European Long-Term Ecosystem and Socio-Ecological Research Infrastructure

D3.2 Governance and data policy for sharing and publishing of data

Authors: Ralf Kunkel (FZ-Jülich), Johannes Peterseil (EAA), Alessandro Oggioni (CNR-IREA), Christoph Wohner (EAA), John Watkins (CEH), Vladan Minić (BSI), Jürgen Sorg (FZ-Juelich
Lead partner for deliverable: FZJ
Other partners involved: NERC/CEH, EAA, BSI, CNR; UGR

H2020-funded project, GA: 654359, INFRAIA call 2014-2015

Start date of project: 01 June 2015 Duration: 53 months

Version of this document: 1.0 Submission date: 30th June 2019

Dissemination level

PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	
CI	Classified, as referred to in Commission Decision 2001/844/EC	

Version control	Edited by	Date of revision
Created - V1	Ralf Kunkel	2018-06-12
Internal review	Johannes Peterseil	2018-12-20
Internal review		
Revised – V2	Ralf Kunkel	2019-05-21
Revised – V3	Johannes Peterseil	2019-06-05
Revised – V4	Ralf Kunkel	2019-06-28
Reviewed	Johannes Peterseil	2019-06-30
Revised – V5		
Signed off – coordinator	Johannes Peterseil	2019-06-30

Publishable Executive Summary

The site network of LTER Europe provides an important source of data for the assessment and understanding of ecosystem processes and their relation to environmental pressures and threats. LTER is organised by different national networks sharing a common conceptual basis, but managed by different organisations with their own funding regimes and organisation structures. The eLTER H2020 project aims to develop a common information management infrastructure for making data from the different distributed resources available, not only for the scientific expert but also for a more general use. Efforts are taken to harmonise measurements with the aim of building a sustainable research infrastructure for long term ecosystem observations. Innovative solutions are needed in order to allow the discovery and access of the data in near real time. This includes the documentation of the observation context as well as facilitating data services to provide online access to the data.

Within eLTER H2020, an information infrastructure has been implemented to document and share the data. The system builds on a network of distributed data services integrated by a central discovery portal and federated data access components.

This document addresses the social and related socio-technical aspects of data publishing within a network of distributed data sources.

In the first part, data policies from the project partners as well as from thematically related research infrastructures and networks, together with Open Research Data Policy are evaluated. As a result, a common eLTER Data Policy and data sharing agreements within and beyond the project has been developed.

In the second part of the document, governance issues of common data sharing principles are addressed. An overview on the actions taken to provide the basis of decisions for the design and the implementation of the eLTER data infrastructure are given. Also, the basic high-level user stories defined in order to provide a basis for the evaluation of different architectures and the related existing or planned tools are described.

Contents

1 Introduction

LTER Europe[1], established in 2003, aims to improve harmonisation and standardisation of long term observation in the ecosystem domain. With its network of sites and researcher it is one of the European scale infrastructures in place providing data and expertise in order to bridge the gap between science and decision making. Getting an overview and access to information on available data and scientific observation infrastructures are therefore one of the central points to the success for the implementation of data intensive science.

The curation of data and the provision of metadata are key processes in each of the environmental projects. Getting an added value on the comparison of the ecological data, also between different projects, requires that the data are available and discoverable for usage by researchers. A distributed infrastructure allows each site to maintain data that has been collected, but a structured life cycle and information technology (IT) tools must be defined.

The project eLTER (H2020) works towards further implementation of services for the LTER community and enhancing data accessibility. Focusing on the requirements and conceptual framework for data integration is part of the tasks to implement a eLTER data infrastructure. By this the IT requirements of various user groups were collected, ranging from data users (e.g. research communities), data providers (e.g. LTER site managers), or data networks (e.g. GEOSS harvesting metadata). This includes the development of a common data policy in order to reduce data access restrictions to the degree possible.

European scale infrastructure projects like ENVRI and ENVRI+[2] support the work on the development of eLTER data infrastructure providing general guidelines. Fostering not only the machine-to-machine communication sharing metadata in a standardised manner using metadata standards (e.g. ISO 19115/19139) but also the sharing of data is an important step.

The eLTER data infrastructure will be an additional data node in the European and global e-infrastructure on environmental data. The evaluation of relevant data infrastructure, either providing best practise or being a relevant node to connect to, is part of the work. European scale regulations, like INSPIRE, or global data networks, like LifeWatch or DataOne, were taken into consideration.

This document addresses the social and related socio-technical aspects of data publishing and use within a network of distributed data sources.

In the first part, data policies from the project partners as well as thematically related research infrastructures and networks and the Open Research Data Policy will be evaluated. As a result, a common eLTER Data Policy and data sharing agreements within and beyond the project is developed.

In the second part of the document, governance issues of common data sharing principles are addressed. An overview on the actions taken to provide the basis of decisions for the

[1] see https://www.lter-europe.net/

[2] See http://www.envriplus.eu/ and https://confluence.egi.eu/display/EC/ENVRI+Reference+Model

design and the implementation of the eLTER data infrastructure are given. Also, the basic high-level user stories defined in order to provide a basis for the evaluation of different architectures and the related existing or planned tools are described.

The work was also guided by previous developments in the LTER Europe context. A guiding principle was to focus on open source products as well as to foster an open access policy wherever possible and useful.

2 Setup of the eLTER infrastructure

Development of a data policy and a concept for long-term governance of the eLTER data infrastructure has to be done based on its design. This chapter will give an overview of the eLTER data infrastructure setup as much as it is relevant for this task.

Network of networks

LTER-Europe is the European network for Long-Term Ecosystem Research (LTER) comprising of the different national LTER networks. The network is the result of a 15 years harmonisation and integration process of ecosystem research infrastructures in 30 countries, which resulted in formal LTER networks in 21 countries with well-established national and European governance structures and embedded in the global LTER network (ILTER). Since LTER-Europe was heavily involved in developing the concept of Long-Term Socio-Ecological Research (LTSER), it now comprises also larger LTSER platforms.

In addition, LTER Europe is comprised of a network of national networks, which are running the different long-term monitoring sites. Funding and organisation of the work is dedicated to the different local institutions coordinated by the national as well as the European LTER network.

Network of sites

The eLTER infrastructure comprises about 420 ecosystem research sites (65% terrestrial, 26% aquatic and 9% transitional waters LTER Sites) and 35 LTSER Platforms for socio ecological research at the regional scale. Around 100 different institutions operate the infrastructures. LTER-Europe has condensed research sites originally set up in varying contexts (projects and networks driven by national/institutional strategies and domain specific requirements). All sites and national networks comply with a refined site classification reaching from highly instrumented main sites (19%), to regular LTER sites (44%), extensive (24%) and emerging sites (4%).

Network of data management

The management of data for most of the LTER sites is the responsibility of the respective site. This creates a high heterogeneity in the data management processes and tools. Basic information on the data management at the different LTER sites was provided using DEIMS-SDR. In the section on Infrastructure and Data Management relevant characteristics are listed. This included data storage, standards, metadata and data provision. Management of data within the LTER network is distributed. Data are usually managed by the local sites and shared e.g. with international programmes using the respective data sharing formats.

Looking at the storage and management of long-term data at the different LTER sites it shows, that the vast majority of information is stored in structured files or spreadsheets (45,5%) followed by relational databases (32,4%). Spatial information is either stored file-based (21,7%) or in a geodatabase (19,0%). The majority of data is managed in a structured way (structured files or databases). Nevertheless, a minority of the data still is managed either in proprietary file formats (e.g. of a given sensor) or on paper.

3 Compliance

"Open science is the idea that scientific knowledge of all kinds should be openly shared as early as is practical in the discovery process." Michael Nielsen[3]

"There's now a clear understanding that we need robust data to drive democracy and development- and a lot of it." Open Data Barometer 2016[4]

"Data has become a key asset for the economy and our societies similar to the classic categories of human and financial resources. Whether it is geographical information, statistics, weather data, research data, transport data, energy consumption data, or health data, the need to make sense of data is leading to innovations in technology, development of new tools and new skills." European Commission[5]

The availability and reusability of data to address both understanding ecosystem processes in their relation to environmental threats and changes as well as to forecast future behaviour based on scenarios is a core prerequisite for any environmental research. In the past decades both monitoring efforts, e.g. based on remote or in-situ sensing, as well as the technical capabilities to transfer and ingest large volumes of data have advanced. Nevertheless, accessibility and moreover interoperability is not always granted despite many efforts in this direction. Especially in the scientific domain long term archiving and securing the availability is still an issue. Projects like OpenAIRE[6] are tackling this issue providing recommendations, trainings and tools to support.

A number of directives implemented on the global (e.g. Aarhus Convention[7]) and regional scale (e.g. Environmental Information Directive (2003/4EC) [8] addressed access to environmental data but did not address open data and science per-se. In recent years this aspect became more prominent in order to enable smart, sustainable and inclusive growth. On the European scale the Digital Single Market (DSM) forms one of the ten political priorities of the European Commission and, as part of the Digital Agenda for Europe (DAE), is also one of the 7 flagship initiatives set by the "Europe2020" strategy. DSM envisages a market where *"free movement of goods, persons, services and capital is ensured … under conditions of fair competition, and a high level of consumer and personal data protection, irrespective of their nationality or place of residence."* This endorses open science and open access to scientific results. The objective is to provide European science, industry and public authorities with excellent digital infrastructure - supercomputing and data storage. In addition

[3] see http://michaelnielsen.org/blog/open-science-2/
[4] see https://opendatabarometer.org/3rdedition/report/
[5] see https://ec.europa.eu/digital-single-market/en/big-data
[6] see https://www.openaire.eu/
[7] see https://www.unece.org/fileadmin/DAM/env/pp/documents/cep43e.pdf
[8] see https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2003:041:0026:0032:EN:PDF

to directives implemented on the European scale Initiatives, e.g. FAIR, OpenScience or GEOSS, provide principles in order to ensure a FAIR exchange of information and data.

For LTER Europe as well as for the newly established eLTER research infrastructure the implementation of these principles both on the network as well as on the technical level are crucial for the long-term implementation and were addressed in the architectural design (see Watkins et al. 2016).

This chapter aims to provide an overview on relevant legal binding directives as well as community-based principles in order to foster open data and science. The selection was done on their relevance and applicability for eLTER data flows and the eLTER Research Infrastructure.

3.1 Relevant principles fostering open data and science

Starting with defining levels of open data (Berners-Lee, 2006) in the recent years a number of relevant initiatives like the FAIR principles (Wilkinson et al., 2016) emerged aiming to provide general guidelines for the implementation of open data. Common across the different initiatives is the definition of "*open data*" as a core principle and prerequisite for open data and open science. The key features for openness are (Open Knowledge Foundation, 2019):

- ***Availability and Access:*** *the data must be available as a whole and at no more than a reasonable reproduction cost, preferably by downloading over the internet. The data must also be available in a convenient and modifiable form.*
- ***Re-use and Redistribution:*** *the data must be provided under terms that permit re-use and redistribution including the intermixing with other datasets.*
- ***Universal Participation:*** *everyone must be able to use, re-use and redistribute - there should be no discrimination against fields of endeavour or against persons or groups. For example, 'non-commercial' restrictions that would prevent 'commercial' use, or restrictions of use for certain purposes (e.g. only in education), are not allowed.*

The resulting key principles relevant in the eLTER context are the FAIR principles defined by FORCE11 as well as the Open Science and Open Data Principles defined by the European Commission.

3.1.1 FAIR principles

The European Commission encourages all research funded in Horizon 2020 to be open access, enabling everybody to openly access scientific publications and to validate the published results. Open access means that research results and research data are stored in a research data repository so that it is possible to access, mine, exploit, reproduce and disseminate the data, free of charge for any user. Examples of research data include statistics, results of experiments, measurements, observations resulting from field work, survey results, interview recordings and images Data provision itself should follow the FAIR data principles[9] (Wilkinson et al.; 2016), meaning that all research data are Findable,

[9] https://www.go-fair.org/fair-principles

Accessible, Interoperable and Reusable. These principles refer to data (or any digital object), metadata (information about that digital object), and infrastructures.

- To use data, it is necessary to find them both for humans and computers. Metadata are a crucial part of any data archive since they ensure that the data can be understood later. It is essential that metadata richly describe the data with a plurality of accurate and relevant attributes and are submitted at the same time as the data sets to which they pertain. Metadata should be registered or indexed in a searchable resource, be machine-readable to enable automatic discovery of datasets and services and should be, like the data itself, assigned by a globally unique and persistent identifier. In addition, metadata clearly and explicitly include the identifier of the data they describe.
- After finding data and metadata it is required that both are accessible. Metadata are valuable in and of themselves, when planning research, especially replication studies. Even if the original data are missing, tracking down people, institutions or publications associated with the original research can be extremely useful. Therefore, metadata need to be accessible, even when the data are no longer available. Data and metadata are should be retrievable by their identifier using a standardised communications protocol, which is open, free, and universally implementable. If necessary, the protocol allows for an authentication and authorisation procedure.
- In order to be able to integrate data with other data, data and metadata need to interoperate with applications or workflows for analysis, storage, and processing. Therefore, a formal, accessible, shared, and broadly applicable language for knowledge representation as well as controlled vocabularies following the FAIR principles needs to be used. If applicable, metadata and data will include qualified references to other (meta)data.
- To make Metadata and data reusable, they should be well-described and meet domain-relevant community standards so that they can be replicated and/or combined in different settings. Clear and accessible licenses are required when releasing data and metadata to allow proper use.

3.1.2 Open Science and Open Data

Open Science represents a new approach to the scientific process based on cooperative work and new ways of diffusing knowledge by using digital technologies and new collaborative tools (European Commission, 2016b:33). Open Science is frequently defined as an umbrella term that involves various movements aiming to remove the barriers for sharing any kind of output, resources, methods or tools, at any stage of the research process. The main six principles of Open Science are :

- Open data: Make data freely available
- Open Source: Use open source technology (software and hardware) and open your own technologies
- Open methodology: document the application of methods as well as the entire process behind as far as practicable and relevant documentation
- Open peer review: Transparent and traceable quality assurance through open peer review

- Open Access: Publish in an open manner and make it accessible and accessible to everyone
- Open educational resources: Use Free and Open Materials for Education and University Teaching

Even though, especially for the library and information domain, the focus is usually placed on two of these movements: Open (research) Data and Open Access to scientific publications.

The aim of Open Access (OA) is the maximisation of the dissemination of scientific and scholarly information. According to the Budapest Open Access Initiative OA literature is freely availability on the public internet, permitting any users to read, download, copy, distribute, print, search, or link to the full text of these articles, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. The only constraint on reproduction and distribution, and the only role for copyright in this domain, should be to give authors control over the integrity of their work and the right to be properly acknowledged and cited.

The Open Definition sets out principles that define “openness” in relation to data and content. It makes precise the meaning of “open” in the terms “open data” and “open content” and thereby ensures quality and encourages compatibility between different pools of open material. It can be summed up in the statement that:

> *“Open means anyone can freely access, use, modify, and share for any purpose (subject, at most, to requirements that preserve provenance and openness).”*

3.1.3 GEOSS Data Sharing Principles

One of the first accomplishments of the Group on Earth Observations (GEO) was the acceptance of a set of high-level Data Sharing Principles as a foundation for GEOSS. Ensuring that these principles are implemented in an effective yet flexible manner remains a major challenge. The 10-Year Implementation Plan says "The societal benefits of Earth observations cannot be achieved without data sharing" and sets out the GEOSS Data Sharing Principles (GEO, 2010):

- There will be full and open exchange of data, metadata and products shared within GEOSS, recognizing relevant international instruments and national policies and legislation;
- All shared data, metadata and products will be made available with minimum time delay and at minimum cost;
- All shared data, metadata and products being free of charge or no more than cost of reproduction will be encouraged for research and education.

