European long-term ecosystem, critical zone and socio-ecological systems research infrastructure PLUS

Concept and extension of registering research objects

Deliverable D11.4

13 January 2023

Editorial team:

Christoph Wohner - EAA
Hanna Koivula - CSC

Contributing authors:

Alessandro Oggioni - CNR
Paolo Tagliolato - CNR
Cristiano Fugazza - CNR
Vladan Minic - BioSense
John Watkins - CEH
Johannes Peterseil - EAA

Prepared under contract from the European Commission
Grant agreement No. 871128
EU Horizon 2020 Research and Innovation action

Project acronym:	**eLTER PLUS**
Project full title:	European long-term ecosystem, critical zone and socio-ecological systems research infrastructure PLUS
Start of the project:	Feb 2020
Duration:	60 months
Website:	https://elter-ri.eu/elter-plus
Deliverable title:	Concept and extension of registering research objects
Deliverable n°:	D11.4
Nature of the deliverable:	Demonstrator
Dissemination level:	Public
Citation:	Wohner, C., Koivula, H. (2023). *Concept and extension of registering research objects*. Deliverable D11.4 EU Horizon 2020 eLTER PLUS Project, Grant agreement No. 871128.

Deliverable status:

Version	Status	Date	Author(s)
1.0	Final	13 Jan 2023	Christoph Wohner (EAA)
0.95	Review	3 Jan 2022	Anssi Kainulainen (CSC), Holger Villwock (SLU)
0.9	Draft	9 Dec 2022	Christoph Wohner (EAA)

Table of Contents

Summary

This deliverable describes current concepts and frameworks for the description of research objects and their relevance to eLTER. In this regard, a research object is defined as a "*method for the identification, aggregation and exchange of scholarly information on the internet*". The document also lists all types of research objects that can already be described using the different components of the eLTER Information System at the time of writing and which types of ROs should be documented in the future. It further discusses which frameworks, models and formats would aid the documentation process as a whole or particular aspects of it as well as support the usage of the research objects in the community and how this would benefit all of eLTER.
The deliverable also delineates work done in the scope of task 11.2 consisting of the subtasks 11.2.1, 11.2.2 and 11.2.3 as well as task 11.3 consisting of 11.3.1 and 11.3.2.
While datasets are an important type of RO, they are mostly excluded from this document because they will primarily be addressed by deliverable D11.5 "Framework for data documentation and provenance using semantics".

Glossary

API	Application Programming Interface
CESSDA	Consortium of European Social Science Data Archives
CORDIS	Community Research and Development Information Service
DAR	Digital Asset Registry
DEIMS-SDR	Dynamic Ecological Information Management System - Site and Dataset Registry
DIP	Data Integration Portal
eLTER IS	eLTER Information System
EOSC	European Open Science Cloud
FAIR	Findable Accessible Interoperable Reusable
FDOF	FAIR Data Object Framework
GBIF	Global Biodiversity Information Facility
GIS	Geographic Information System
INSPIRE	Infrastructure for Spatial Information in the European Community
INSPIRE EF	Environmental Monitoring Facilities
JSON	JavaScript Object Notation
JSON-LD	JavaScript Object Notation - Linked Data
OGC	Open Geospatial Consortium
OWL	Web Ontology Language
REST	Representational State Transfer

RDA	Research Data Alliance
RDF	Resource Description Framework
RDFa	Resource Description Framework in Attributes
RDFS	Resource Description Framework Schema
RI	Research Infrastructure
RO	Research Object
SOS	Sensor Observation Service
SOSA	Sensor, Observation, Sample, and Actuator ontology
SSN	Semantic Sensor Network
UUID	Universally Unique Identifier
W3C	World Wide Web Consortium
WFS	Web Feature Service
WMS	Web Mapping Service

1 Introduction

1.1 Purpose of the document

The purpose of this document is to list existing types of research objects (ROs) in the current eLTER Information System (eLTER IS) as well as to define desirable ROs to be established in the eLTER IS in the future. Furthermore, it discusses existing concepts, frameworks and formats for the description of ROs. It also describes current components in the eLTER IS used for the generation of ROs, how these systems can be queried, potential external sources of ROs and how that information can be connected to create an added-value, e.g. by the creation of site information clusters or provenance chains. The conclusions drawn from this work will be the basis for future development roadmaps and implementation steps.

1.2 Terminology, motivation and status quo

Before discussing the various aspects of registering research objects, clear terminology is needed. According to Belhajjame et al. (2015) and Wikipedia (2021) a research object (RO) is defined as “a method for the identification, aggregation and exchange of scholarly information on the Web.” The goal of following the RO approach is “to provide a mechanism to associate related resources about a scientific investigation so that they can be shared using a single identifier”. ROs are not one specific technology, but are instead guided by a set of principles of identity, aggregation and annotation (Bechhofer et al. 2010).
Another definition provided by the European Open Science Cloud (EOSC) Interoperability Framework (IF) talks about "data at large" or “digital objects” that are subject to FAIR principles as: "[...] kind of objects that allow binding all critical information about any entity”. Furthermore, it is stated that - in the scope of the IF - this encompasses: “research data, software, scientific workflows, hardware designs, protocols, provenance logs, publications, presentations, etc., as well as all their metadata." (European Commission, Directorate-General for Research and Innovation, Corcho, O., Eriksson, M., Kurowski, K., et al., EOSC interoperability framework 2021).

