID-ID (see <https://data.lter-europe.net/deims/person/documentation>).
- Network metadata describing the related national LTER Networks (as the organisational frame conditions for the research sites) (see <https://data.lter-europe.net/deims/network/documentation>).
- Publication metadata give the possibility to link to publications. Links to standard sharing platforms (e.g. DOI, ResearchGate, etc.) should be possible. This is in order to provide full link to the information and also link descriptive articles to research sites and datasets.

The link to the resulting publications (being scientific or non-scientific) should be done in a automated way via the use of PIDs for the data objects. This is showing the added value and use of the research infrastructure.

With regard to the items mentioned above the implementation in LTER Europe is the following

- Observation system à part of the dataset MD and the research site MD
- Data processing system à part of the method description in the dataset MD
- Observation event and collected sample à part of the description in the dataset MD
- Data processing event à part of the description in the dataset MD
- Data product à part of the method description in the dataset MD
- Paper or report product à publication which can be linked to the research site or dataset
- Research objects or feature interest (e.g. site, taxa, ...) à part of the description in the dataset MD (e.g. reference to the research site and taxonomic reference (EML))

1. For each used or required catalogue, consider the following questions:

i. Item descriptions:

- Which fields do you use for describing items?
- Which standards do you apply for these fields (format or standard model)?
- Do you use controlled vocabularies for these fields? If so, please cite the vocabulary service providers.
- Do you maintain a cross-link or inter-links between:
 - Catalogue items (e.g. between observation system - observation event - result dataset)?
 - Fields for item description and actual items (e.g. between dataset fields - dataset access services or between sample fields - label on the sample)?
 - Which repositories/software do you use to manage your metadata?

Depending on the MD model (see above) the items for the MD description are defined. The following standards are used in the LTER Europe context

- Research site MD model – mapping to INSPIRE EF Data Model (see http://inspire.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_EF_v3.0rc3.pdf) – see MD model description at <https://data.lter-europe.net/deims/research-site/documentation>
- Dataset MD model – community metadata profile based on EML and ISO19115; see MD model description <https://data.lter-europe.net/deims/dataset/documentation>.
- Sensor MD model - the use of SensorML for sensor based data (linked to SOS service) is part of the current eLTER project and was already tested in the EnvEurope project.

For most of the MD fields defined reference lists are used, which will, in the next version of EnvThes, be managed by the common controlled vocabulary. Currently the keyword field is referencing to the common vocabulary EnvThes (see <http://vocabs.ceh.ac.uk/evn/tbl/envthes.evn>).

Cross links are provided from the MD records to the specific MD entities, e.g. for the dataset the related research site is referenced. In addition the link to the digital data object (e.g. file or data service) is provided by the 'online distribution link'. This allows to – if open access – directly access the data sources from the metadata record. The terms of use and the data access policy is part of the dataset MD.

MD are created and managed using DEIMS (<http://data.lter-europe.net/deims/>). Additionally other MD editors and repositories can be used within the domain of LTER Europe. Mainly geonetwork is used. Often MD are only available as textual description at the local research site. Standardisation and harmonisation in the field of MD is one of the primary goals of LTER Europe.

ii. Inputs:

- **Human inputs:** Do you provide/need facilities for editors/reviewers to maintain the metadata in the catalogues (e.g. forms, validation workflow, etc.)? If so, please describe them briefly.

MD are mainly edited directly by humans using DEIMS or similar MD editors. DEIMS provides forms to view, edit and update MD records and the links to the data resources. DEIMS is a webbased MD management tool and repository which includes forms for MD editing and pages to discover, view and download MD. Links to the data objects (if provided by the data provider) are part of the MD record and can be accessed by the user.

User authenticities are provided for the different national LTER networks in order to manage their metadata.

- **Machine inputs:** Do you use/ need automated harvesting to populate your catalogues? If so, which protocol do you use (e.g. csw, oai-pmh, other, specific)?

Harvesting of MD is done using harvesting lists (EML) and metadata as well as geonetwork (ISO19115) in order to distribute and harvest MD from different nodes into the discovery portal. For the second standard CSW is used in order to exchange MD.

OAI-PMH as MD exchange standard will be evaluated by LTER Europe in order to harvest and integrate MD for the LTER Europe Data Integration Portal. This work is done within the eLTER and the EUDAT2020 project. In addition EUDAT B2FIND as MD harvesting service will be evaluated. B2FIND is using OAI-PMH standard.

- How do you manage duplicates? i.e. Do you apply governance rules in a network of catalogues, do you use unique identifiers, or take other actions?

Each MD record get a unique identifier in order avoid duplicates (field metadata identifier, see <https://data.lter-europe.net/deims/dataset/documentation#D2>).

iii. Human outputs:

- What specific feature is provided/required in your web discovery function (multi-criteria search, graphical selection components (e.g. map, calendar), facets, keyword or natural language)?

DEIMS supports multi-criteria and faceted search. Only English is supported currently for data description and discovery. With the LTER Europe Data Integration Portal the integration of different MD sources and MD nodes will be done.

Discovery focuses currently on faceted search of MD content. Temporal and spatial search are somehow implemented, but needs to be improved.

Other MD catalogues (e.g. TERENO data portal) are providing search according to temporal and spatial facets.

- Do you evaluate the accessibility, quality, and usage of your catalogue by using a dashboard or value-added products? If so, do you provide/need:
 - Popularity or Usage feedback?
 - Any other synthesis, indicators or statistics?
 - If so, please describe them shortly.

Within DEIMS feedback mechanisms were installed, but not frequently used. Therefore with the new version of DEIMS the possibility of feedback was made more user friendly. Feedback is collected in the new version as JIRA tasks which are assigned to different persons to solve them.

Performance or quality indicators are not collected at the moment.

- Is the catalogue freely readable or do you apply a specific authorization scheme? If you are applying a specific authorization scheme, please cite the authentication system (SSO) and the authorization rules.

The MD platform DEIMS is freely readable. No authorisation rules are implemented for the viewing and discovery. To edit and create MD records user authentication is needed.

iv. Machine outputs:

- Do you provide/need machine interfaces for accessing your catalogues? If so, which protocols are implemented?
- Do you need to fit in Applicable Regulations requirements (e.g. INSPIRE) or embedding frameworks (GEOSS, OBIS)? If so, please cite the regulation, applicable interface requirements, impacts (format, performance, access policy, etc.) on your catalogues.

Metadata on datasets from LTER Europe needs to be compliant to EML in order to allow harvesting within ILTER (and DataOne). EML standard is used to exchange MD.

LTER Europe aims to provide INSPIRE compliant MD and in further steps data services in order to provide an added value for its members to be INSPIRE compliant.

The link to GEOSS, EU-BON, etc. is recommended and will be evaluated in terms of the development and implementation of the LTER Europe Data Integration Platform (DIP).

Formalities (who & when)

Go-between	@Barbara Magagna
RI representative	@Johannes Peterseil
Period of requirements collection	20160118
Status	collected