

MAKING eLTER DATA FAIR

Well documented, quality controlled and curated datasets are a core prerequisite to enable scientific analysis and forecasting. eLTER provides access to a wealth of data dealing simultaneously with ecosystem characteristics and environmental pressures. The application of FAIR principles for research data management is key for eLTER to support Open Science. The FAIR principles ensure Findability, Accessibility, Interoperability (e.g., between RIs and project partners), and Reusability and also Reproducibility for data and information. They facilitate joint data analysis and secondary data usage. eLTER works on harmonised data access and community profiles for standardised metadata and data provision as well as centralised support for data quality control for a plethora of data types.



Providing quality controlled, well-documented, and standardised data is the basis for reliable scientific analysis and is needed for evaluating existing environmental policies. Long-term ecosystem monitoring data are major prerequisites for research not only in Europe but also on a global scale. In order to foster information exchange and sharing, data must be findable, accessible, interoperable and reusable following the FAIR principles. This requires proper documentation of datasets and data services as well as the existence of infrastructure allowing the discovery, access and integration of data in a web-based environment.

The eLTER Research Infrastructure (RI) capitalises on the wider network of LTER-Europe, its sites, platforms and organisational structure. It implements tools and services for advanced data management and fosters the adoption of common standards. This includes the facilitation of controlled vocabularies and metadata standards to enable the interoperability of the eLTER long-term observation data.

Common eLTER Community Profile for datasets and sites

By providing descriptive metadata for the datasets and services information on the context of the observation is provided enabling the discoverability and accessibility. In addition

provenance information enables the understanding of data processing steps, its ownership and origin of data. The eLTER Community Profile for datasets is based on ISO19115/19139 and EML supporting basic INSPIRE compliant metadata provision. For eLTER Sites and Platforms, the INSPIRE Environmental Monitoring Facilities (EF) and OGC Observations & Measurement model are used to ensure interoperability and compliance with European and global standards. eLTERs Digital Asset Register and DEIMS-SDR support the implementation of defined standards allowing the harvesting of metadata to several federated catalogues using standardised interfaces (e.g. OGC CSW, REST-API, OAI-PMH).

Common format for providing eLTER datasets

eLTER developed a simple data exchange format. This includes the most essential attributes for ecological observations following the recommendations for environmental observations (OGC Observation & Measurements) as well as best practices from related networks and monitoring programmes. The eLTER Data Reporting Format (Peterseil & Geiger, 2020) enables easy extension in terms of temporal (e.g. daily, weekly), spatial (e.g. station, plot, transect, profile) or thematic (e.g. variables, observed entities) information without changing the basic reporting structure. It allows the provision of additional information,



e.g. methods applied or quality flags and referencing related information like the eLTER Site via the DEIMS.ID. Based on the modular structure, easy transformation into other formats (e.g. import to OGC Sensor Observation Service, SOS) is enabled.

Common controlled vocabulary

Semantic consistency through the use of controlled terms is an important aspect to ensure interoperability across data. eLTER collaborates with the long-term ecosystem monitoring community to develop the controlled vocabularies (<https://vocab.ter-europe.net/en/>), covering the main needs for keyword tagging and parameter naming both for sites and datasets. The vocabularies are established based on a collaborative effort inviting experts and users to contribute to its development. eLTER vocabularies (EnvThes) re-use a number of existing vocabularies (e.g. US LTER Vocabulary) as well as conceptual models (e.g. OGC Observation & Measurements, OBOE and SERONTO) with main design principles

from the Complex Properties Model. The recommendations developed by the RDA I-ADOPT working group are implemented to enhance semantic interoperability.

Persistent identification (PIDs) of sites and datasets

Persistent identification of the digital assets of the eLTER RI is important to ensure reproducibility. For eLTER in-situ facilities (sites and platforms), a persistent and unique identifier (DEIMS.ID) is implemented, allowing to reference sites and related data across infrastructures and networks. The DEIMS.ID enables a “FAIR” representation of each site in research publications, reports, and presentations and makes related data and information available and reusable to a broader audience. For the operational eLTER RI, datasets will have assigned DOIs (where appropriate) from the Digital Asset Registry. This registry will relate identifiers from data in different repositories (such as EUDAT) to eLTER data products through linked PIDs.

The screenshot shows the eLTER Vocabularies website. The main heading is 'EnvThes - Thesaurus for long term ecological research, monitoring and experiments'. There is a search bar with 'Content language English' and a search button. On the left, there is a list of terms under 'Alphabetical' and 'Hierarchy' tabs. The selected term is 'area of slope 1% arable land of catchment'. The detailed view on the right shows the following information:

- variable > area of slope 1% arable land of catchment
- PREFERRED TERM: **area of slope 1% arable land of catchment**
- BROADER CONCEPT: variable
- CREATOR: 0000-0003-2195-3997
- HASCONSTRAINT: slope 1%
- HASCONTEXTOBJECT: catchment
- HASOBJECTOFINTEREST: http://purl.obolibrary.org/obo/ENVO_01001177
- HASPROPERTY: area
- URI: <http://vocab.ter-europe.net/EnvThes/30217>
- DOWNLOAD THIS CONCEPT: RDF/XML TURTLE JSON-LD
- Created 3/28/22, last modified 3/27/22

Making eLTER data FAIR

- Implement a **community profile for metadata documentation**, balancing usability and efforts
- Implement a **community profile for data provision and reporting** enabling harmonised virtual access to data
- Develop and reuse **common controlled vocabularies** for the annotation of metadata and data
- Provide a **common interface and services** for data and metadata discovery, visualisation and access

References

- Peterseil, J., Geiger, S. (2020) Field specification for data reporting. Technical Document TechDoc.01. EU Horizon 2020 eLTER PLUS Project, [<https://doi.org/10.5281/zenodo.6373410>]
- Wohner, C. et al. (2020) Towards interoperable research site documentation – Recommendations for information models and data provision. Ecological Informatics (Vol. 60), [<https://doi.org/10.1016/j.ecoinf.2020.101158>]
- Wohner, C., Peterseil, J., Klug, H. (2022) Designing and implementing a data model for describing environmental monitoring and research sites. Ecological Informatics (Vol. 70), [<https://doi.org/10.1016/j.ecoinf.2022.101708>]
- Wohner, C. et al. (2019) DEIMS-SDR – A web portal to document research sites and their associated data. Ecological Informatics (Vol. 51), [<https://doi.org/10.1016/j.ecoinf.2019.01.005>]