By this the GEOSS Data Sharing principles are focusing mainly on the aspect of discoverability, accessibility and reusability.

3.2 Relevant European directives and initiatives

3.2.1 Aarhus Convention

The Aarhus Convention, adopted in 1998, established a number of rights with regard to the environment, including rights on access to information. This can include information on the state of the environment, but also on policies or measures taken, or on the state of human health and safety where this can be affected by the state of the environment. Applicants are entitled to obtain this information within one month of the request and without having to say why they require it. In addition, public authorities are obliged, under the Convention, to actively disseminate environmental information in their possession. This Convention has been translated into EU legislation through a series of Directives and Regulations binding Member States as well as European Institutions and Bodies[10]. Neither Directive on Public Access to Environmental Information (2003/4/EC) nor the UNECE Aarhus Convention are dealing with open data in general or open science data in particular.

3.2.2 Public Sector Information (PSI) Directive

The Directive on the re-use of public sector information[11] (PSI) is built around two key pillars of the internal market: transparency and fair competition. It sets minimum rules for the re-use of PSI throughout the European Union. In its recitals it encourages Member States to go beyond these minimum rules and to adopt open data policies, allowing a broad use of documents held by public sector bodies[12].

In 2006, the European Commission adopted a Decision on the re-use of Commission information, going beyond the PSI directive rules. On 12 December 2011, the Commission proposed to update the 2003 Directive on the re-use of public sector information by:

---

[10] see Directive 2003/4/EC of the European Parliament and of the Council of 28 January 2003 on public access to environmental information and repealing Council Directive 90/313/EEC, OJ L 41, 14.2.2003, p. 26–32.
Directive 2003/35/EC of the European Parliament and of the Council of 26 May 2003 providing for public participation in respect of the drawing up of certain plans and programmes relating to the environment and amending with regard to public participation and access to justice Council Directives 85/337/EEC and 96/61/EC, OJ L 156, 25.6.2003, p. 17–25.

Regulation (EC) No 1367/2006 of the European Parliament and of the Council of 6 September 2006 on the application of the provisions of the Aarhus Convention on Access to Information, Public Participation in Decision-making and Access to Justice in Environmental Matters to Community institutions and bodies, OJ L 264, 25.9.2006, p. 13–19.

[11] see Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the re-use of public sector information, OJ L 345, 31.12.2003, p. 90–96.

[12] see Quick guide to 2003/98/EC Directive : http://ec.europa.eu/information_society/policy/psi/docs/pdfs/ directive/quick_guide_directive.pdf

- Making it a general rule that all documents made accessible by public sector bodies can be re-used for any purpose, commercial or non-commercial, unless protected by third party copyright;
- Establishing the principle that public bodies should not be allowed to charge more than costs triggered by the individual request for data (marginal costs); in practice this means most data will be offered for free or virtually for free, unless duly justified.
- Making it compulsory to provide data in commonly-used, machine readable formats, to ensure data can be effectively re-used.
- Introducing regulatory oversight to enforce these principles;
- Massively expanding the reach of the Directive to include libraries, museums and archives for the first time; the existing 2003 rules will apply to data from such institutions.

This update is currently being discussed in order to be adopted by the European Parliament and the Council to become the new PSI directive. The future Open Data and Public Sector Information Directive will make more data more available and reusable. The new rules will[13]:

- Identify, by way of adoption of an implementing act, a list of high-value datasets such as geospatial or statistics data to be provided free of charge. These datasets have a high commercial potential, and can speed up the emergence of value-added EU-wide information products and services and the development of AI.
- Stimulate publication of dynamic data and uptake of Application Programme Interfaces.
- Limit the exceptions which currently allow public bodies to charge more than the marginal costs of dissemination for the re-use of their data.
- Enlarge the scope of the Directive to
    - *data held by public undertakings*, under a specific set of rules. In principle, the Directive will only apply to data which the undertakings make available for re-use. Charges for the re-use of such data can be above marginal costs for dissemination
    - *research data resulting from public funding* – Member States will be asked to develop policies for open access to publicly funded research data. New rules will also facilitate the reusability of research data that is already contained in open repositories.
- Strengthen the transparency requirements for public–private agreements involving public sector information, avoiding exclusive arrangements.

By this the PSI is strengthening the right to access environmental data and is affecting other directives link the INSPIRE directive focusing on harmonised access to spatial data from federal bodies.

[13] see https://ec.europa.eu/digital-single-market/en/news/ factsheet-building-data-based-economy-eu-agreement-future-open-data-and-public-sector

3.2.3 INSPIRE Directive

The INSPIRE directive (2007/2/EC)[14] came into force on 15 May 2007 and was implemented in various stages, with full implementation required by 2021 based on the roadmap[15]. The INSPIRE directive aims to create a European Union (EU) spatial data infrastructure. This should enable the sharing of environmental spatial information among public sector organizations and better facilitate public access to spatial information across Europe. A European Spatial data infrastructure assists in policy-making across boundaries. Therefore, the spatial information considered under the directive is extensive and includes a great variety of topical and technical themes. INSPIRE is based on a number of common principles:

- Data should be collected only once and kept where it can be maintained most effectively.
- It should be possible to combine seamless spatial information from different sources across Europe and share it with many users and applications.
- It should be possible for information collected at one level/scale to be shared with all levels/scales; detailed for thorough investigations, general for strategic purposes.
- Geographic information needed for good governance at all levels should be readily and transparently available.
- Easy to find what geographic information is available, how it can be used to meet a particular need, and under which conditions it can be acquired and used.

INSPIRE is based on infrastructures for spatial information established and operated by the 27 Member States of the European Union. The Directive addresses 34 spatial data themes needed for environmental applications, with key components specified through technical implementing rules. This makes INSPIRE a unique example of a legislative "regional" approach.

3.2.4 Database Directive

The most important directive regarding rights on data and information is Directive 96/9/EC of 11th March 1996 on the legal protection of databases[16]. The whole copyright law framework is also relevant in this matter as we will see in the last paragraph with the caveat that copyright protection does not apply in most jurisdictions to mere facts or Earth physical parameters as they are not original and in the public domain.

This lack of protection led to the adoption of the database Directive which grants a sui generis right, bearing some distant similarities with copyrights, to the person investing in the compilation of a database. However, the circumstances opening the right to the sui generis

[14] see https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32007L0002

[15] see http://inspire.ec.europa.eu/index.cfm/pageid/44

[16] see Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases; OJ L 77, 27.3.1996, p. 20–28.

right were restrictively interpreted by the European Court of Justice in 2004[17], making a distinction between the investment needed for collecting the data and the investment needed for creating the data. Only the first category of investment is a condition for granting the sui generis rights to the maker of the investment. This interpretation by the Court of Justice would leave the investment in a data collection infrastructure outside the criteria for sui generis protection.

Databases may also be protected by copyrights when by reason of the selection or arrangement of their contents, they constitute the author's own intellectual creation. The copyright protection of databases does not however extend to the database elements, i.e. data and information. These elements can be separately protected by copyrights if they fulfil by themselves the criteria for protection. When no EU directive harmonizes the criteria for protection of these elements, a legal analysis must be conducted at national and international levels as the legislation on copyrights is not fully harmonized at EU level.

3.2.5 Software Directive

Directive 2009/247EC of the European Parliament and of the Council of 23 April on the legal protection of computer programs[18] is the frame for copyright protection for computer programs as literary works in the EU.

The term computer program includes programs in any form, including those which are incorporated into hardware. The term also includes preparatory design work leading to the development of a computer program provided that the nature of the preparatory work is such that a computer program can result from it at a later stage.

Only the expression of a computer program is protected and ideas and principles which underlie any element of a program, including those which underlie its interfaces, are not protected by copyright. So to the extent that logic, algorithms and programming languages comprise ideas and principles, those ideas and principles are not protected, only the expression of those ideas and principles.

The author of a computer program is the natural person or group of natural persons who has created the program or, where the legislation of the Member State permits, the legal person designated as the right holder by that legislation. Where a computer program is created by an employee in the execution of his duties or following the instructions given by his employer, the employer exclusively shall be entitled to exercise all economic rights in the program, unless otherwise provided by contract.

[17] see Cases C-46/02, Fixtures Marketing Ltd v. Oy Veikkaus Ab, ECR 2004, p. I-10365 and C-203/02, The British Horseracing Board Ltd and Others v William Hill Organization Ltd, ECR 2004, p. I-10415.

[18] see Directive 2009/24/EC of the European Parliament and of the Council of 23 April 2009 on the legal protection of computer programs, OJ L 111, 5.5.2009, p. 16–22.

4 Data Policy

An essential aspect of eLTER H2020 is the exchange and availability of the data within eLTER and with the public. In this context, the Data Sharing Policy (DSP) provides a legal framework for the distribution and reuse of primary and secondary research data. The eLTER Data Sharing Policy defines main data sharing issues from the viewpoint of very different actors and is intended to provide a guideline for the definition and implementation of local data policies and licenses.

4.1 Actors in data sharing

Data user

The main interests of a data user are to find, query, discover and access data. Therefore, the data sharing policy has to specify

- which data can be accessed and how access is granted
- the scope for using, redistributing and/or citing the data (licensing)

Data provider

The data provider offers his data to the community. He has to describe the data by “rich” metadata, convert it into formats supported/required by the infrastructure, perform required processing and quality assessment and publish data and metadata either by own or central infrastructure services. A major concern of data providers are questions of data ownership, intellectual property, citation of data. Often, there are concerns that once knowledge was shared, it is no longer possible to publish findings in a journal. In addition, often there is a lack of individual incentives e.g. recognition and awareness of scientist‘s work, e.g. by proper citation of data in case of reuse. Data providers may also have concerns on missing control on data distribution and re-use by the data provider and, finally, have missing knowledge on how to share data technically and do not want to invest time and effort on this issue. From the viewpoint of a data provider, the data sharing policy specifies:

- which data and metadata have to be provided
- which data and data exchange formats/interfaces/vocabularies need to be used
- how the data, its processing and quality assessment procedures have to be described
- how the data have to be cited
- under which access conditions and licenses the data will be provided
- in which time periods data have to be provided

Data manager

The data managers are responsible for the publication of data and metadata. They provide access to data on a technical level, set up the necessary web services forward licensing information to the user and implements data access constraints into the infrastructure. From the viewpoint of a data manager, the data sharing policy specifies:

- what data metadata formats are supported

- what information on data and metadata is published to the data users
- if there are any retention times after which data will be published automatically
- data privacy issues to be considered
- access constraints to be implemented

Although most of the general issues on data sharing is defined in the **Data Sharing Policy (DSP)**, information on how the data are shared within the infrastructure is specified in the **Data Management Plan (DMP)**.

4.2 Survey of existing data policies

In order to develop an eLTER DSP, a survey of existing data policies was performed, which is described in detail in this chapter. A number of data policies from a number of relevant projects, networks or research infrastructures are evaluated. Main focus was laid on the main issues of access to data and metadata, data delivery, cost and licensing policy. This evaluation was the basis for the revision of the eLTER Data Sharing Policy.

4.2.1 CarboEurope

The aim of the CarboEurope-IP is to understand and quantify the terrestrial carbon balance of Europe and associated uncertainties at local, regional and continental scale. In order to achieve this strategic objective, the project addresses questions of the spatial-temporal carbon balance of the European continent and the controlling mechanisms of carbon cycling in European ecosystems. CarboEurope developed an infrastructure consisting of distributed databases and a central database for the integrated project. A common data policy was developed[19]

Metadata access	Accessible via the internet for external users
Data access	Data access from a Web-GIS platform. Data are freely accessible without any restrictions. Internal and external access depends on data processing and publication status.
Data delivery	Depends on data types between every half year or within one month after end of campaign
Cost	Free of charge or no more than cost of reproduction and delivery
Licensing	Acknowledgement of data sources, distribution to third parties not allowed without prior consent from Data Manager and Project Leader. Data provider can choose to give access to the data with an open license (CC BY-SA or CC-0)

[19] see http://carboeurope.org/

4.2.2 COCONET

COCONET –towards COast to COast NETworks of marine protected areas (from the shore to the high and deep sea), coupled with sea-based wind energy potential (CoCoNET), funded within the FP7 (EU 7th Framework Programme for Research and Technological Development) Program, is meant to enhance policies of effective environmental management, also to ascertain if the existing MPAs are sufficient for ecological networking and to suggest how to design further protection schemes based on effective exchanges between protected areas. The coastal focus is widened to offshore and deep-sea habitats, comprising them in MPAs Networks. These activities will also individuate areas where Offshore Wind Farms might become established, avoiding too sensitive habitats but acting as stepping stones through MPAs[20].

Metadata access	Freely accessible without any restrictions
Data access	Freely accessible without any restrictions. Requires acceptance of user license; data access form provided. Focal point is a WebGIS platform
Data delivery	not mentioned
Cost	free of charge or no more than cost of reproduction and delivery
Licensing	Acknowledgement of data sources Distribution to third parties not allowed without prior consent from Dtaa Manager and Project Leader. Data provider can choose to give access to the data with open licenses (CC BY-SA or CC-0)

4.2.3 ECOPOTENTIAL

Ecopotential is a project funded under the H2020 research programme The work focuses on the integrated use of Earth Observations from remote sensing and field measurements, data analysis and modelling of current and future ecosystem conditions and services in order to provide data for informed management decisions and improve PA management workflows. ECOPOTENTIAL considers cross-scale geosphere-biosphere interactions at regional to continental scales, addressing long-term and large-scale environmental and ecological challenges. The Ecopotential data policy distinguishes between external (generated outside the project context) and internal (generated within the project activities) data. The data sharing policy[21] follows the GEOSS Data Sharing Principles.

[20] see http://coconetgis.ismar.cnr.it/documets/COCONET_DataPolicy.pdf
[21] see https://www.ecopotential-project.eu/images/ecopotential/documents/datapolicy.pdf

Metadata access	Full and open access to all metadata through the Ecopotential In-situ Catalogue[22] harvested by the GEOSS Data Collection of Open Resources for Everyone (Data-CORE)
Data access	Internal data are free and open as defined in the online distribution link in the metadata; external data might be subjected to restrictions of data use and sharing
Data delivery	All shared data, products and metadata will be made available with minimum time delay. The data sharing policy defines a possible embargo period of maximum 6 month.
Cost	Free of charge or at no more than cost of reproduction
Licensing	ECOPOTENTIAL does not assert any intellectual property rights to data and knowledge generated through ECOPOTENTIAL project (= internal data). For internal data and information that qualify as works in the sense of copyright are dedicated to the public domain (by CC-0-Waiver or Open Data Commons Public-domain-declaration) or made available under a CC-BY 4.0-License: (a) CC0, under which data are made available for any use without restriction or particular requirements on the part of users; (b) CC-BY, under which data are made available for any use provided that attribution is appropriately given for the sources of data used, in the manner specified by the owner; and (c) CC-BY-NC, under which data are made available for any use provided that attribution is appropriately given and provided the use is not for commercial purposes

4.2.4 ENVEUROPE

EnvEurope was conceived synergistically dealing with several key targets and in response to the challenges of research within the Long-Term Ecosystem Research in Europe site network (LTER-Europe). The project has been structured to play a role in the conceptual and operative context of the Shared Environmental Information System (SEIS) promoted by the European Commission and in the development of some components of the Global Monitoring for Environment and Security (GMES), a joint initiative of the European Commission and European Space Agency. The project combines existing data with new data generated during the project to achieve a deeper understanding of ecosystem functioning, improved environmental management and support for the development of EU environmental policies and conservation planning through the integration of objectives, resources and disciplines. EnvEurope developed a data policy to ease collaboration among the participants of the EnvEurope project, to ensure timely submission of data, to protect the researcher's rights to publish their results, to provide rules for the use of the data and, finally, to provide the broader scientific community with an easy access to the data available within the EnvEurope project.