Now that the terminology has been defined, it has to be explained why the documentation of ROs is necessary and worth the effort. Why do we want to implement the registration of research objects or extend dataset documentation to track data provenance?

One of the cornerstones of scientific endeavour is reproducibility of results. As Hanson et al. (2011) state: “We must all accept that science is data and that data are science, and thus provide for, and justify the need for the support of, much-improved data curation”. This statement also applies to biodiversity data which should be published, cited, and peer reviewed (Costello et al. 2013). Therefore, anything that aids or eases the process of reproducing scientific results must be in the interest of everyone involved in the scientific process. The RO approach is primarily motivated by a desire to improve reproducibility of scientific investigations (Bechhofer et al. 2010) and hence is fully in line with this process.
In order to increase the reproducibility of scientific work, data and all processing steps need to be clearly documented and available as possible to everyone. ROs should be treated as mutable objects with an object history (provenance), meaning that they can be updated as new knowledge becomes available while also keeping former RO versions accessible. Capturing and versioning the provenance, i.e. RO lineage consisting of each part of the scientific process leading to a result or statement, require proper references and are therefore citable. By providing a citation for all existing ROs, reproducibility of ROs increases and provides participants in the scientific process the opportunity to be credited.
eLTER as an emerging research infrastructure focusing on environmental data thrives to incorporate open science and data concepts broadly and therefore envisions the RO concept for implementation in the data generation workflow. Section 2.1 provides insight into already existing components that are suitable for implementing the RO concept or partly already implemented it.

Currently in eLTER, the following different types of ROs can be distinguished with various degrees of implementation and usage in the different systems. These RO types are listed in Table 1. The table also provides a partial mapping to the DataCite resource types (DataCite Metadata Working Group 2019) as there is not always a fully matching resource type. Other mappable ontologies could not be identified.

Table 1 List of Research Objects types as of December 2022

Category	Research Object Type	Documentation possible in current eLTER information system	Corresponding DataCite resource type
physical infrastructure	site	yes	~ Service
physical infrastructure	sensor	yes	~ Service
collected evidence	observation/measurement	no	~ Event
collected evidence	sample	no	PhysicalObject
collected evidence	specimen	no	PhysicalObject
collected evidence	digital evidence*	no	Audiovisual; Image; Sound

human resource	person	yes	~ Service
human resource	network	yes	~ Service
digital resource	method	no	~ Text
digital resource	dataset	yes	Dataset
digital resource	workflow	yes	Workflow
digital resource	model	yes	Model
digital resource	code	partly through the use of external services	Software

*E.g. voice recording of a bird or photo trap of a deer

Some of these RO types can already be documented in the current eLTER IS. Sites, persons and sensors (in a beta version) can be described using DEIMS-SDR. Datasets, web services (e.g. OGC WMS/WFS), models and code/computational notebooks can be described using the Digital Asset Registry (DAR). These components and how they expose RO information are described in detail (see section 2.1).
Code can be described, published and shared using public repositories such as GitHub, GitLab or Zenodo. eLTER aims to make use of well-established and open resources and is therefore not planning to implement an in-house software tool for such applications.
There is currently no dedicated system for the documentation of methods, samples, specimens and digital evidence in eLTER.
So while certain types of ROs can already be documented, they are currently oftentimes documented separately without the means to connect that information. Therefore, the different ways ROs can be documented and connected have to be evaluated.

1.3 Evaluation of frameworks and formats for the registration of research objects

In order to give statements on which concepts and frameworks for the registration and management of ROs are suitable to be implemented, a list of existing standards, frameworks and formats was compiled and evaluated. There are also a number of emerging standards and recommendations, which may not be ready to be implemented yet, but should be followed by the eLTER RI closely in order to connect to other existing or emerging RIs when new implementations are established.

Core questions, which are aligned with the general needs of eLTER, to be addressed are:

- Can data lineage/provenance be documented? If so, how?
- Which types of persistent identifiers (PID) can be used for different ROs?
- Which international standards can be identified and applied?

Criteria to be evaluated are:

- Expected amount of effort necessary to implement given standard or framework
- Compatibility with current components of the eLTER IS
- Suitability for eLTER use cases expected in the future

Furthermore, components and frameworks that can be used for addressing particular aspects of RO documentation such as persistent identification are also evaluated. Additional thought is also given to the maturity of the stage of lifecycle for each framework or component.

1.3.1 Research Object Crate (RO-Crate)

RO-Crate is a community effort to establish a lightweight approach to packaging research data with their metadata. It is based on schema.org annotations in JSON-LD, and aims to make best-practice in formal metadata description accessible and practical for use in a wider variety of situations, from an individual researcher working with a folder of data, to large data-intensive computational research environments (Soiland-Reyes et al. 2022). A dedicated goal of RO-Crate is to make data findable, accessible, interoperable and reusable (FAIR; Wilkinson et al. 2016).