[22] see http://data.ecopotential-project.eu:8081

Metadata access	Free for use to any party within and outside the EnvEurope community
Data access	Only for ENVEUROPE members within the project runtime. Users outside of EnvEurope have to be dealt by E-Mail with data provider
Data delivery	As soon as possible. Time periods for data delivery set for different kinds of data
Cost	Free of charge or at no more than cost of reproduction
Licensing	No specific license given

4.2.5 EPOS

EPOS, the European Plate Observing System, is a long-term plan to facilitate integrated use of data, data products, and facilities from distributed research infrastructures for solid Earth science in Europe. Through integration of the existing national and trans-national research infrastructures EPOS will contribute to the access and use of the multidisciplinary data recorded by the solid Earth monitoring networks, acquired in laboratory experiments and/or produced by computational simulations. The EPOS infrastructure consists of three complementary elements: National Research Infrastructures, Thematic Core Services and Integrated Core Services[23].

Metadata access	Metadata are free and available at any time and a CC-BY 4.0 license, even for restricted and embargoed data.
Data access	Depending on type of data either open, restricted to certain community or embargoed for a certain period of time.
Data delivery	Minimum time delay.
Cost	Free of charge or at no more than cost of reproduction
Licensing	All published data have to be connected to a license. As a default, CC-BY 4.0 and CC-BY-NC 4.0 licenses are used. Third parties may provide own licenses

4.2.6 EEA

The European Environment Agency (EEA) is an agency of the European Union, whose task is to provide information on the environment. The EEA aims to support sustainable development by helping to achieve significant and measurable improvement in Europe's environment, through the provision of timely, targeted, relevant and reliable information to policy making agents and the public. The European environment information and observation network (Eionet) is a partnership network of the EEA and its member and

[23] see https://gnss-metadata.eu/Guidelines/EPOS-Data_Policy.pdf

cooperating countries. Through EIONET, the EEA brings together environmental information from individual countries concentrating on the delivery of timely, nationally validated, high-quality data.

This knowledge is made available through the EEA website and forms the basis of both thematic and integrated environmental assessments. This information serves to support environmental management processes, environmental policymaking and assessment, as well as citizen participation[24].

Metadata access	Not explicitly mentioned
Data access	Fully, freely and openly available for others for EEA data
Data delivery	As soon as possible and with minimum delay both for third parties and the EEA
Cost	Free of charge or no more than cost of reproduction and delivery
Licensing	All data are ODC Attribution License (ODC-By) or similar as a standard. Third parties may provide their own licenses

4.2.7 IAGOS

In-service Aircraft for a Global Observing System (IAGOS) is a European Research Infrastructure for global observations of atmospheric composition from commercial aircraft. IAGOS combines the expertise of scientific institutions with the infrastructure of civil aviation in order to provide essential data on climate change and air quality at a global scale. Data are collected into a common database and published via a common web portal. IAGOS delivers a time and spatially resolved multi-component dataset on (ECV) and Air Pollutants[25].

Metadata access	Not explicitly mentioned
Data access	Free and unrestricted access after registration and upon acceptance of the data protocol.
Data delivery	Not explicitly mentioned
Cost	Free of charge or no more than cost of reproduction and delivery
Licensing	All data can be used free for scientific (non-commercial) use. Appropriate citation and an offer to co-authorship if the data play a significant role in a publication are required

[24] see https://www.eea.europa.eu/legal/eea-data-policy
[25] see http://www.iagos-data.fr/#CMSConsultPlace:DATA_POLICY

4.2.8 INTERACT

INTERACT is an infrastructure project under the auspices of SCANNET, a circum-arctic network of currently 86 terrestrial field bases. INTERACT specifically seeks to build capacity for research and monitoring in the European Arctic and beyond, and is offering access to numerous research stations through the Transnational Access program. INTERACT aims to the setup up of an interoperable distributed data infrastructure using the FAIR principles. The INTERACT data policy[26] specifies the delivery and access to data to/by the INTERACT Data Management System.

Metadata access	Free and open access to data to any person or organisation who requests them
Data access	Free and open access to data to any person or organisation who requests them. Restrictions may be applied to data supported by the exceptions on disclosure in the INSPIRE directive and the Directive 2003/4/EC
Data delivery	With minimum time delay given by the time needed for preparing data for publication, including quality assurance
Cost	Free of charge or no more than cost of reproduction and delivery
Licensing	no explicit licence mentioned, however, the attribution of data is required

4.2.9 LifeWatch-ERIC

LifeWatch ERIC is a European Infrastructure Consortium providing e-Science research facilities to increase knowledge and deepen our understanding of Biodiversity organisation and Ecosystem functions and services in order to support civil society in addressing key planetary challenges. LifeWatch ERIC, established as a European Research Infrastructure Consortium in 2017, runs three data portals for data discovery and access. A data policy is available for the Italian branch[27].

Metadata access	Freely accessible without any restrictions
Data access	Free and open access to data after registration. Embargo periods for data may be applied
Data delivery	not explicitly mentioned
Cost	Free of charge
Licensing	Metadata are available under a CC-0 license. The default data

[26] see https://eu-interact.org/app/uploads/2019/02/D4.4.pdf
[27] see http://www.lifewatchitaly.eu/web/lifewatch-italia/datapolicy

	license is CC BY-NC-SA 4.0. However, individual data providers may define own licenses

4.2.10NERC

The UK Natural Environment Research Council (NERC) commissions new research, infrastructure and training that deliver valuable scientific breakthroughs. NERC has a network of environmental data centres that provide a focal point for NERC's scientific data and information. These centres hold data from environmental scientists working in the UK and around the world. The data centres are responsible for maintaining environmental data and making them available to all users. NERC has a well-established data policy that outlines the roles and responsibilities of all those involved in collecting and managing environmental data[28].

Metadata access	Not explicitly mentioned
Data access	Full and open for NERC funded data for anybody. The only restrictions on access are those supported by the exceptions on disclosure in the Environmental Information Regulations (2004). If it is proposed to restrict access to any data it is explained why.
Data delivery	All environmental data of long-term value generated through NERC-funded activities must be submitted to NERC for long-term management and dissemination within two years of collection. Embargo times for special cases of maximum 2 years allowed
Cost	Free of charge. Large or complex requests where we may charge the cost of supply, or where third-party licence conditions either prevent such free supply, or require us to make specific charges
Licensing	All those who use data provided by NERC are required to acknowledge the source of the data. All environmental data made available by the NERC Environmental Data Centres will be accompanied by a data licence (UK OGL by default). Data originally provided to NERC by a third-party may have their own access and licence conditions which restrict how or when we can make data available to others, in which case our data licence conditions will reflect these.

4.2.11NitroEurope

NitroEurope is a project for integrated European research into the nitrogen cycle. NitroEurope is part of the EU's Sixth Framework Programme for Research and Technological Development, and ran from 2006 until 2011. NitroEurope integrates an observing system for N fluxes and pools, a network of manipulation experiments, plot-scale C-N modelling, landscape analysis, European up-scaling and uncertainty and verification of

[28] see https://nerc.ukri.org/research/sites/data/policy/

European estimates. The NEU science activity is divided into six components that will deliver different types of data to the NEU distributed database located at three dedicated data centres (DCs). The DC's responsibilities include storing, checking for completeness, maintenance and distribution of NEU data[29].

Metadata access	Not explicitly mentioned
Data access	Fully, freely and openly available for others for EEA data
Data delivery	As soon as possible and with minimum delay both for third parties and the EEA
Cost	Free of charge or no more than cost of reproduction and delivery
Licensing	All data are ODC Attribution License (ODC-By) or similar as a standard. Third parties may provide their own licenses

4.2.12TERENO

For the investigation of consequences of Global Change for terrestrial ecosystems and the socioeconomic implications TERENO created observation platforms on the basis of an interdisciplinary and long-term aimed research program with a close cooperation between several facilities of the Helmholtz-Gemeinschaft. TERENO provides long-term statistical series of system variables for the analysis and prognosis of Global Change consequences using integrated model systems, which will be used to derive efficient prevention, mitigation and adaptation strategies. A data policy[30] outlines the principles adopted by the TERENO initiative regarding its data policy. The data rights statement in this document provides general directives and rules on how to share, reuse and cite all TERENO data.

Metadata access	Fully, freely and openly available
Data access	Fully, freely and openly available for quality checked data after self-registration by the user. A part of the data may be restricted to certain user groups through additional registration. Embargo times for special cases are allowed
Data delivery	As soon as data is quality controlled and prepared for exchange, but not later than three months after its acquisition.
Cost	Free of charge
Licensing	Users may copy, distribute, display and perform the work for non-commercial and scientific purposes provided that they follow defined rules regarding citations and acknowledgements. Users may make derivative works based on TERENO data.

[29] see http://www.nitroeurope.eu/sites/nitroeurope.eu/files/neu_data/Component7/web/NEU_Data_Policy.pdf

[30] see http://teodoor.icg.kfa-juelich.de/downloads-de/TERENO%20Data%20policy.pdf

	They agree to make their results available to TEODOOR as far as they are able to do so.

4.3 Data Sharing Policy

On European as well as on global scale the LTER network collect and analyse long-term data sets to better understand patterns and processes in ecosystems and the interactions between ecosystems and socio-economic systems. In order to enable this, eLTER supports and encourages the full and open sharing of ecological and socio-economic data free of charge to advance research and education in the eLTER and beyond. Documentation and archiving of data for the benefit of present and future generations is one of the major goals as global environmental research becomes increasingly data driven.

eLTER policies for data sharing and accessibility consider principles and policies developed by other global networks, including the Global Earth Observation System of Systems (GEOSS, 2005), the Organization for Economic Co-operation and Development (OECD, 2007), as well as recommendations by the European Commission on Open Data (EC, 2017)[31] and INSPIRE (2007/2/EC)[32] that support sharing and open access to publicly funded research data.

Recommendation

eLTER encourages the public sharing of data and metadata. Data collected as the result of public funds being awarded to an eLTER entity should be made available online free-of-charge, with as few restrictions as possible, on a non-discriminatory basis. eLTER scientists should make every effort to release data in a timely fashion and with attention to accurate and complete metadata.

A first draft of a DSP was prepared by the eLTER data management team in 2016 and was handed over to eLTER project management. During the project, a discussion process was initiated between core eLTER partners, the eLTER project management and the eLTER data management in order to refine and to revise the draft document. The final draft document is attached as appendix to this deliverable.

The ***eLTER Data Sharing Policy*** is intended to provide ***general guidelines and recommendations*** for the development and implementation of a local data sharing policy on the level of member networks or research sites. It is also the main input for the further development of the data sharing policy and licenses of the eLTER RI.

[31] see http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/_h2020-hi-oa-pilot-guide_en.pdf

[32] see https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32007L0002&from=EN

Metadata access	**Metadata is always freely accessible**, without any restriction through the eLTER Data Portal or any related catalogue. Metadata access and usage is guaranteed under an open license agreement applying the Creative Commons Public Domain Dedication ("**CC0 1.0 Universal**"[33]). Metadata available through the eLTER Data Integration Portal (DIP)
Data access	Full and open for any data collected under the eLTER H2020 project or if requested by national research data obligations. Restrictions in data access might occur for third party data not collected in the funding frame of eLTER H2020.
Data delivery	All shared data, products and metadata will be made available with minimum time delay. The data sharing policy defines a possible embargo period of maximum 12 month.
Cost	Not addressed as this will be defined in the eLTER PPP
Licensing	**Datasets are openly available applying a Data Use License to all Datasets**, if not otherwise specified. The policy is implemented through the licensing of individual data sets by the originators, where such licenses exist, or through the general principles of this policy, namely - That all data are provided under the terms and conditions compliant with a CC-BY license unless stated otherwise.

[33] see https://creativecommons.org/publicdomain/zero/1.0/deed.en. This license allows to copy, distribute and perform work, even for commercial purposes, all without asking permission.

5 Governance for long-term implementation

5.1 Data Governance

In order to ensure the continued sustainable operation of the eLTER information system built during the eLTER H2020 project beyond the runtime of the project, a number of governance roles need to be defined interacting at different levels of scale. The following section provides a proposal in order to setup the governance structure for data management based on the lessons learned from long term monitoring projects and related research infrastructures. This encompasses roles as well as governance structures which need to be considered by future eLTER projects as well as the planned eLTER RI implementation programme.

Sufficient funding is required to establish and maintain the eLTER data infrastructure on a long term. This includes funding for personnel, hard- and software, cost for travel and networking events, education and training. A sufficient number of qualified staff is required to implement and operate the infrastructure through a clear system of governance.

The proposed structure has been tested within the operational arrangements of the TERENO network and is given as a potential starting point for establishing governance arrangements with eLTER RI head office. These arrangements would start if and when an eLTER Preparatory Phase Project (PPP) is awarded. However, implementation of these data governance structures needs to be considered alongside other roles and governance structures within such an eLTER RI (e.g. legal frameworks such as ERIC, contractual arrangements between RI head office and topic or service centres, and relationships with national site networks).

5.1.1 Roles and responsibilities

The basic task of the eLTER data governance is to ensure and organise an effective implementation of the eLTER data sharing infrastructure and the enforcement of the decided rules. In order to implement and run the eLTER Information System in the long run, a number of roles need to be defined and implemented. Figure 1 provides a first overview on the different roles as well as the potential interactions and communication flows.

Project Management (PM)

The eLTER Project Management is the representative of the project consortium. It takes the final decision on the appointment of positions in the DAB, DMSC, DCB and DC and is working in close cooperation with these boards.

Data Coordination Board (DCB)

The eLTER Data Management is managed operatively by the eLTER Data Coordination Board (DCB), a group of coordinators lead by one eLTER Data Management Coordinator, which is responsible for the coordination and operation of the eLTER data management

amongst the partners. Working together with the eLTER Data Management Steering Committee, eLTER Data Advisory Board, the eLTER Data Coordination Office and Data Centre (DC) and the working groups DCB develop plans for implementing the eLTER Data Management and ensure that the plans are carried out. DCB will plan and facilitate integration and synthesis exercises, including workshops, meetings and conferences. Furthermore, it will supervise the capacity building, knowledge transfer and external communication in conjunction with the DMSC. The DCB has the authorization to issue and enforce instructions for data sharing and provision on a technical level.

The DCB members are appointed by the eLTER Project Management after being put forward by the eLTER Data Managers Assembly. The number of members of the DCB, the term of office and the exact modalities of operation are established by the RI management.

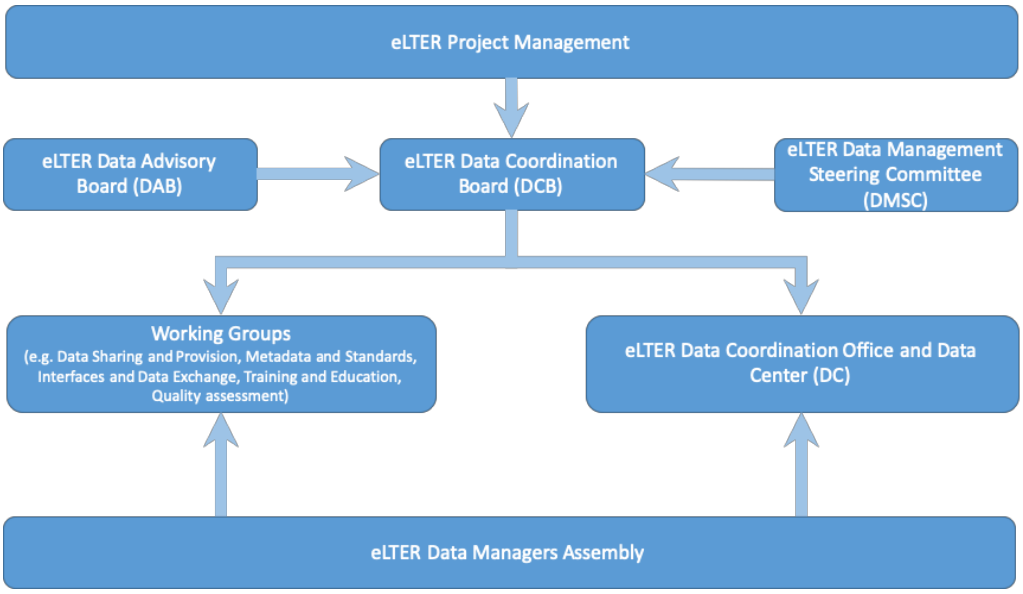


Figure 1 Organisational roles of eLTER data governance

Data Management Steering Committee (DMSC)

The Data Management Steering Committee (DMSC) consists of representatives from the eLTER project. Its main task is to decide the basic configuration of the eLTER data infrastructure, including the networking with external partners, and the integration of the infrastructure in national and international programs as well as the strategic planning future activities. New DMSC members are nominated by the eLTER Project Management. The number of members have to be fixed by the eLTER Project Management

Data Advisory Board (DAB)

The Data Advisory Board (DAB) consists of 12 members, which are nominated by the eLTER Project Management. The Advisory Board has the power to determine up to 3 co-opted members as appropriate. The members of the Advisory Board are appointed for a term of office amounting to 2 years with the possibility to be re-nominated. The Advisory Board nominates or elects a Chair and a Vice Chair from its members. Their term of office is

2 years with the possibility to be re-elected for a further term. The Advisory Board advises the Data Management Steering Committee (DMSC) with respect to current activities and infrastructural strategies of eLTER data management. The basis for these recommendations are yearly documents provided by the DMSC and the annual meetings at which the DMSC will present the current state of development in eLTER and provide an overview on current and planned projects and activities.

Data Coordination Office and Data Center (DC)

The eLTER Data Coordination Office and Data Center (DC) supports the DCB and is responsible for the central administration of eLTER Data Management. The office is the interface between the eLTER data managers, the working groups, the DCB, DMSC and DAB and the eLTER Project Management as well as the eLTER researchers, public and private partners. The DCO holds and operates the central eLTER data infrastructure components and web portals. Among the tasks handled are:

- Central facilities, databases, web portal
- Organization of central events and regular meetings
- Public Relations (Website, press-releases, etc.)
- Contact to external partners
- Corporate Design
- Quality Management

Data Managers Assembly (DMA)

Each eLTER partner should nominate at least one responsible contact person dealing with eLTER related data management issues in the institution. Periodically, at least once a year, an assembly of the responsible data managers will be convened to discuss current issues in eLTER data management and to decide on related issues.

Working Groups (WG)

The Working Groups provide the particular scientific and technical knowledge which is essential to plan, realize and develop the eLTER data infrastructure. They will be composed of data managers, data scientists or experts from the eLTER partner institutions, but might also include experts from universities and other research organisations. The working groups work closely together with the DCB and DC. The most important working groups proposed here are:

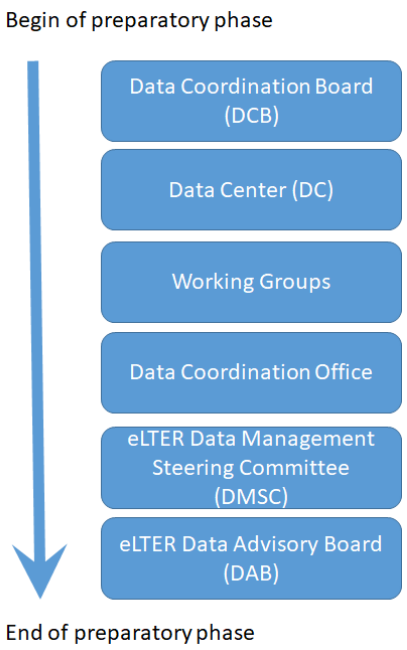
- The working group "*Training and Education*" conceptualizes and plans measures to train and educate the eLTER partners in questions of data sharing, provision, standards, discovery and usage. The group plans outreach activities with regard to knowledge transfer and the development of information products.
- The working group "*Interfaces and Data Exchange*" works on solutions to for the implementation and usage of interfaces to share data and metadata within or outside of the eLTER data infrastructure and on data sharing and provision issues on a technical level.
- The working group "*Metadata and Standards*" defines data and metadata standards and develops metadata profiles for the documentation and sharing of (meta)data and within eLTER. The group is also responsible for the selection of common vocabularies. A variety of different thematic groups within eLTER (atmosphere, land surface, soil, water, remote sensing, ecology, etc.) should be represented as group members.

- The working group "*Data Sharing and Provision*" deals with questions of data policy related issues concerning data sharing and provision. The group develops workflows e.g. for data provision, control mechanisms for policy compliance.
- The working group *"Quality assessment"* develops concepts for the quality assessment of metadata and data within eLTER

Other working groups than the ones addressing above might be set up.

5.1.2 Implementation

Implementation of the governance structure and its final design will be negotiated with and set into force by the eLTER Project Management. As a recommendation based on the experiences from other data infrastructures (e.g. TERENO, WASCAL), implementation of the proposed governance structure should be done in consecutive steps during the preparatory phase as indicated in the right figure.



In the beginning of eLTER PPP, the Data Coordination Board is implemented by the Project Management. In parallel, the Data Center is founded. Once both are implemented, the working group topics will be selected and implemented in decreasing order of priority for infrastructure implementation. With increasing coordination efforts, a Data Coordination Office will be implemented. At the end of the Preparatory phase both the Steering Committee and the Data Advisory Board will be founded.

At the end of the Preparatory phase, all required institutions for data management governance should be set up and operative.

5.2 Data Sharing Policy

5.2.1 Implementation

As already described in chapter 4, a first draft of a Data Sharing Policy (DSP) was prepared by the eLTER data management project partners and handed over to eLTER project management in the initial phase of the eLTER H2020 project. During the project, a participatory process was initiated between the eLTER partners, the eLTER project management and the eLTER data management to refine and to revise the draft document. The final draft DSP is attached as ***Appendix*** to this document. However, the data policy was not set into force during eLTER H2020 project.

Target of the PPP is the negotiation and implementation of a common eLTER Data Sharing Policy. For this purpose, a specific working group within eLTER will be created. In a first step, the relevant DSPs from project partners (if available) will be collected and evaluated with respect to metadata access, data access, data delivery, cost and licensing. In the next step, a standardized license for data sharing attached to each data set and a workflow how to handle different licenses issued by individual partners will be negotiated.

If necessary, the existing draft eLTER Data Sharing Policy will be modified and handed over again to eLTER project management, which will initiate the final acceptance of the document through signing by current eLTER partners and putting the DSP into force.

On a long term, the Data Coordination Board develops workflows for the ratification, implementation and surveillance of the eLTER DSP by new eLTER partners. If required, the Data Coordination Board evaluates the DSP periodically for necessary changes and updates.

5.3 Data management plan

Data Communication and Data Management are key issues of interdisciplinary research projects. The overall success depends on the well-organized data management and data exchange between the involved partners. Data management facilitating the acquisition, provision, integration, management and exchange of heterogeneous digital resources within a scientific and non-scientific multiuser (distributed) environment has to

- reflect the organizational/institutional constraints of the project members and stakeholders
- consider both data processing (research) and data security requirements
- ensure the long-term persistence of the data sets
- be flexible and upgradeable
- meet common data and metadata standards
- meet common data transfer protocols and mechanisms for project data and metadata

Data Management Plan and data policy together describe how data are provided, stored, shared and managed within the eLTER infrastructure. Whereas the Data Sharing Policy focuses on policy issues, the DMP relates more to technical issues on data management and data sharing like:

- General architecture of the data infrastructure is described
- Data management responsibilities and user groups within the distributed eLTER research infrastructure will be defined
- Initial description of the overall data stock to be produced
- Data types and data formats to be supported within eLTER based on the FAIR principles. Open formats will be preferred.
- Standards and metadata: file naming conventions and exchange formats depending on data analysis and modelling requirements are defined. The data will be indexed by meta data using common standards.
- Data curation processes like the establishment of homogenized quality control

- Data provision and sharing: workflows are established to harmonize preparation and submission of data and metadata to the data bases. Mechanisms to tackle data ownership, copyright and intellectual property rights issues are addressed.
- Archiving and preservation workflows (including storage and backup)

The target audience for this plan is all project members providing and using data and data products. The plan acts as an agreement between data users and data providers/creators in the framework of eLTER, on how to prepare, use, store, disseminate and publish data generated by the project. On the basis of the data management plan the technical eLTER infrastructure will be implemented and set up in each partner institution. However, the federated structure of eLTER requires a modular system of data management plans.

A ***General Data Management Plan*** [gDMP] for eLTER as a whole defines the general setup of the infrastructure and the basic general rules to be applied for the central and federated components, like interfaces, standards, metadata provision, common vocabularies, registries, quality control, data provision and sharing.

Partner specific Data Management Plans [pDMP] describe in detail, which kind of data are collected at the individual sites, how data management is organized and how the data are processed, described, managed and shared within or outside of eLTER. The Partner DMPs have to refer and have to be compatible to the requirements of the General DMP. Each eLTER site/project/partner contributing to the infrastructure has to provide an individual DMP, which needs to be approved by the eLTER Data Coordination Board.

5.3.1 Implementation

A first draft of the eLTER Data Management Plan has already been created in the onset of the eLTER H2020 project as a basis for data infrastructure implementation, data integration and data sharing. However, since infrastructure development is a continuous process, the DMP needed to be modified as well. The current version of the DMP is attached to this deliverable as ***Appendix*** to the current deliverable.

During the eLTER PPP project, further modifications of the main eLTER DMP and adaptations to the current implementation are necessary. For this purpose, a working group “Data Management Planning” will be implemented by the DCB in the beginning of the PPP project. This working group is responsible for periodic updates of the Main DMP in close cooperation with other relevant working groups, the DCB and the DC. Secondly, the group evaluates and selects available tools for the generation of modular DMPs (e.g. RDMO, DMPOnline, DMPTool) to be filled by the eLTER partners and defines templates for the modular DMPs. These tools and templates will be made available to the eLTER partners by the DC. The working group and the DCB will develop workflows on the delivery, assessment, verification and approval of the modular DMPs.

On a long-term, the frequency of necessary modifications of the DMP and the amount of new modular DMPs will decrease. Therefore, the DCB and/or DCMB may decide, whether or not the responsibilities for updating and managing DMPs will be transferred to the Data Centre.

5.4 Integration of data and data services

eLTER is designed a federated data infrastructure, so integration of data services is a key part of delivering functionality to both creators and users of data. eLTER assumes that data will be stored in existing partner data systems that must interface to the eLTER infrastructure as well in the central data stores of eLTER (e.g. DEIMS, DIP and Central Data Node (CDN)). This design is based on the selection and operation of key data standards for communication and transfer of data between the different components of the infrastructure. To help partners that do not have the resources and/or skills to host standardised data infrastructure, cloud-based data storage and services are made available for hosting data through the CDN. The use of recognised data standards and the provision of cloud-based facilities to provide eLTER data services will improve the sustainability of the eLTER infrastructure as we will enable users of the services to take advantage of ongoing development in these standards and tools as eLTER evolves.

The specification of the data services required and the appropriate standards and host providers will be taken through the eLTER Data Coordination Board guided by the Advisory Board. The key central services for the cataloguing, registration, evaluation and access of data will be controlled by the eLTER Data Centre as part of the ESFRI Head Office services. The refining of service development will have to take place in close consultation with eLTER stakeholders through steering groups and the envisage Topic Centre on the eLTER RI. Within the design of any such RI will be the specification of the quality of service required from the eLTER data infrastructure and establish service level agreements with service providers. This will include cloud services providers as envisaged by the EOSC that eLTER wishes to work with in designing new EOSC data services.

The sections below describe the principles and current implementation of the eLTER data infrastructure and how existing service standards allow data to move through from data depositor to data user.

5.4.1 Site registry Service

The registration of sites in eLTER is done using DEIMS-SDR. DEIMS-SDR (Dynamic Ecological Information Management System - Site and dataset registry, https://deims.org) is an information management system that allows discovering long-term ecosystem research sites around the globe, along with the data gathered at those sites and the people and networks associated with them. DEIMS-SDR describes a wide range of sites, providing a wealth of information, including each site's location, ecosystems, facilities, parameters measured and research themes (Wohner et al. 2019).

Accounts for DEIMS-SDR can be requested by anyone and are provided upon approval by the eLTER Data Coordination Office. The necessary workflows will be developed and implemented during the preparatory phase. Once an account has been provided, one or multiple site records can be created, which should contain the following information.

- General characteristics ('General Site Description', 'Coordinates', 'Site Type', and 'Size')
- Climatic characteristics ('Mean Annual Air Temperature' and 'Sum Annual Precipitation')
- Topographic characteristics ('Elevation Range (minimum – maximum)')

- Ecosystem characteristics ('Biome' and 'Ecosystem and Land Use')
- Scientific characteristics ('Purpose of Site', 'Research Topics', 'Design of Observation', 'Scale of Observation', 'Design of Experiments', 'Scale of Experiments' and 'Observed parameters')
- Operation characteristics ('Year Site was established', 'Site Status (active, inactive, closed)', 'Permanent Operation', 'Accessible All Year', and 'Permanent Power Supply')
- Data management ('Data Request Format' and 'Data Storage Location')
- Metadata information ('Metadata provider')

Once a comprehensive site record including LTER network affiliation has been provided, it can be accredited by the network manager of the respective LTER national network. The resulting information of all eLTER site records not only allows more efficient information management, but also allows in-depth analyses of the capabilities of the network (Mollenhauer et al. 2018).

It is also possible to add information datasets and so called data products to each site. This can but does not have to include the actual data itself or the links to external services or storage locations.

Site information can be displayed in a human readable way as a regular html page (https://deims.org/8eda49e9-1f4e-4f3e-b58e-e0bb25dc32a6) or in a machine-readable way in either Inspire EF or ISO19139. This information is also provided to all subsequent systems that are connected to DEIMS-SDR and the Data Integration Portal (DIP), such as the GEO Portal (http://www.geoportal.org/) and DataONE (https://www.dataone.org/). Rudimentary information about sites is also provided using WMS and WFS (https://deims.org/geoserver/web/).