RO-Crate allows to document the provenance (Sefton et al. 2022). To specify which equipment was used to create or update a Data Entity, the RO-Crate JSON-LD should have a so-called Context Entity for each item of equipment. Software used to create files can also be described using the entity of type SoftwareApplication.
The https://workflowhub.eu/, which is described in greater detail further down below, accepts upload by RO-Crate and generates RO-Crate to improve reproducibility of computational workflows that follow the Workflow RO-Crate profile. Furthermore, a dedicated Python for creating and consuming RO-Crates is available (rocrate Python Package 2022).

If all necessary ROs are already available in a machine-readable way, an RO-Crate can be implemented relatively easily. An example for how an RO crate can look like when applied to an existing eLTER dataset (Kobler et al. 2021) is provided in Figure 1.


```
{ "@context": "https://w3id.org/ro/crate/1.1/context",
  "@graph": [
    {
        "@type": "CreativeWork",
        "@id": "ro-crate-metadata.json",
        "conformsTo": {"@id": "https://w3id.org/ro/crate/1.1"},
        "about": {"@id": "./"}
  },
  {
    "@id": "http://hdl.handle.net/11304/c74f367d-ce04-4f26-afb5-1c2bcc74eb45",
    "@type": "File",
    "name": "AT_ZOEBELBODEN_PLANT_TREE_2003_2019_V20210420.7z",
    "description": "Long-term forest inventory and aboveground tree biomass data from LTER Zöbelboden Austria (2003-
2019)",
    "encodingFormat": "7z"
  }
]
}
```


Figure 1 Example for an RO-Crate describing an eLTER dataset on EUDAT B2SHARE

1.3.2 FAIR Data Objects (FDO)

The FAIR principles (Findable, Accessible, Interoperable, Reusable; Wilkinson et al., 2016) require metadata to be “rich” and to adhere to “domain-relevant” community standards. The richness of metadata and data can be increased by annotating (and referencing) the digital content with vocabularies that are also available as FAIR Digital Objects (FDO) themselves.
The FAIR Digital Object Framework (FDOF) is a framework, which defines a model to represent objects in a digital environment (FDOF 2022). The framework consists of a predictable identifier resolution behaviour, a mechanism to retrieve the object's metadata and an object typing system. The main goal of the FDOF is to define a set of features to provide foundational support for the FAIR principles.
To be classified as a FAIR Digital Object (FDO), a digital object is identified by a globally unique, persistent and resolvable identifier with a predictable resolution behaviour, described

by related metadata records (FDOs themselves) and classified by the FDOF typing system. As of January 2023, the Object Typing System is to be completed. Further changes or extensions to the framework can therefore be expected. The simplified model of FDO is presented in Figure 2.

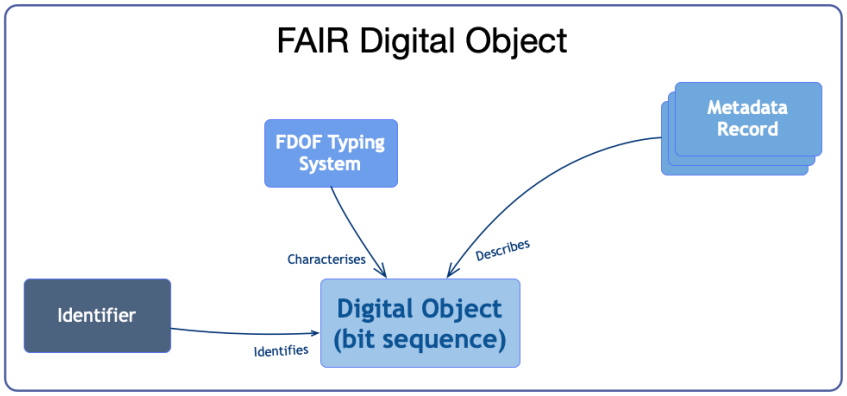

Figure 2 Simplified model of FAIR Digital Objects (FDOF, 2022)

According to the official documentation, no encoding of provenance information can be done at the time of writing this deliverable. If all necessary ROs are already available in a machine-readable way, an implementation of FDO is expected to be done easily.
For the eLTER IS and their ROs to be compliant with the FDOF it is therefore necessary for all relevant entity types (e.g. sites, sensors, samples, etc.) to have such globally unique, persistent and resolvable identifiers.

1.3.3 Resource Description Framework (RDF)

RDF is a framework for representing information on the Web allowing information to be easily identified, disambiguated and interconnected and thus making it machine actionable. The core structure of the abstract syntax is a set of triples, each consisting of a subject, a predicate and an object. A set of such triples is called an RDF graph (RDF Working Group 2014b). An RDF graph statement is represented by: 1) a node for the subject, 2) an arc that goes from a subject to an object for the predicate, and 3) a node for the object. Each of the three parts of the statement can be identified by a Uniform Resource Identifier (URI).
SPARQL is a standard query language for RDF graphs. Also other technologies, like OWL or SKOS (W3C) build on RDF and provide language for defining structured, web-based ontologies, which can complement the discipline specific community standards and provide cross-domain integration and interoperability of data.

An RDF representation of ROs is also suggested in the aforementioned FDO framework, which mandates RDF as the default metadata representation of digital objects. FDOF's Identifier Record must be represented in RDF (FDOF 2022).