The management of site information using DEIMS-SDR is also positioned in other projects than eLTER, e.g. LTER CWN, EcoPotential, ENVRI Plus and ENVRI Fair as well as in the ILTER community, ensuring the sustainable long-term usage of the collected data beyond the scope of the eLTER project and providing this information to stakeholders inside and outside the LTER community, e.g. Zilioli et al. 2019. Long-term usage of each site record is further fostered by the provision of a persistent, resolvable and unique identifier, called DEIMS.ID, that exceeds the frame of the eLTER project as well as the LTER community allowing unique, cross-RI identification.

5.4.2 Federated metadata catalogue and visualisation

All external data and metadata providers are connected by the eLTER Data Integration Portal (DIP)[34]. The DIP is hosted at BioSense Institute and is based on a modified version of 52^0 N SOS server and GeoNetwork. Frontend runs on a virtual machine, database on another, and it is exposed publicly through third VM which acts like a proxy and firewall. The DIP itself does not require a user login in order to search/download the data/metadata. All data will be retrieved from remote servers on demand, which means that there is no data replication. In order to provide faster search, metadata are harvested from different sources like DEIMS-SDR, CDN, etc. The DIP if compliant with the GDPR policy, because all metadata are harvested from known sources which also takes care about GDPR.

[34] see http://dip.lter-europe.net/.

The DIP is also harvested by other systems like DataONE, GEOSS etc.

5.4.3 Data nodes

eLTER currently uses the UK Science Cloud (JASMIN) hosted at STFC Harwell to provide virtual data infrastructure. The virtual machines use containerised components to provide a software tools such as the 52^0 N for SOS services and GeoNode for CWS services. The services provided comply with the EU INSPIRE regulations implemented through OGC standards. The eLTER Central Data Node is provided as virtual machines within this facility to host partner site data and also pan-network data products such as CORDEX regional climate projections. For data curators who want to use the CDN to allow their data to be accessible through INSPIRE services, a series of forms are provided in which they can register as CDN users, register time series data services within the CDN and upload data to though services. Once registration and upload is completed, these data can be visualised and access through SOS services clients such as the eLTER DIP or any other compatible data portal.

This configuration provides options for sustainability of the data infrastructure through services provided by local host institutes and central cloud facilities depending on what is most appropriate for data providers to meet eLTER policy for continued data access. Data providers can start by using the eLTER CDN and move to host institution infrastructure as available or visa versa if host resources become unavailable. For control of key pan-network data products, eLTER will maintain CDN services to ensure access to these data.

5.4.4 Semantic and workflow harmonization

In order to be able to share data within a distributed network, not only common interfaces but also and in particular standards with respect to semantics and workflows have to be defined. Semantical standards may be common vocabularies, like already existing thesauri (e.g. EnvThes) or new and/or extended thesauri for special purposes, common metadata profiles (ISO 19115, sensorML) or common file structures and/or formats.

Similarly, common workflows have to be developed with respect, e.g. to quality control of data and metadata, metadata publication, compliance control for data, metadata and services, publication issues or data provision. In particular, workflows to handle violations of regulations and agreements (data policy, data and metadata provision, license violations, standards etc.) need to be developed.

During the preparatory phase, the eLTER Data Coordination Board together with the Data Center will initiate the process to define and implement common standards and workflows. If necessary, Working Groups will be established to elaborate detailed aspects.

5.4.5 Implementation

Currently, data integration takes place in the description of data sets and services within the metadata standards within catalogue services and data delivery services using SOS and WFS within DEIMS-SDR, DIP and CDN. This means that data sets and related sites and

sensors can be described consistently and shared between eLTER data infrastructure components as well as with other portals using ISO standards. The data themselves can be visualised together using the SOS client tools in the DIP and downloaded as CSV files containing time series data.

These basic services for finding, integration and access to data will be extended to spatial data types (using WFS) and particular modelling community data types such as delivery of NetCDF files in future RI work. Currently any data types can be catered for through deposit of data files within the DEIM-SDR archive with a suitable ISO data description. Metadata records relating to this archive are harvested by the eLTER DIP so can be discovered through this portal. The DIP also uses the WFS service from DEIMS to access geographic information about sites and data locations.

These data services shared between eLTER components are also currently open to access by any users wanting programmatic access (e.g. through Python or R). This functionality has been presented at eLTER workshops but is not more widely advertised at present. If future development of RI service governance, we will implement API management systems to control access and quality of service of third-party access to these services. However, encouraging the use of these APIs to create analytical workflow is seen as a crucial element of data integration across eLTER data and also with other API accessible data (e.g. EO or wider in-situ measurements such as biodiversity).

5.5 Integrating new partners

A central issue in developing a data infrastructure is the integration of new partners. In eLTER H2020 different access points into eLTER RI have, e.g. DEIMS, DIP, Data Nodes) have been defined, developed and implemented to allow data providers with different skills easy and adapted connection points to eLTER. The system was tested on the basis of “showcase” data providers.

The process of integration new partners will be initiated by a formal application of the partner. After a principle decision of the Data Coordination Board the integration process will be initiated. Integrating new partners into eLTER RI may be a problematic issue due to several reasons. In the beginning new partners needs to be trained and educated with respect to the options for data provision on a technical level, but also on the requirements for data management, adherence of common standards and profiles and data policy issues. It can be expected that there will be serious distinctions and incompatibilities between the requirements by eLTER and the standard data provision by the new partner.

Based on a Data Management Plan created by the new partner a negotiation process will be initiated with the aim to identify the adaptations to be made, either by eLTER or the partner, and the resources and time periods needed to adapt the systems. The Data Center is in charge to guide through this process. An integration plan will be created by the Data Center and the new partner and will be implemented after approval by the Data Coordination Board.

New partners will assess whether they want to connect existing data infrastructure services from their institution directly to eLTER (e.g. through OGC services like SOS) or make use of the CDN to provide such services for their data. If they decide to use the CDN, they will use

DEIMS-SDR as usual to register sites and sensors but point these descriptions to services provided by the CDN. They will create these CDN services through a set of CDN registration and data upload forms provided and once completed, these data will be accessible through the eLTER DIP.

5.6 Training and capacity building

Training and capacity building is an important issue to connect the eLTER partners, i.e. data providers, data managers and data users. Training and capacity building measures will be initiated and organized by the eLTER Data Coordination Office on behalf of the Data Coordination Board. If necessary, a Working Group will be established to develop a catalogue of measures for training and capacity building. This catalogue may, include workshops, trainings, user guides as well as an online help. In many cases, personal support (on-site support) by data management staff will be required, especially in the initial phase of eLTER PPP.

5.7 Planning and risk management

Medium-term (three to five years) and long-term (> five years) plans will be created to ensure the continued availability and accessibility of the data. In particular, both the response to rapid changes of circumstance and long-term planning should be described, indicating options for relocation or transition of the activity to another body or return of the data holdings to their owners (i.e., data producers). Important issues to be addressed are what will happen in the case of cessation of funding of a particular partner or who could take over responsibilities for data publication in case one or more eLTER sites stop data delivery.

6 References

Berners Lee, T. (2006), Linked Data - Design Issues. Online http://www.w3.org/DesignIssues/LinkedData.html [Last Accessed 12.6.2019 14:30]

GEO (2010) GEOSS Data Sharing Action Plan. Document 7 (rev2). Online https://www.earthobservations.org/documents/geo_vii/07_GEOSS%20Data%20Sharing%20Action%20Plan%20Rev2.pdf [Last Accessed 12.6.2019 15:00]

Mollenhauer, H., Kasner, M., Haase, P., Peterseil, J., Wohner, C., Frenzel, M., … Zacharias, S. (2018). Long-term environmental monitoring infrastructures in Europe: observations, measurements, scales, and socio-ecological representativeness. Science of The Total Environment, 624, 968–978. https://doi.org/10.1016/j.scitotenv.2017.12.095

Open Knowledge Foundation (2019) Open Data Handbook. Online http://opendatahandbook.org/guide/en/ [Last Accessed 12.6.2019 14:45]

Watkins, J., Ciar, D., Wohner, Ch. Peterseil, J., Schentz, H., Oggioni, A., Lanucara, S., Minić, V., Škrbić, S., Bodroški, Z., Kunkel, R., Sorg, J. (2016) D8.1 eLTER Information Architecture Report. eLTER Project report (H2020, GA-Nr. 654359), [http://www.lter-europe.net/document-archive/elter-h2020-project-files/d8-1-it-design, last accessed 11.4.2018]

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M. , Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, Ph.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, Ch.T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J.G., Groth, P., Goble, K., Grethe, J. S., Heringa, J., 't Hoen, P. A.C, Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., ons, A., Packer, A. L., Persson, B., Rocca-Serra, Ph., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, Th., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., Mons, B. (2016): The FAIR Guiding Principles for scientific data. management and stewardship. Scientific Data 3 [2016/03/15/online; http://dx.doi.org/10.1038/sdata.2016.18]

Wohner, C., Peterseil, J., Poursanidis, D., Kliment, T., Wilson, M., Mirtl, M., & Chrysoulakis, N. (2019). DEIMS-SDR – A web portal to document research sites and their associated data. Ecological Informatics. https://doi.org/10.1016/j.ecoinf.2019.01.005

Zilioli, M., Oggioni, A., Tagliolato, P., Pugnetti, A., & Carrara, P. (2019). Feeding Essential Biodiversity Variables (EBVs): actual and potential contributions from LTER-Italy. Nature Conservation, 34, 477–503. https://doi.org/10.3897/natureconservation.34.30735

European Long-Term Ecosystem and Socio-Ecological Research Infrastructure

Data Management Plan (Draft)

Authors: eLTER Data Management Team
Lead partner for deliverable: Umweltbundesamt GmbH
Other partners involved: UKRI-CEH, BSI, FZJ, CNR

Contributor

*Peterseil Johan*nes, Umweltbundesamt GmbH, Austria (WP3 lead)
Watkins John, NERC-CEH, UK (WP8 co-lead)
Kunkel Ralf, FZ Jülich, Germany
Wohner Christoph, Umweltbundesamt GmbH, Austria
Minic Vladan, BioSense Institute, Serbia
Rennie Sue, UKRI-CEH, UK
Oggioni Alessandro, CNR-IREA, Italy (WP3 co-lead)
Crnojevic Vladimir, BioSense Institute, Serbia (WP8 lead)

Content scheme

The current version of the initial data management plan is based on the recommendations of the Data Curation Center [DCC, 2013] and the provided Data Management Planning Tool "DMP Online"[1].

H2020-funded project, GA: 654359, INFRAIA call 2014-2015
Start date of project: 01 June 2015 Duration: 48 months
Version of this document: Version 2.0 Submission date: 2019-06-30

Dissemination level

PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	
CI	Classified, as referred to in Commission Decision 2001/844/EC	

[1] Based on https://dmponline.dcc.ac.uk/projects/elter-h2020/plans/11410/edit

Version control	Edited by	Date of revision
Created - V1	Data Management Group	2015-11-10
Internal review	Data Management Group	2016-05-31
Internal review		
Revised – V2	Data Management Group	2016-06-109
Revised – V3	Data Management Group	2019-05-09
Revised – V4		
Reviewed		
Revised – V5		
Signed off – coordinator	Data Management Group	2019-06-30

Publishable Executive Summary

Data Communication and Data Management are key issues of interdisciplinary research projects. The overall success depends on the well-organized data management and data exchange between involved partners. Data management facilitating the acquisition, provision, integration, management and exchange of heterogeneous digital resources within a scientific and non-scientific multiuser (distributed) environment has to

- reflect the organizational/institutional constraints of the project members and project stakeholders
- consider both data processing (research) and data security requirements
- ensure the long-term persistence of the data sets
- be flexible and upgradable
- meet common data and metadata standards to be compatible with other data management systems
- meet common data transfer protocols and mechanisms for project data and metadata

A first draft of the eLTER Data Management Plan has been created already in the onset of the eLTER H2020 project as a basis for data infrastructure implementation, data integration and data sharing. However, since infrastructure development is a continuous process, the DMP needed to be modified. The document provides version of the Data Management Plan following the recommendations of Data Curation Centre [DCC, 2013].

Content

Glossary

B2SHARE	EUDAT Data Repository Service
CDI	Central Data Infrastructure
cDN	eLTER Central Data Node
CSW	OGC Catalogue Service Web
CV	Controlled vocabulary
DEIMS-SDR	Dynamic Ecological Information Management System - Site and Dataset Registry
DIP	Data Integration Portal
DN	Data Node
eDN	eLTER existing Data Node
EF	INSPIRE Environmental Monitoring Facilities
EF	Environmental Monitoring Facility
eLTER IS	eLTER Information System
EML	Ecological Metadata Language
GET-IT	Geographic Enabled ToolkIT (GET-IT) Starter Kit ®
MD	Metadata
OGC	Open GeoSpatial Consortium
PID	Persistent Identifier
SOS	OGC Sensor Observation Service
SWE	OGC Sensor Web Enablement
TA	Physical Transnational Access to research sites
UUID	Unique Identifier
VA	Virtual Access to data
WCS	OGC Web Coverage Service
WFS	OGC Web Feature Service
WMS	OGC Web Mapping Service

1 Project context

The eLTER H2020 project will serve as the flagship for the further development of the Long-term Ecosystem Research infrastructure and community in Europe closely interacting with the eLTER ESFRI process as well as LTER Europe network (and the national LTER networks). The overall aim of the eLTER project is to advance the European network of Long-Term Ecosystem Research sites and socio-ecological research platforms to provide the highest quality services for multiple use of a distributed research infrastructure. Major objectives are to:

- identify user needs for the research infrastructure in relation to major societal challenges
- streamline the design of a cost-efficient pan-European network, able to address multiple ecosystem research issues
- develop the organisational framework for data integration and enable virtual access to LTER data
- foster the societal relevance, usability and multiple use of information, data and services through partnerships with providers of remotely sensed data, analytical services and scenario testing models, and via the adoption of new measurement

The specific role of eLTER H2020 is to catalyse conceptual and service developments of a distributed, highly integrated and widely used research infrastructure to support a wide range of ecosystem and critical zone research questions, such as:

- How are ecosystems and biodiversity changing or adapting to global-change stresses?
- What are the determinants of ecosystem resilience?
- What are the threshold interactions resulting in system shifts?
- How can we respond locally, nationally and internationally to support systems that are more resilient to global change effects?

Inter-alia the documentation of the European long term research infrastructure as well as access to existing long term data is aimed by the eLTER H2020 project. eLTER H2020 will use existing observation facilities[2] in order to newly generate and provide access to existing data. eLTER H2020 comprises the following components with relation to data delivery and data use:

- Concerted data delivery of selected 162 LTER sites (Virtual Access);
- Supported access to 18 top ecosystem research sites (Transnational Access);

[2] See http://www.lter-europe.net/elter/about

- Research infrastructure integration (nationally, Europe, globally).

Data usage and access is designed along with the four exemplary scientific use cases with increasing level of complexity to assess LTER data quality and services. These will also act as proof of concept for the resulting technical solutions. The architecture of the common eLTER Information System and tools will be developed alongside the use cases. The results will feed into the further specification of the eLTER design and the planning eLTER ESFRI implementation.

2 Overview on the eLTER Information System

2.1 Data architecture

LTER Europe (by the means of eLTER H2020) aims to provide a frame for an integrated discovery and access to metadata resulting from the distributed network of data providers within the LTER Europe network. The Data Integration Platform (DIP) is developed as a "one-stop-shop" interface for end users (being researcher or expert) integrating metadata from the different distributed data sources (the data providers / nodes).