However, different kinds of digital objects usually need discipline specific schema specifications, e.g. so called community standards and vocabularies for expressing and linking the RO's with others. Examples for such community-specific standards are:

- OGC Sensor Web Enablement (SWE) standards maintained by an OGC working group (https://www.ogc.org/node/698)
- The Interest Group on Physical Samples and Collections in the Research Data Alliance is currently working on recommendations which will be taken into consideration once published (Research Data Alliance FAIR Data Maturity Model Working Group 2020)

- Ecological Metadata Language (EML) using an XML-schema for ecological (meta)data (Matthew B. Jones et al. 2019)
- Darwin Core (DwC) and its extensions (are available as RDF-schemas; Darwin Core Task Group 2009)

Section 3.2 provides an example implementation of eLTER DEIMS-SDR (see section 2.1.2) entities, such as networks, site, activities, datasets and sensors, mapped and represented in the RDF format. This acts as a proof of concept for a potential RDF implementation for other components of the eLTER IS and would enable functionality as outlined in section 3 Potential applications.

1.4 Evaluation of related frameworks and systems

1.4.1 Provenance

Provenance is information about entities, activities and people involved in producing a piece of data or object, which can be used to form assessments about its quality, reliability or trustworthiness. PROV-DM is the conceptual data model that forms a basis for the W3C provenance (PROV) family of specifications. PROV-O is an OWL2 ontology allowing the mapping of the PROV data model to RDF (W3C 2013).

To be useful, an RO needs to be equipped with the artefacts (i.e. components expressing status or different phases/versions of the objects) used in a given execution including the workflow inputs, outputs, and, most importantly, the retrospective provenance. The process also needs to be described with vocabularies and/or (community) standards to allow creating documentation and versioning about processes.

Common Workflow Language (CWL) is an open standard for describing how to run command line tools and connect them to create workflows. Different communities, including OGC and Elixir, have piloted or adopted the use of CWL. Tools and workflows described using CWL are portable across a variety of platforms that support the CWL standard. Adopting CWL would hence enable eLTER RI to utilise workflow tools from related RIs. An example for shared workflows can be found at LifeWatch metadata catalogue, where five workflow descriptions from Oceanographic research are available at https://metadatacatalogue.lifewatch.eu/. The available workflows can be downloaded in the RDF format from the repository.

The DataCite DOI metadata schema allows tracking DOI metadata provenance (provenanceMetadata) by describing the history of a particular DOI metadata record (DataCite 2023b), i.e. what changes were made when and by whom. This information is stored and provided via an API for all DOI registrations since March 10, 2019 (DataCite Metadata Working Group 2019). This option is recommendable for ROs that need to be published and persistently citable from outside of the RI.

The schema.org schemas/standards and vocabularies are used along with the Microdata, RDFa, or JSON-LD formats to add information to your web content. It is also possible to use schema.org vocabularies for tracking provenance of data. However, this option better suits multimedia objects and other research communication outputs as well as training materials available on the Internet.

1.4.2 Persistent identification

Persistent identifiers (PID) are needed for referencing ROs in order to make them reusable and enable repurposing and giving credit to published ROs. A number of recommendations for the identification and referencing of different kinds of ROs have been published, but as

most of them are relatively new, only some of them have been piloted or implemented at this point.

According to the EOSC PID Policy (European Commission, Directorate-General for Research and Innovation, Hellström, M., Heughebaert, A., Kotarski, R., et al. 2022), a PID is an identifier that is globally unique, persistent, and resolvable and supports and enables research that is FAIR. PIDs can be born digital (e.g. documents, data, software, services - otherwise known as digital objects - and collections made of them), physical (e.g. people, instruments, artefacts, samples), or conceptual (e.g. organisations, projects, vocabularies).

The RDA Data Citation Working Group's recommendation for Data citation of evolving data (Rauber et al. 2021) and Parland-von Essen et al. (2018) give recommendations for identifying research data types and ways to assign different kinds of PIDs for different kinds of data to enable data citation. Research data can be categorised into three data types: operational data, generic research data and research data publications. Operational data are active, can be raw, and continuously updated. Due to its dynamic nature, it needs to be made stable before it can be cited. Citable data types are generic research data that are validated and can be cumulative, i.e. data can be added and cited with timestamp without versioning. However, if it is dynamic it should be versioned and each version is citable. The categorisation and usage of identifiers enable researchers and data centres to identify and cite data used in experiments and studies. The definitions of data levels and data product described in the Deliverable D11.2 EU Horizon 2020 eLTER PLUS Project (Koivula, H., et al. 2022) are aligned with these guidelines.

Resolvability is an essential part of data citation and tracking data provenance. Resolvable identifiers should be used when data are cited from outside the RI. A digital object identifier (DOI) is a persistent identifier or handle used to uniquely identify various objects, standardised by the International Organization for Standardization (ISO). DOI and other handle systems are resolvable with a separate resolving mechanism.
A Uniform Resource Name (URN) is a Uniform Resource Identifier (URI) that uses the URN scheme. URNs are globally unique persistent identifiers assigned within defined namespaces so they will be available for a long period of time, even after the resource which they identify ceases to exist or becomes unavailable. URNs cannot be used to directly locate an item and need not be resolvable, as they are simply templates that another parser may use to find an item. However, the resolver address can be added in front of the URN, (i.e. http://urn.fi) to make it resolvable, for example: http://urn.fi/urn:nbn:fi:fsd:T-FSD3424 .

A Cool URI is an identifier in the form of a web link that is always (re)directed to the same resource, even when the location has changed. For example, Linked Data is structured data, which uses Cool URIs to interlink with other data so it becomes more useful through semantic queries. It builds upon standard web technologies such as HTTP, RDF and URIs.

Systems that do not use resolver applications offer other types of identifiers. They can be based on international standards, de facto standards or might be local solutions.

Examples of identifiers that can be used to identify ROs:

- DataCite DOIs for any FAIR Data Object. Enables tracking provenance and crediting data originator.
- PIDINST schema for the metadata describing the instrument instance to be registered in the PID infrastructure (Stocker et al. 2020). Version 1.0 of the schema has been endorsed as a RDA recommendation (Krahl et al. 2021). Furthermore, the repository provides a mapping of the PIDINST schema onto the current DataCite Metadata Schema 4.4.

Existing PIDs and metadata schemas for different kinds of reference data are:

  - OrcID for researcher identification
  - CrossRef Funder registry (CrossRef 2023)
  - ROR for research organisations
  - EU Participant Identifier Code (PIC)
  - International Generic Sample Number (ISGN) ID for samples (i.e. physical objects; DataCite 2023a)
  - QID - WIKIDATA (can be used for research organisations)
  - ISNI is an identifier system for uniquely identifying the public identities of contributors to media content
  - ISBN identifiers for books
  - As well as recommendations developed by the aforementioned RDA working group for identifying evolving data (Rauber et al. 2021)

Examples of internal non-resolvable PIDs:

- Accession numbers in data archives, i.e.:
  - NP_999856.1 (protein)
  - rs141296055 (mutation)

Each of those types of identifiers might be used for assigning a suitable identifier for an RO within eLTER.

2 Defining interfaces and endpoints for data exchange

2.1 Description of available eLTER components

This section provides an overview of the current components in the eLTER IS that either document or process ROs or that might be used for documenting or managing ROs in the future.

2.1.1 Digital Asset Registry (DAR)

The Digital Asset Registry (DAR) is a web-based catalogue of digital assets generated by the eLTER research network. The types of assets that can be catalogued are datasets, web services (e.g. OGC WMS/WFS), models and code/computational notebooks (e.g. Jupyter). Additional assets can be added if required. Once assets are recorded, a faceted search interface allows users to find and access the data they need. These assets can be connected to eLTER sites that are described in DEIMS-SDR. By cataloguing and providing search facilities, the DAR helps to enable the 'Findable' and 'Accessible' elements of the FAIR principles.

All of DAR's core functionality that is available in the web interface can also be used programmatically through the DAR's RESTful API. This means that metadata can be searched, edited and manipulated by external code if required. The available information is exposed in multiple formats including JSON, XML and RDF. The following query exemplifies this functionality.

- https://catalogue.lter-europe.net/elter/documents?facet=elterDeimsUri%7Chttps://deims.org/8eda49e9-1f4e-4f3e-b58e-e0bb25dc32a6&format=json)

It returns all datasets associated with the eLTER site “Zöbelboden”. The query uses the persistent, unique and resolvable identifier of the site issued by DEIMS-SDR (https://deims.org/8eda49e9-1f4e-4f3e-b58e-e0bb25dc32a6).

2.1.2 DEIMS-SDR

DEIMS-SDR (www.deims.org) is a web-based information portal for integrated ecological information that comprises detailed descriptions of sites where research is carried out, including the technical infrastructure, ecosystem properties and research activities and foci. The available information is exposed through various formats and services. It acts as the primary site registry for the various LTER networks, most notably ILTER and eLTER (Wohner et al. 2022). It also allows the description of data collection activities and sensors. The sensor functionality is in a beta version that is currently not aligned on a greater eLTER level. While DEIMS-SDR also allows documenting datasets, the DAR is developing the functionality to replace the dataset function. This is due to the DAR's capability of allowing documentation of a range of digital assets (e.g. data, web services, code bases) together with their provenance. As of December 2022, DEIMS-SDR stores around 1200 site records, ca. 620 of them are eLTER sites. It offers different APIs to expose this site information, such as WFS, WMS, CSW or a REST-API. Examples for querying the DEIMS-SDR REST-API are given below. These examples illustrate how site records can be queried by:

- name
  - https://deims.org/api/sites?name=z%C3%B6belboden
  - https://deims.org/api/sites?name=lago
- country (using the two-digit ISO code), e.g.
  - https://deims.org/api/sites?country=at
  - https://deims.org/api/sites?country=it
- network using the UUID assigned to the network as well as if the sites should be verified members, e.g.
  - https://deims.org/api/sites?network=735946e0-4e9e-484a-acee-85e31f4e2a2e&verified=true
- observed property using the URI issued by the EnvThes (EnvThes 2023)
  - https://deims.org/api/sites?observedproperty=http://vocabs.lter-europe.net/EnvThes/22160

Through the use of these APIs, DEIMS information can be shared with others with relative ease. Site information, for instance, can be queried for all eLTER sites, integrated with different information sources and imported in different systems. These existing functionalities can also be applied with any subset of sites described on DEIMS, enabling custom functionality tailored to the subset of sites. This is also in line with the site information clusters concept discussed in (Peterseil et al. 2022). EcoSense, which is described in the following section, illustrates such a workflow.