Those data provider / nodes can either provide

- an existing service endpoints (in terms of MD endpoint, data storage and services; advanced capabilities; e.g. TERENO) or
- a deployment/clone of the standard virtual data node (DN), which has been developed during the H2020 eLTER project. The Central Data Node is a centrally managed instance of these virtual data nodes which can be used across the network an RI.

The LTER Europe Information System aims to implement standards wherever possible, being ISO19115/19139 and EML for MD, ISO19156 (O&M - Observation and Measurement) for observation data, and INSPIRE EF (Environmental and Monitoring Facility) data specification for site documentation.

The *eLTER* Information System is composed of different components. Figure 1 show a logical architecture with an overview of the components as well as their linkages. The main components are:

The controlled vocabulary **Environmental Thesaurus** [EnvThes][3] (and related services) is be used by the different components of the eLTER Information System. EnvThes provides the main semantic source for keywords and parameter names. It is used to annotate keywords to datasets during the metadating process. In future, also the central registration of reference lists should be possible.

The **DEIMS-SDR - Site and Dataset Registry** [DEIMS-SDR][4] is a central service provided in order to manage metadata on relevant information entities of the research infrastructure (e.g. network, research site, dataset, person, sensor, station).

The **Data Repository** or **Data Node** [DN][5] (either local or central) provides services to manage, archive and share data from long term observation and experimentation. The data

[3] see http://vocabs.lter-europe.net/edg/tbl/EnvThes.editor

[4] see https://deims.org/

[5] see https://cdn.lter-europe.net/

can either be stored in a database, single files stores or already published applying time-series or geo-spatial services (e.g. TERENO using OGC SOS)

The **Data Integration Portal** [DIP][6] harvests and integrates metadata from different provider nodes and acts as "one-stop-shop" for the discovery and access of metadata and data. The DIP also could include the digestion of data into intended data formats (based on user requests). The final specification and implementation of the DIP will be done during the H2020 eLTER project.

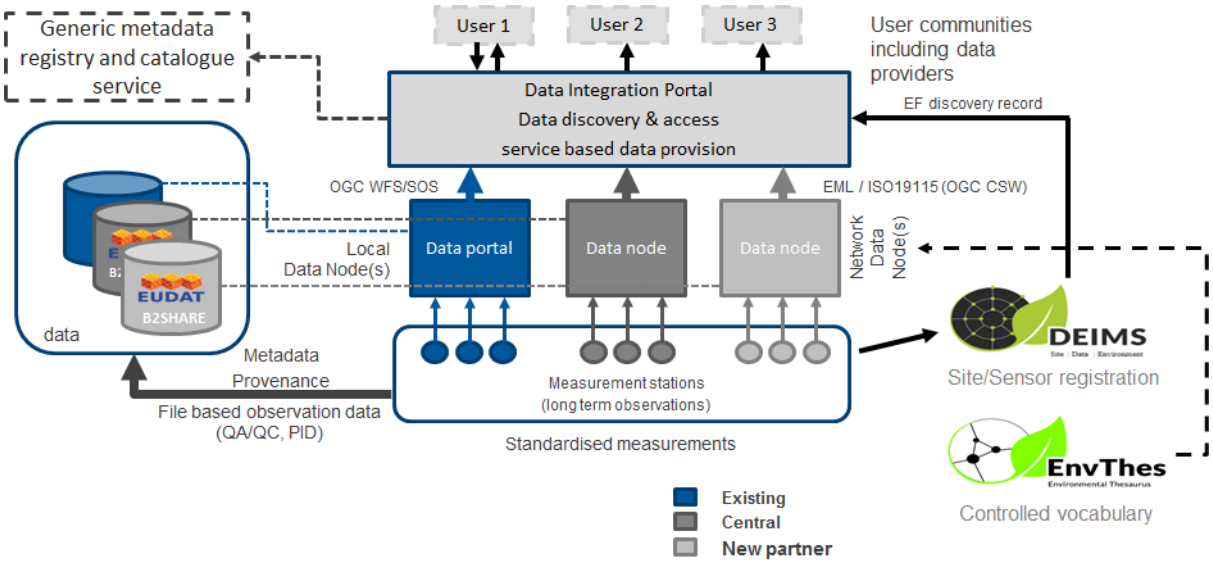


Figure 1 Overview of logical architecture for eLTER

Communication and information transfer between the data nodes and the central discovery portal is done via standardized service interfaces (e.g. OGC CSW for metadata) enabling machine-to-machine communication. This logical architecture (see *Figure 1*; Watkins et al. 2016[7]) aims to provide a scalable solution for data management and data exchange within the LTER network. In this way, different data provider / nodes can be registered in other networks, e.g. European or national sectoral portals without any additional efforts. The eLTER Data Integration Platform (DIP) will provide the central node to link LTER data into other networks (e.g. DataOne, EUDAT, GEOSS, etc.)

A central component to support the user needs is the development of a software suite which is coupled into a **eLTER Central Data Node** [cDN] which is composed of different components:

[6] see http://dip.lter-europe.net/

[7] see https://www.lter-europe.net/document-archive/elter-h2020-project-files/d8-1-it-design

Metadata Component providing standard metadata for the discovery and re-use of the data. The Site documentation is based on DEIMS-SDR. The documentation of the dataset (discovery) can be done either via DEIMS-SDR or shared through a local metadata system using e.g. pyCSW. Exchange of metadata is based on OGC CSW standard enabling metadata harvesting via Rest-API or OAI-PMH. Description of sensor metadata is done locally using SensorML standard.

Data Storage Component storing the observation data in defined dataset or time series. This can be composed of a database (e.g. for time series sensor observation and geospatial data using 52°North Software Suite) or/and a file repository (linking e.g. to EUDAT B2SHARE).

Data Web Service Component providing an service based access to data via e.g. OGC SOS) for dedicated data. Only time series data shared through the eLTER CDN can be accessed as OGC SOS service.

The different software tools are packaged as virtual machine, being able to be distributed to data providers.

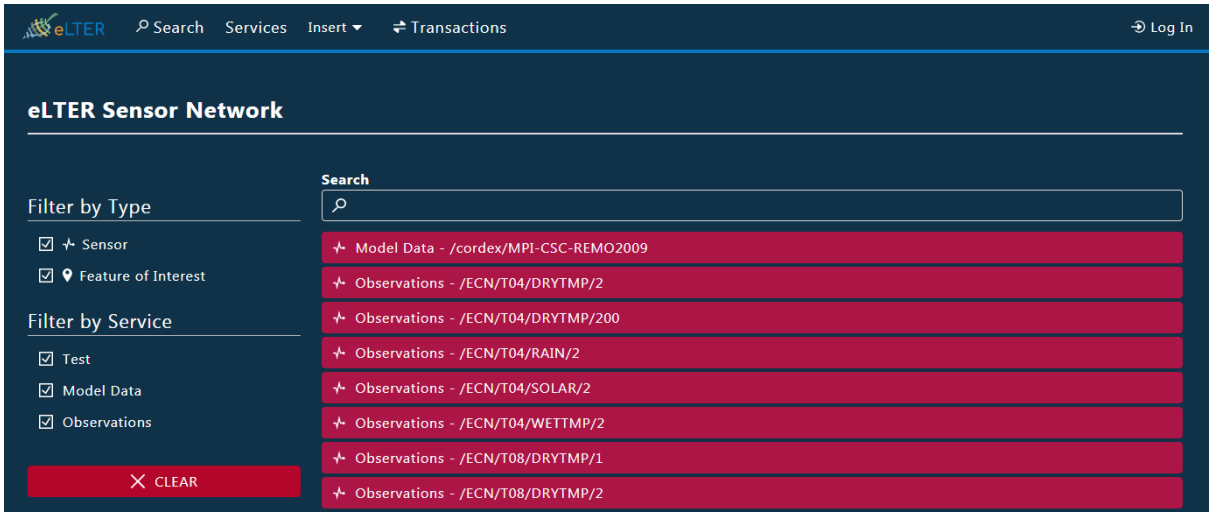


Figure 2 Landing page for the eLTER Central Data Node [CDN]

Existing data nodes [eDN] provided by partners (e.g. TEODOOR (LTER Germany), NextData (LTER Italy), iEcoLab (LTER Spain)) are composed of similar components and are providing standard interfaces (e.g. OGC CSW and OGC SOS) to link and ingest metadata and datasets into the eLTER Information System (see Figure 1).

The Data Integration Portal [DIP] harvests existing metadata from the network of data nodes, including the DEIMS-SDR and eLTER CDN as central components, and provides an integrated view on the metadata. In addition it provides user interfaces based on 52°North Helgoland client to visualise and subset data based on OGC SOS web services. The eLTER DIP is built on existing technology having geonetwork, EUDAT B2FIND and 52°North

Helgoland client as the main underlying software modules (Minic et al.2018)[8] . For harvesting of the metadata the CSW standard will be used.

The eLTER Information System is under development, with different development stages of the different components.

2.2 User architecture

Within the project context until the full implementation of the data management infrastructure of the eLTER RI a simplified model will be implemented. These are defined in the current DMP.

2.2.1 User roles and responsibilities

In the eLTER context four basic roles can be defined describing the range of central data management and distributed repositories. The specification and contacts for the different roles is given in Annex 1 - Central roles and functions.

Data Management Team [DMT] - responsible for the overall coordination and maintenance of the eLTER Information System. Within the DMT two sub-roles can be defined:

> **DMT Coordination** - the coordination team is compiled of WP leads and main contributors from WP3 and WP8 of the eLTER H2020 project. Until the final implementation of the eLTER RI, the DMT Coordination is the main body to make the final decision on all subjects regarding data management in cooperation with LTER Europe and the eLTER RI implementation team.
>
> **DMT Technical Support and Maintenance -** features the maintenance and modification of the components of the eLTER Information System. For each of the component one person is dedicated to the "Technical Support and Maintenance" team. This is true for the following components: a) DEIMS-SDR, b) eLTER CDN, c) eLTER DIP, and d) EnvThes.

Local data manager [LDM] - any party or entity responsible for the management of a partner data node. The partner nodes are linked to the eLTER Information System using standardised interfaces and services (OGC CSW and OGC SOS). The Local Data Manager is the main contact point for any socio-technological issues related to the partner node.

Data provider [DP] - any party or entity providing data (of any resolution) to the eLTER Information System (either being contracted within the eLTER H2020 project or voluntarily).

[8] see https://www.lter-europe.net/document-archive/elter-h2020-project-files/d8-3-dip

This is the contact point to any data related questions. The data provider needs to specified for each dataset and data services in the discovery metadata.

Data user [DU] - any entity using data provided by the eLTER Information System.

2.2.2 Target audience and permissions

The usage of the eLTER Information System with special regard to the provision of data is intended for and open to everyone involved in long-term ecosystem research on local, national, regional or global scale. Nevertheless, for the different components and subsystems of the eLTER Information System special rules may apply.

EnvThes – The common and shared vocabulary is open to be used by any interested person. New terms can be submitted to the EnvThes Editors team which is responsible for the maintenance and development of EnvThes.

DEIMS-SDR - DEIMS-SDR is open for everyone. The target audience of DEIMS-SDR are scientists and site manager in the environmental domain. Accounts will be provided to everyone that requests one and agrees to the terms of use[9] as well as the data policy[10] of DEIMS-SDR.

Central Data Node (CDN) - The CDN allows data providers, who don't have a local node, to register sensors and upload data to the eLTER Information System. Accounts will be provided to any national data providers who requests one and agree to the terms of use on the website.

eLTER Data Integration Portal (DIP) - The target audience of the eLTER DIP is mainly focused on the scientific community in order to provide access to observation data. The eLTER DIP is open for any data users without any restrictions. Registration and authentication is not needed and requested.

3 Data types

The basic entity for the data collection is the LTER Site and LTSER Platform[11]. The LTER research components are described using the DEIMS-SDR - Site and dataset registry[12]. The LTER-Europe network is comprised of 25 member networks with a pool of around 450 LTER

[9] See https://deims.org/terms

[10] See https://deims.org/privacy

[11] See http://www.lter-europe.net/lter-europe/infrastructure/sites-platforms/categories

[12] See http://deims.org/site/

Sites and 35 LTSER Platforms. LTER-Europe is a regional group within the global ILTER network. It provides multiple networking activities with permanent governance structures. Data are collected at these long term observation infrastructures.

The eLTER RI as permanent funded research infrastructure on European scale is built on the site network of LTER Europe and will further develop tools and services to be used by the whole LTER community.

In the eLTER context a distinction between two main types of data is made:

- **Internal data** - which encompass all data which are generated within the nearer project context (e.g. TA or scientific WP) or are subjected to VA. These data are seen as a project result and are therefore defined as "*open scientific data*" which need to be shared in a timely manner with the scientific community in accordance with EU Open Data Policy.
- **Third party and external data** - which encompass all data which are generated outside the nearer project context (e.g. historic data or third party data) either by previous projects not funded through the H2020 programme or similar funding schemes or are collected by any third party (e.g. national meteorological institutes. For these data, the access property rights are defined by the data owner and cannot be overruled by the projects data sharing policy.

Whereare for the first category of data full access to metadata and data can be granted, for the latter the eLTER Information System can only provide access to metadata. The primary data user has to acknowledge the specific data license restrictions defined by the third party.

Within the eLTER H2020 project the main focus was on the provision and sharing of time series data for

- **mesoclimatic observations** (meteorological data) - the data are provided as structured tabular data, documented by a linked MD record. The data are either collected within the current eLTER H2020 project or provided via the VA contract. The data are can be shared through OGC SOS services using the eLTER CDN or file based.
- **species diversity** of terrestrial ecosystems (vegetation data) - the data are provided as structured tabular data, documented by a linked MD record. The data are either collected within the current eLTER H2020 project or provided via the VA contract. The data are mainly shared file based using the eLTER data repository B2SHARE.
- **nitrogen deposition** - the data are provided as structured tabular data, documented by a linked MD record. The data are either collected within the current eLTER H2020 project or provided via the VA contract. The data are can be shared through OGC SOS services using the eLTER CDN or file based.

The detailed specification for the data requests is given in the Data Specification for Virtual Access (Peterseil et al. 2018)[13] and the accompanying Field Specification for Data Reporting (Peterseil & Wohner 2018)[14]. For completeness in the following only short information on the different data types is given.

The thematic scope of the eLTER H2020 data collection was based on the user requirements. The scope of data to be considered in the eLTER Information System will extend with the implementation of the eLTER RI.

4 Standards and metadata

In order to ensure data interoperability and re-usability for several applications/models, implementation of standards and regulations for particular data groups, which are not covered by common standardization rules, are needed. These harmonisation rules are specified in cooperation with the scientific experts in their fields. Integration between the data being gathered in the project and pre-existing data sources are made on several levels, as far as appropriate:

- ***Observable property level*** - where needed, transformation of units and recalculations of values, as well as mapping of local parameter naming to common naming of the parameters (using EnvThes, see section 4.1)
- ***Format level*** - application of standardized formats where applicable (see section 4.2)
- ***Documentation level*** - use of standardized metadata profiles (see section 4.3)
- ***Geographic reference system level*** - transformation to standardized coordinate systems (see section 4.4)
- ***Temporal and spatial scales*** - done based on the analysis needs.