2.1.3 EcoSense

EcoSense (EcoSense 2023) is a visualisation tool, initially developed in the H2020 e-shape (e-shape 2019) project. It is a web GIS portal, based on the AgroSens (Agrosens 2022) platform, designed for the visualisation of eLTER site metadata (fetched from DEIMS-SDR), time-series data (OGC 2012) and selected remote sensing (RS) data products. While EcoSense itself does not register ROs, it visualises ROs described coming from the aforementioned systems DAR and DEIMS-SDR. EcoSense will become a part of the (Extended) Data Integration Portal (DIP 2023). The majority of available data products originate from WP4, hosted on Central Data Node (CDN 2022). Based on eLTER PLUS project needs, it was decided to replace some components in order to support more data types that could be visualised. The user interface will be revised as well in order to support a larger amount of layers that can be searched and displayed. This process is currently on-going. Apart from data generated in eLTER, other (global) layers were added such as: Digital Elevation

Model (Copernicus DEM 2016), selected SoilGrids layers (SoilGrids 2023) or CORINE Land Cover (CORINE Land Cover 2018).

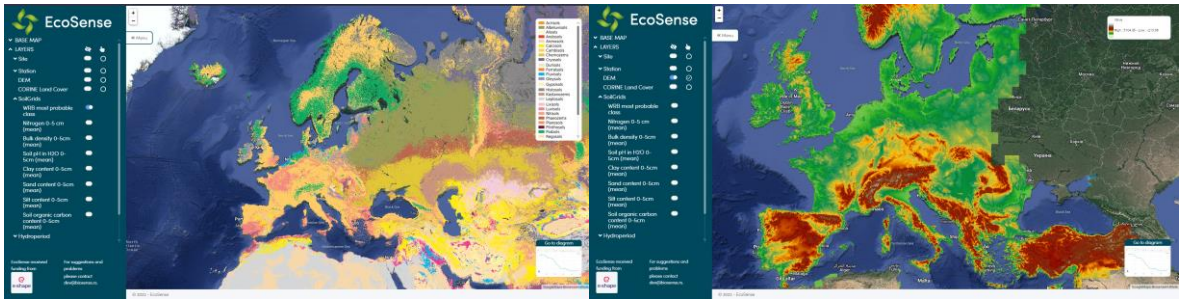

Figure 3 One of SoilGrids layers (left) and DEM layer (right) visualised in EcoSense

The aim of this platform is to enable visualisation of large amounts of data without a need for downloading them. Users may compare data retrieved from different sources and even be able to add their own data source (GeoServer/ArcGIS server, or SOS server). When satisfied with the discovered data regarding the type, resolution (time and spatial), outliers etc., they may use the original data source in DataLabs or their own Python or R environment for further processing.

2.2 Potential external sources of research objects

In addition to the ROs generated in eLTER, various external sources of ROs might be relevant and suitable for eLTER purposes. These sources range from a variety of pre-processed datasets, live data services as well as general online workflow services tools. In particular, they encompass:

- External pre-processed datasets, e.g.
  - E-OBS (Copernicus Climate Change Service 2020)
  - Eurostat population data
  - ICOS data products (ICOS 2023)
- External data services
  - GBIF API
  - Consortium of European Social Science Data Archives (CESSDA) Data Catalogue (CESSDA Data Catalogue 2023)
  - The Community Research and Development Information Service (CORDIS) providing information on all European Union (EU) funded research projects as well as persistent identifiers for these projects
- External tools and services
  - Workflowhub.eu
    - A registry for describing, sharing and publishing scientific computational workflows

In general, any service providing suitable APIs or datasets that can be used in a data workflow might be a source of ROs. The prerequisites are similar to the ones for ROs generated in eLTER, namely:

- Unique, persistent and resolvable identification
- Machine-readable representation that are queryable using an API
- Legal right to use the RO in the context of eLTER

Should these requirements be fulfilled, a usage of the RO in eLTER is possible. To illustrate this with an example: The CORDIS database provides information on EU research projects, such as eLTER PLUS and issues DOIs for each project, e.g. https://doi.org/10.3030/820852.

Data generated within such a project could store this information and the associated DOI as part of its metadata and hence provide additional context information for the dataset.

3 Potential applications

The following section illustrates possibilities to fetch ROs from different sources and connect them to create a benefit and enable functionality that goes beyond what a single component could do by itself.

3.1 Connecting research objects to create site information clusters

Deliverable D4.1 of the eLTER PLUS project (Peterseil et al., 2022) already describes the concept of site information clusters. The main idea behind the concept of site information clusters is to simplify the harvesting and user uptake of data from multiple information sources, thereby facilitating the integration with eLTER data by making use of existing data and services. For this, a selection of relevant data sources is needed. This is a process informed by the framework of the eLTER Standard Observations (SO; Zacharias 2021).
However, describing a generic workflow for data source identification and retrieval as well as the identification of data types and sources are just the first step in the construction of Information Clusters (Peterseil et al. 2022).

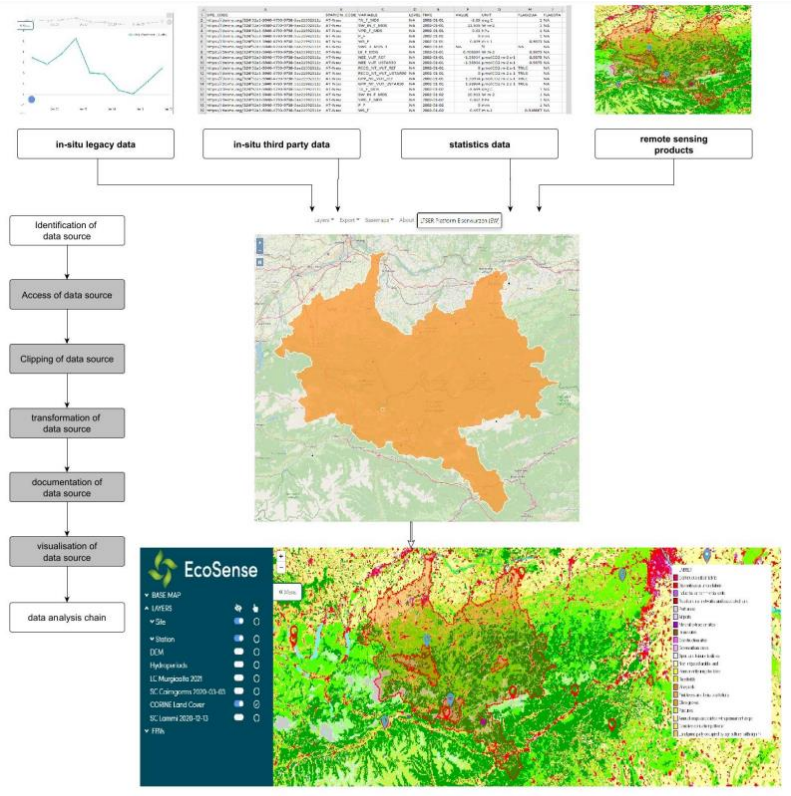

Figure 4 Schematic concept of information clusters (Peterseil et al. 2022)

The design and implementation of the workflows is intended to assist the development of new services, by establishing the structures for incorporating information into the extended eLTER Data Integration Platform (DIP) mentioned in section 2.1.3.
In order to create site information clusters for eLTER sites, existing site information from DEIMS-SDR can be taken and extended with information coming from other sources. For this deliverable, a prototype for such a site information cluster was built (Site Catalogue Github

2023) using the eLTER DataLabs illustrating functionality that goes beyond what the eLTER DIP or EcoSense do at the moment.
The prototype fetches information on all eLTER sites from DEIMS-SDR, i.e. a subset of site data, joins them with socio-economic indicators calculated using the eLTER Cookie Cutter (eLTER Cookie Cutter 2023) and provides an interface to display and query this information. The general principle for this is illustrated in Figure 5.

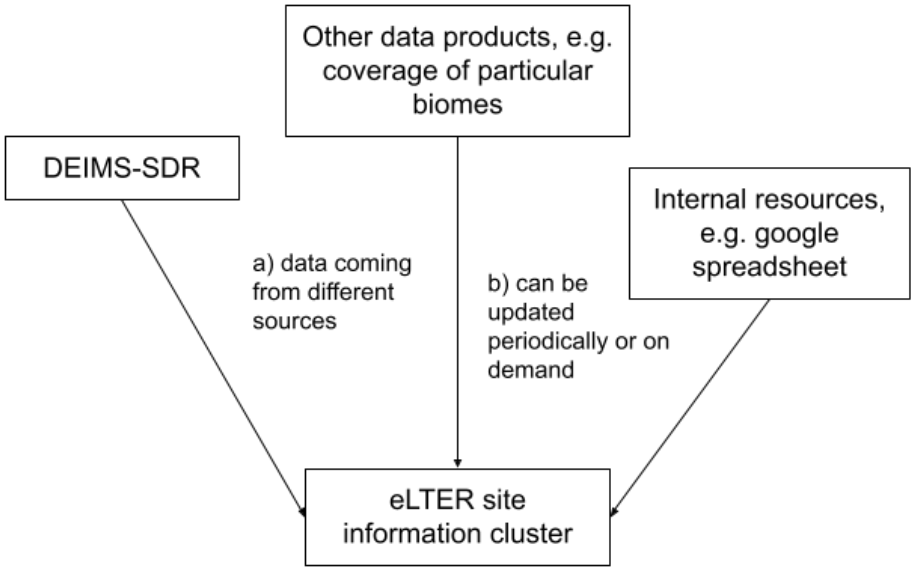

Figure 5 Data sources for an eLTER site information cluster

The resulting tool is a web application and service where eLTER sites can be visualised and queried based on the linked data sources Figure 6.