4.1 Core Controlled Vocabulary

eLTER is using EnvThes[15] as core vocabulary. EnvThes is applied for keyword tagging as well as harmonisation of parameter names. All reference lists used in the context of DEIMS-SDR will be managed and described in EnvThes to ensure a) the ability to provide a unique identifier for the concept, b) the ability to provide an online reference, and c) the ability to translate the concept (multi-linguality).

EnvThes is developed as a community effort of the long term monitoring community and related projects (e.g. eLTER H2020 and EcoPotential). Synchronisation of the controlled

[13] see https://www.lter-europe.net/document-archive/elter-h2020-project-files/d-3-4-data-spec-va

[14] see https://www.lter-europe.net/lter-europe/data/eLTER_T3.4_VA_TemplateSpecification_V0.7_Short.pdf

[15] See http://vocabs.lter-europe.net/edg/tbl/EnvThes.editor

vocabulary with the MD edit framework DEIMS-SDR will be done regularly after major updates of EnvThes. This will be decided by the EnvThes working group of LTER Europe.

4.2 Data format

Data storage and data formats highly depend on the types of data, which can be very heterogeneous. Basically, data of interest in eLTER H2020 can be classified as follows:

Structured (quantitative) data:

- Time series data provided by online in-situ or remote sensors, to be stored in relational databases
- Tabular (statistical) data derived from field campaigns, surveys or generated by simulations, to be stored in files or relational databases

Unstructured (qualitative) data

- e.g. descriptive documents, audio- and video data, pictures, stored in files

Geo-referenced data stored in files or relational geo-databases

- Vector-based geospatial data
- Raster based geospatial data

Except for unstructured information, all data are being interpreted by computer programs to make them understandable and are - by their very nature - software dependent. Data are thus endangered by the obsolescence of the hardware and software environment on which access to data depends. The best option to warrant interoperability of data between varieties of applications and over the long term is to convert data in non-proprietary and standardized open-formats. This will be provided as a recommendation to the data users.

Nevertheless, archiving of data in proprietary or quasi-proprietary formats (e.g. MS Word, MS Excel) can be required in cases where meaningful content would get lost when storing in open formats, e.g. formulas and diagrams in MS Excel tables when exporting in comma-separated values (ASCII/CSV) files. In order to save storage space on the server's hard-drives, uploaded data may be stored as open compressed data file packages (e.g. .zip, .7z). Local data managers may be contacted if data from non-proprietary or not supported file formats need to be converted into supported formats.

An overview of available data formats is given in Table 1. The list encompasses an initial overview of available and considered data formats which is based on evaluations done by Kunkel et al (2013). Additional data formats might be added to the list, if needed.

Table 1. Table of relevant data formats to be considered in long term observations (adapted from [Kunkel 2013])

Type of data	Supported formats for exchange and reuse
Qualitative textual data (textual documents)	● Rich text format (rtf) ● Hypertext markup language (html/htm) ● Plain text (txt) ● Extensible markup language (xml) according to an appropriate Document Type Definition (DTD) or schema (xsd) ● **Portable Document Format (pdf)** ● Microsoft Office Word Document (doc/docx)
Qualitative tabular data (with or without column labelling, variable names and metadata)	● **Comma separated values (csv)** ● Tab delimited file (tab) ● Extensible markup language (xml) according to an appropriate Document Type Definition (DTD) or schema (xsd) ● NetCDF (nc/cdf) ● Microsoft Office Excel (xls/xlsx) ● Microsoft Office Access (mdb/accdb)
Digital image data	● Bitmap (bmp) ● **Joint Photographic Experts Group (jpg/jpeg)** ● Portable network Graphics (png) ● Tagged Image File (tif/tiff)
Geo-referenced vector data	● **ESRI Shape File (package with shp/shx/dbf/prj)** ● ESRI Personal Geodatabase (gdb) - not suggested ● Geographic Markup Language (gml) ● Keyhole Markup Language (kml/kmz)
Geo-referenced raster data	● GeoTIFF (tif/tiff) ● NetCDF (nc/cdf/netcdf) ● **ESRI ASCII GRID (asc)** ● ESRI Binary GRID (package) - not suggested
Digital Video Data	● MPEG-4 (mp4)
Digital Audio Data	● MPEG-3 (mp3)
Compressed file formats	● **Zip-Format (zip)** ● 7-zip-Format (7z) ● rar-Format (rar)

Sensor based time series data (timestamp - value pairs) will be mostly not stored in simple file based formats but in relational database systems. For data storage in the database and registration of the sensor metadata, an underlying data model for time series data is used, which stores observation data along with sufficient metadata to provide traceable heritage from raw measurements to usable information in the Central Data Node [cDN].

4.3 Data documentation

Metadata are defined a descriptive information on information objects, which support the exchange of information between data provider and data user in standardised way. If not otherwise provided the collection of metadata will be based using DEIMS-SDR[16].

This includes the collection of meta-information on the following elements a) Research site , b) Dataset, c) Person, d) Network, and e) Sensor. The metadata models are based on existing relevant community standards, which are described in the following.

4.3.1 Site documentation

The Site metadata model describes the metadata elements of a site. A 'Site' is defined as an in-situ observation or experimentation facility, delimited in space, but varying in size and complexity of the internal organizational and observational design, for the collection of data covering e.g. bio-geophysical, biotic or socio-ecological characteristics.

The metadata model is based on the requirements defined by target stakeholder groups (e.g. ILTER/LTER-Europe) and research projects (e.g. EnvEurope (EnvEurope, 2019), ExpeER (ExpeER, 2019), eLTER and EcoPotential) with the basic metadata elements being in line with the data specification of INSPIRE EF. It defines metadata elements about the organisation (e.g. contact, information and networks), the location, the observation characteristics (e.g. climate, habitats) or available equipment. Additionally, there are fields about the focus and design of a site, network affiliation and information about data policies and data management.

A unique identification of the Site is provided by using the **DEIMS.ID** *(e.g. https://deims.org/8eda49e9-1f4e-4f3e-b58e-e0bb25dc32a6)* and a network specific identification (e.g. *https://deims.org/site/LTER_EU_AT_003*) as alternative identifier. The template for the network specific identifier is defined by the LTER Europe network. While the DEIMS.ID is a persistent identifier, the identification using the network code is volatile as the site codes are managed by each national network and can change over time.

4.3.2 Dataset documentation

The dataset metadata model (DSMM)[17] is based on requirements defined by the target stakeholder groups. The user requirements for the content were harmonized with the INSPIRE metadata regulation and implemented using Ecological Metadata Language (EML, version 2.1.1). The MD model focused on the EML metadata standard (stemming from the US LTER activities) but also provides a mapping to ISO compliant metadata.

[16] See http://deims.org

[17] See https://deims.org/models

Dataset metadata can be exposed according to the following MD standards:

- Ecological Metadata Language (EML, version 2.1.1)[18]
- INSPIRE MD Regulation (based on ISO19115/19139)[19]

A unique identification of the dataset MD record is provided by using a **UUID** for the identification of the MD record, (e.g. *https://deims.org/dataset/75a7f938-7c77-11e3-8832-005056ab003f*). This UUID is also used for MD harvesting.

4.3.3 Sensor documentation

The sensor description is based on minimum requirements to generate valid SensorML. It is based on SensorML 2.0. The sensor documentation is implemented in the eLTER CDN and is linked to the uploaded data.

4.3.4 Other metadata

In addition to the metadata on sites and datasets also information on persons, networks and sensors stored. The MD records provide a reference to this information. The MD models are proprietary and are described within DEIMS-SDR.

4.4 Geographic reference

If not otherwise specified World Geodetic System 1984 (WGS 1984, EPSG:4326) is used.

Alternatively (especially for raster data) the ETRS-LAEA (ETRS89, EPSG:3035) for European scale data can be used as well:

> *WGS84 Bounds: -10.6700, 34.5000, 31.5500, 71.0500*
>
> *Projected Bounds: 2426378.0132, 1528101.2618, 6293974.6215, 5446513.5222*
>
> *Scope: Single CRS for all Europe. Used for statistical mapping at all scales and other purposes where true area representation is required.*
>
> *Last Revised: Feb. 2, 2007*
>
> *Area: Europe - ETRS89*

[18] See http://knb.ecoinformatics.org/software/eml/eml-2.1.1/

[19] See http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32008R1205:EN:NOT

4.5 Standard Data Model (SDM)

For standard data products (e.g. air temperature) standard data models (SDM) will be developed and implemented. This process is still in progress and not completed now. The SDM will apply for qualitative and quantitative tabular data (either file based or databases). The current version applied for the data flows is provided on the LTER Homepage[20].

For sensor based time series data the data model according to OGC Observation and Measurement (Version 1.0 and 2.0), as applied by the 52°North SOS Server will be implemented for the CDN.

5 Data curation processes

5.1 Data quality control

The level of data quality, following the common definition of data quality i.e. "*The state of completeness, validity, consistency, timeliness and accuracy that makes data appropriate for a specific use*"[21], is dependent on the real world phenomena represented by the data and data products. To carry out data quality assessments, expert knowledge on the field of data generation is demanded. Therefore, the responsibility for data quality assessments and assurance measures is by the nominated responsible persons working on the data. These persons are also responsible to describe the data quality assessment measures and methods and to provide an evaluation on the trustworthiness of the data within the Metadata Catalogues.

For eLTER H2020 the Data Quality Control process is delegated to the data provider who has to provide sufficient information on the quality of the data (at least on the level of the metadata). The dataset should not only contain the values, but also information on the data quality (e.g. [data quality flag]). As data are resulting from different sources a harmonisation of the data quality flags is not possible.

As requirement the data quality information needs to be provided in English. Common generic data quality flags still need to be developed.

[20] see https://www.lter-europe.net/lter-europe/data/eLTER_T3.4_VA_TemplateSpecification_V0.7_Short.pdf

[21] See http://en.wikipedia.org/wiki/Data_quality

5.2 Data processing levels

For eLTER H2020 three different levels of data processing are identified. This processing status is characterised by the different processing steps which are applied to the dataset. Nevertheless, if data are shared - regardless of the data processing level - metadata need to be provided.

Depending on the parameters being controlled, the following data processing levels are identified (taken from Kunkel et al. 2013):

- **Raw data** - unprocessed or preliminary pre-processed data and data products, that have not undergone quality control.
- **Quality controlled data** - data that have passed quality control procedures (e.g. visual or automatic inspection, calibration or correction). Quality information needs to be provided with the data (see chapter 5.1).
- **Derived data products** - data that have been derived from quality controlled data (e.g. by aggregation or modelling).

The data processing level needs to be specified within the method section of the dataset MD.

6 Data sharing

6.1 Data identification

Depending on the storage location of the dataset, different options for the identification of the 'dataset' can be implemented for the eLTER Information System. A dataset can be encompassed here also as data service (e.g. OGC WFS or OGC SOS).

6.1.1 File based data representation

Metadata records for research sites and datasets can be referenced using the UUID. The base URL for the identification for the sites is http://deims.org (without any additional suffix) and https://deims.org/dataset/ for the datasets.

If the data file is uploaded to B2SHARE, also a B2SHARE Handle (PID) is provided to the user uniquely identifying the data document. This PID is also referenced in the MD record for the dataset.

If the data file is uploaded to any cloud data repository a persistent identifier might exist (e.g. DOI). If existing, this identifier can be used in addition to the MD UUID.

For long term data archiving a unique identifier for the metadata record as well as for the dataset needs to be provided. The issue of dynamic data is currently not addressed in the eLTER Information System.

6.1.2 Service based data representation

For service based data representation (e.g. OGC WFS or OGC SOS) the datasets are uniquely identified by the MD UUID provided in the dataset MD record.

6.2 Data rights management

Data rights management includes the regulation on privacy and intellectual property. This is subject to legal regulation and will be attached to the data management plan in a separate document defining the eLTER Data Sharing Policy.

6.3 Data sharing and citation

eLTER is fostering FAIR and open data sharing but aims to implement three principles of FAIR data use. This is a) citation of data used, b) acknowledgement of data providers and collectors, and c) invitation to further collaboration.

- **Citation**. It is considered a matter of professional ethics to acknowledge the work of other scientists. Thus, the Data User are expected to properly cite the Dataset in any publications or in the metadata of any derived data products that were produced using the Dataset. Citation should take the following general form: Creator, Year of Data Publication, Title of Dataset, Publisher, Dataset identifier, URL, Date of dataset access or as defined in the metadata field "Recommended citation" in DEIMS-SDR[22] . For example:

  Scotti A, Tappeiner U, Bottarin R (2019) Stream benthic macroinvertebrates abundances over a 6-year monitoring period of an Italian glacier-fed stream. Biodiversity Data Journal 7: e33576. https://doi.org/10.3897/BDJ.7.e33576 (Online access 1.4.2019)[23]

  or if no recommended citation is provided using the dataset title and DEIMS-SDR Metadata UUID as reference, for example:

[22] See https://deims.org/models

[23] See https://deims.org/dataset/628e6286-e7e0-4b61-b3ea-d019f985f27d

LTER Zöbelboden, Austria, Air chemistry, 2012 [https://deims.org/dataset/cd1fb6f8-5e57-11e3-aa73-005056ab003f] (Online access 1.4.2019)

- **Acknowledgement**. The Data User should acknowledge any institutional support or specific funding awards referenced in the metadata accompanying this Dataset in any publications where the Dataset contributed significantly to its content. Acknowledgements should identify the supporting party, the party that received the support, and any identifying information such as grant numbers. For example:

  Datasets were provided by eLTER. Significant funding for these data were provided by the H2020 (Grant 654359).

- **Collaboration**. The Dataset has been released in the spirit of open scientific collaboration. Data Users are thus strongly encouraged to consider consultation, and collaboration with the Dataset Creator especially through further transnational access arrangements to research facilities.

7 Archiving and preservation (including storage and backup)

Archiving and data preservation is subjected to aggregated data products which shared using the eLTER Information System. Raw data are managed by the data provider and are subject to local archiving and preservation regulations.

Depending on the data policy and use regulations, data are either only short-term stored for data analysis reasons (see chapter 7.1) or are archived and shared using the eLTER Information System (see chapter 7.2).

7.1 Short term data archiving

If data are not publicly shared but used in the project context (e.g. census data which are subjected to IPR and data use regulations) these data might be stored short term on the eLTER ftp-site. The original reference and citation for this data must be provided in the derived data products, but the data files need to be deleted after the finalisation of the data analysis.

These data are secured by password and the ftp-repository is backed once a week manually by the data management team.

7.2 Long term data archiving

7.2.1 Site documentation

DEIMS-SDR is regularily archived and backuped by UKRI CEH.

7.2.2 File based data provision

Data provided as data file can be deposited using the B2SHARE repository (https://b2share.fz-juelich.de/ and https://b2share.eudat.eu/). This data are identified by a persistent identifier using B2HANDLE (e.g. http://hdl.handle.net/11304/1f29ee76-eac1-11e5-9bb4-2b0aad496318) and linking to the data file. Data shared via B2SHARE are centrally backup and archived by the EUDAT CDI and FZ Jülich. Metadata are linked to the data file within the repository.

7.2.3 Service based data provision

The CDN has been implemented and tested within the eLTER H2020 project. Based on 52°North SOS, it provides a simple interface for registering sensors and uploading data from these sensors into the CDN. It provides a simple forms-based interface for SensorML for users. Functionality for registering single sensors has been implemented and the ability to register and batch load data from multiple sensors will be tested within the eLTER H2020 project extension.

The database of the eLTER CDN is regularly archived and backup by UKRI CEH.

8 References

DCC (2013). Checklist for a Data Management Plan. v.4.0. Edinburgh: Digital Curation; online avaialbe at http://www.dcc.ac.uk/sites/default/files/documents/resource/DMP/DMP_Checklist_2013.pdf

EnvEurope, 2019. http://www.enveurope.eu/, Accessed date: 16 May 2019.

ExpeER, 2019. http://www.expeeronline.eu/, Accessed date: 16 May 2019

Kunkel Ralf et al. (2013) Establishing a West African Science Service Center on Climate Change and Adapted Land Use (WASCAL); online available at https://icg4wascal.icg.kfa-juelich.de/wascal_searchportal2/downloads/WASCAL-DMP.pdf

Kunkel Ralf et al. (2014) Terrestrial Environmental Observatories (TERENO) Data Management Plan; online available at http://teodoor.icg.kfa-juelich.de/downloads/DMP-V1.0.pdf

9 Annex 1 - Central roles and functions

9.1 eLTER Data Management Coordination Team

- Johannes Peterseil (Umweltbundesamt GmbH, Austria, johannes.peterseil@umweltbundesamt.at)
- John Watkins (UKRI-CEH, UK, jww@ceh.ac.uk)

9.2 eLTER Data Management Support and Maintenance

- DEIMS-SDR Christoph Wohner (Umweltbundesamt GmbH, Austria, christoph.wohner@umweltbundesamt.at)
- eLTER CDN Sue Rennie (UKRI-CEH, UK, srennie@ceh.ac.uk)
- eLTER DIP Vladan Minic (BioSense Institute, Serbia, vladan.minic.ns@gmail.com)
- EnvThes Barbara Magagna (Umweltbundesamt GmbH, Austria, barbara.magagna@umweltbundesamt.at)

10 Annex 2 – Examples for data provision

10.1 Example station description

SCODE	STYPE	westBoundingCoordinate	eastBoundingCoordinate	northBounding Coordinate	southBounding Coordinate	altitude Minimum	altitudeMaximum
IP1	PT	14,4466	14,4467	47,8385	47,8384	950	950

10.2 Example biophysical data

Recommended version basic format

SCODE	SUBST	LEVEL	TIME	VALUE	UNIT	FLAGQUA	FLAGSTA
IP1	TEMP	200	2016-03-15	5.5	°C		X
IP1	PREC	100	2016-03-03	10.2	MM		S
IP1	TEMP	200	2016-02-15	2.5	°C		X
IP1	NH4N	100	2016-03	5.5	mg N/l		W
IP1	SO4S	100	2016-03	10.2	mg S/l		W
IP1	CA	100	2016-03	2.5	Mg/l	L	W
…	…	…	…	…	…	…	…

Alternative version

SCODE	LEVEL	TIME	TEMP	PREC	NH4N	SO4S	CA	TYPE
IP1	100	2016-03	5.5	10.2	2.5	5.5	2.5	Forest
IP1	100	2016-04	5.2	1.2	2.2	5.8	1.2	Forest

Note: Resolution and methods needs to be ddescribed in detail for the single parameter. Additional information for each value (e.g. aggregation level or quality) cannot be provided in this format.

10.3 Example biodiversity data

SCODE	SUBST	TIME	TAXA	VALUE
1	COVE_T1	2016-06-25	FAG SYLV	3
1	COVE_T1	2016-06-25	PIC ABIE	3
1	COVE_S	2016-06-25	FAG SYLV	1
1	COVE_F	2016-06-25	OXA ACET	2
…	…	…	…	…

10.4 Example method documentation

CODE	SAMPLING	FIELD_METHOD	LAB_METHOD	AGG_METHOD
METH_1	providing specification for the sampling methods	providing specification for the field method	providing specification for the analysis method, including the statistical analysis of the dato if relevant	provide specification for the aggregation method
METH_2	example selected	water sample	ICP_OES	

10.5 Example reference list

FIELD_NAME	LIST_CODE	CODE	NAME	DEFINITION
SUBST	CF	NOXN	atmosphere_mass_content_of_nox_expressed_as_nitrogen	"Content" indicates a quantity per unit area. The "atmosphere content" of a quantity refers to the vertical integral from the surface to the top of the atmosphere. For the content between specified levels in the atmosphere, standard names including content_of_atmosphere_layer are used. "Nox" means a combination of two radical species containing nitrogen and oxygen: NO+NO2. The phrase 'expressed_as' is used in the construction A_expressed_as_B, where B is a chemical constituent of A. It means that the quantity indicated by the standard name is calculated solely with respect to the B contained in A, neglecting all other chemical constituents of A.
SUBST	CF	NOYN	atmosphere_mass_content_of_noy_expressed_as_nitrogen	"Content" indicates a quantity per unit area. The "atmosphere content" of a quantity refers to the vertical integral from the surface to the top of the atmosphere. For the content between specified levels in the atmosphere, standard names including content_of_atmosphere_layer are used. "Noy" describes a family of chemical species. The family usually includes atomic nitrogen (N), nitrogen monoxide (NO), nitrogen dioxide (NO2), dinitrogen pentoxide (N2O5), nitric acid (HNO3), peroxynitric acid (HNO4), bromine nitrate (BrONO2) , chlorine nitrate (ClONO2) and organic nitrates (most notably peroxyacetyl nitrate, sometimes referred to as PAN, (CH3COO2NO2)). The list of individual species that are included in a quantity having a group chemical standard name can vary between models. Where possible, the data variable should be accompanied by a complete description of the species represented, for example, by using a comment attribute. The phrase 'expressed_as' is used in the

				construction A_expressed_as_B, where B is a chemical constituent of A. It means that the quantity indicated by the standard name is calculated solely with respect to the B contained in A, neglecting all other chemical constituents of A.
…	…	…	…	…

European Long-Term Ecosystem and Socio-Ecological Research Infrastructure

Data Sharing Policy (Draft)

Authors: eLTER Data Management Group
Lead partner for deliverable: Umweltbundesamt GmbH
Other partners involved: UKRI-CEH, BSI, FZJ, CNR

H2020-funded project, GA: 654359, INFRAIA call 2014-2015

Start date of project: 01 June 2015 Duration: 48 months
Version of this document: Version 2.0 Submission date: 2019-06-30

Dissemination level

PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	
CI	Classified, as referred to in Commission Decision 2001/844/EC	

Version control	Edited by	Date of revision
Created – V1	Data Management Group	June 2016
Internal review	Data Management Group	May 2017
Internal review		
Revised – V2	Peterseil	15th May 2019
Revised – V3	Watkins	7th June 2019
Revised – V4	Watkins	21th June 2019
Signed off – coordinator	Peterseil	30th June 2019

Publishable Executive Summary

The eLTER Data Sharing Policy defines guidance for accession, distribution and re-use of data from the eLTER information infrastructure. This policy applies to data generated by the eLTER H2020 project but also aims to be a common recommendation for related national long-term monitoring networks as well as projects. It aims to foster data re-use through open access with due acknowledgement of data originators. The eLTER Data Sharing Policy is based on best practise examples by related national networks and environmental research infrastructures. The eLTER H2020 project is extending this best practice to support open science and open data policy. This is an important input to the development of the eLTER research infrastructures in the coming years.

The eLTER Data Sharing Policy is a draft document which needs to be agreed by the data provider. Reference to the eLTER DSP can be provided in the dataset metadata.

Contents

1 Preamble

The eLTER Data Sharing Policy (DSP) is defined for the eLTER H2020 project to provide clear guidance for data provision and sharing. It also aims to provide a recommendation for the development and implementation of data sharing policies within related national LTER networks or projects. This fosters interoperability of data licensing across projects, which is an important aspect in the implementation of the FAIR principles. The Data Sharing Policy follows common community practices (e.g. the ILTER Data Policy) as well as best practises from related environmental research infrastructures and projects aiming to reduce the barriers for data publication and sharing. eLTER as H2020 project is dedicating itself to open science data policy.

The eLTER Data Sharing Policy defines guiding principles to be applied when regulating the deposit, access and re-use of data through the services of the eLTER Information System. The eLTER DSP aims to ensure the reuse and reusability of long-term observation data following FAIR principles and to foster the visibility and acknowledgement of data providers in reward for the creation and dissemination of these data.

2 eLTER Principles of Data Sharing

eLTER collects and analyses long-term data sets to better understand patterns and processes in ecosystems and the interactions between ecosystems and socio-economic systems. eLTER supports and encourages the full and open sharing of ecological and socio-economic data free of charge to advance research and education in the eLTER and for wider exploitation. eLTER documents and archives data for the benefit of present and future researchers as global environmental science becomes increasingly data driven. eLTER policies for data sharing and accessibility consider principles and policies developed by other global networks, including the Global Earth Observation System of Systems (GEOSS) (GEOSS, 2005), the Organization for Economic Co-operation and Development (OECD, 2007), as well as recommendations by the European Commission on Open Data (EC, 2017)[1] and INSPIRE (2007/2/EC)[2] that support sharing and open access to publicly funded research data.

eLTER encourages the public sharing of data and metadata. Data collected as a result of public funding to eLTER partners should be made available online with as few restrictions as possible, on a non-discriminatory basis. eLTER scientists should make every effort to release data in a timely fashion allowing for publication and citation of eLTER derived research and supporting data with attention to accurate and complete metadata.

[1] see http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf
[2] see https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32007L0002&from=EN

3 Term definitions

Dataset is defined as quantitative or qualitative attributes of variables or sets of variables that have been gathered (also defined as "observation") following the standard data collection and quality assurance procedures defined by eLTER. Data collectors can be either single scientists or sensors depending on the operation procedure.

Metadata is defined as information about the data description, provenance, quality, processing and maturity level (raw data streams, automated quality control, processed, derivative products), as well as the necessary data collection context in order to support interoperability with other observatories, archives, and databases following common standards.

Data Portal is defined as the IT infrastructure providing access to the data using the eLTER Data Integration Portal (DIP) for data publication and sharing and DEIMS-SDR for metadata provision and exchange.

Data Provider is defined as any entity (person, institution or legal entity) that submits and shares Datasets through the eLTER DIP. Within DEIMS-SDR the Data Provider is specified in the metadata element "Contact".

Data User is defined as any entity (person, institution or legal entity), that requests access to any Dataset shared through the eLTER Data Portal or any linked thematic data portal harvesting the metadata (e.g. eLTER DIP).

4 Elements of the eLTER Data Policy

The eLTER Data Policy intends to provide open and free access to metadata and data using the eLTER Data Portals whenever possible. Metadata and data will be subject to the terms and conditions for use, citation and repurposing. The policy guidance for these is described below.

This policy is implemented through the licensing of individual data sets by the originators, where such licenses exist, or through the general principles of this policy, namely - That all data are provided under the terms and conditions compliant with a CC-BY license unless stated otherwise.

4.1 Data sharing principles

4.1.1 Metadata

Metadata is always freely accessible, without any restriction through the eLTER Data Portal or any related catalogue.

4.1.2 Datasets

Datasets are openly available under the terms and conditions compliant with a CC-BY license unless stated otherwise

4.1.3 Data embargo periods

Data embargo periods will be supported by this policy upon request of Data Providers to protect the right of originators to first publication. This period will be 1 year from data submission to eLTER. Only under particular conditions, Data Providers may request a longer embargo period. During the embargo period, only metadata will be fully available.

4.1.4 Sensitive data

Sensitive data (e.g. occurrence of endangered species or private individual's information) may be subject to additional restrictions that will be negotiated between individual Data Providers and Data User. Indication on sensitive data needs to be specified in the Metadata.

Any principles for unrestricted data access and use in this policy cannot countermand statutory rights to copyright or data privacy (GDPR rights) within the jurisdiction of the data provider's principal place of business.

4.2 Attribution and citation

4.2.1 Citation

It is considered a matter of professional ethics to acknowledge the work of other scientists. Thus, the Data User are expected to properly cite the Dataset in any publications or in the metadata of any derived data products that were produced using the Dataset. Citation should take the following general form: Creator, Year of Data Publication, Title of Dataset, Publisher, Dataset identifier, URL, Date of dataset access or as defined in the metadata field "Recommended citation" in DEIMS-SDR3 . For example:

> *Scotti A, Tappeiner U, Bottarin R (2019) Stream benthic macroinvertebrates abundances over a 6-year monitoring period of an Italian glacier-fed stream. Biodiversity Data Journal 7: e33576. https://doi.org/10.3897/BDJ.7.e33576 (Online access 1.4.2019)*[4]

or if no recommended citation is provided using the dataset title and DEIMS-SDR Metadata UUID as reference, for example:

> *LTER Zöbelboden, Austria, Air chemistry, 2012 [https://deims.org/dataset/cd1fb6f8-5e57-11e3-aa73-005056ab003f] (Online access 1.4.2019)*

[3] See https://deims.org/models
[4] See https://deims.org/dataset/628e6286-e7e0-4b61-b3ea-d019f985f27d

4.2.2 Acknowledgement

The Data User should acknowledge any institutional support or specific funding awards referenced in the metadata accompanying this Dataset in any publications where the Dataset contributed significantly to its content. Acknowledgements should identify the supporting party, the party that received the support, and any identifying information such as grant numbers. For example:

> *Datasets were provided by eLTER. Significant funding for these data were provided by the H2020 (Grant 654359).*

4.2.3 Collaboration

The Dataset has been released in the spirit of open scientific collaboration. Data Users are thus strongly encouraged to consider consultation, and collaboration with the Dataset Creator especially through further transnational access arrangements to research facilities.

4.3 IPR policy

4.3.1 Intellectual property

The intellectual property of the Datasets submitted by an individual Data Provider to the eLTER Data Portal remains exclusively with the individual data providers. Individual Data Providers are the only responsible for any copyright infringement, as well as for the breach of other proprietary right of any person or entity.

4.4 Warranty and liability

Where data are made available through eLTER, except if agreed otherwise, the Data User is expected to possess the appropriate professional skills and knowledge to interpret and make valid use of the scientific information provided. Use of these data indicates their acceptance of the following principles:

- eLTER data is provided “as is” without warranty of any kind, either expressed or implied, including, but not limited to, any implied warranty against infringement of third parties’ property rights, or merchantability, integration, satisfactory quality and fitness for a particular purpose.
- eLTER does not represent or warrant that the data will be error free or uninterrupted, or that all non-conformities can or will be corrected, or that any data is accurate or complete, or that they are of a satisfactory technical or scientific quality.
- eLTER shall not be held liable for any direct or indirect, incidental, consequential or other damages, including but not limited to the loss of data, loss of profits, or any other financial loss arising from the use of eLTER data, or inability to use them, even is eLTER is notified of the possibility of such damages.

- eLTER has no obligation to provide technical support or remedies for the data other than agreed by the eLTER Data Providers and described in the eLTER Data Sharing Policy (DSP) and Data Management Plan (DMP).

5 Acknowledgements

This policy was adapted from the LTER Network Data Access Policy, Data Access Requirements, and General Data Use Agreement (http://www.lternet.edu/data/netpolicy.html) of the US LTER and the ILTER Network Data Access Policy, Data Access Requirements, and General Data Use Agreement5. We also considered best practise examples from related environmental research infrastructures, namely ICOS, ACTRIS or INTERACT, as well as national LTER networks, namely LTER Italy and LTER CWN.

[5] https://www.ilternet.edu/sites/default/files/ILTER%20Network%20Data%20Access%20Policy-8.23.2008-1.pdf