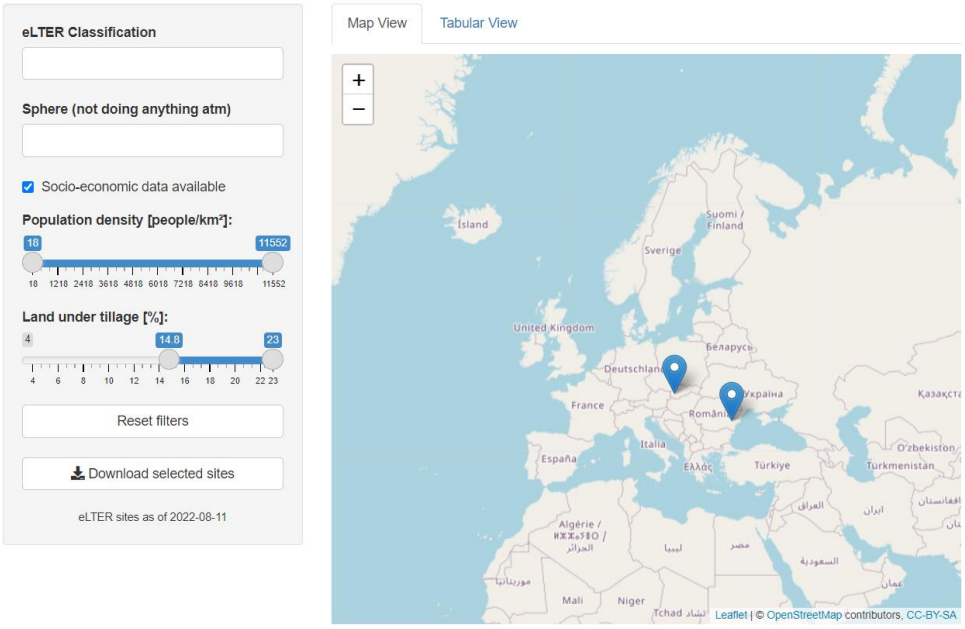

Figure 6 Filtering eLTER sites by socio-economic data using an information cluster

In addition to the socio-economic data, other data sources (or ROs) could be added to the prototype and additional facets could be added. The resulting interface enables more targeted

queries in a user-friendly interface in comparison to standard components of the eLTER IS such as DEIMS-SDR. It also eases working with subsets of sites, e.g. all sites conducting socio-economic research instead of all eLTER sites.

3.2 RDF representation of eLTER metadata and data

This section describes the work done in creating an RDF representation of selected research objects currently available in the eLTER IS. The latest version of the code generating this RDF representation is publicly available (eLTER RDF GitHub 2022) as well as a snapshot version on Zenodo (Alessandro Oggioni et al. 2022) for references.

The FAIR data maturity model (Research Data Alliance FAIR Data Maturity Model Working Group 2020) is a framework by the Research Data Alliance (RDA) for evaluating FAIRness of research data, i.e. the compliance with the aforementioned FAIR principles. Three criteria highlight the suitability of representing (meta-) data as RDF in order to enhance their machine understandability:

1. RDA-I1-02M Metadata uses machine-understandable knowledge representation
2. RDA-I1-02D Data uses machine-understandable knowledge representation
3. RDA-I3-01D Data includes references to other data

RDF meets all of these criteria. Existing ROs described in DEIMS-SDR were therefore mapped and transformed to an RDF representation. This was done by defining an appropriate and machine-actionable mapping of those documents into RDF-encoded ROs following authoritative and well-known ontologies and RDF schemas.
The RO types “site”, “activity”, “dataset” and “sensor” are retrievable as JSON documents using the DEIMS REST-API and were the basis for the mapping. In addition, the entity types “person” and “network” were also used for the mapping. Networks, a group of sites, e.g. all eLTER sites, and persons are part of the site documentation but are not available as separate entities in the REST-API. Nevertheless, these RO types were also used in the generation of RDF by extracting this information from other records.
Table 2 provides an overview of all mapped entities, the ontologies used for their RDF description, along with the authority, consortium or community under which they are developed or exploited.

Table 2 Ontologies used in the RDF mapping of DEIMS entities

DEIMS-SDR entity	Ontologies (namespace - details)	Authority/Consortium/Community of reference
Network	EF- SmOD, Environmental Monitoring Facility Vocabulary (Tasarova, T., Leadbetter, A. and Marine Institute 2017) RDF (RDF Working Group 2014a) RDFS (RDF Working Group 2014b)	INSPIRE, Environmental monitoring facility World Wide Web Consortium (W3C) Sea Data Net
Site	EF DCT - Dublin core (DCMI Usage Board 2020) GeoSPARQL (Knibbe, F., Herring, J., Abhayaratna, J., Bonduel, M., Perry, M., Car, N.J., Cox, S.J.D., & Homburg, T. 2021)	Dublin Core™ Metadata Initiative Open Geospatial Consortium (OGC)

	OWL - Web Ontology Language 2004 DCAT 2020 RDFS Vcard (McKinney, J. & Iannella, R. 2014)	
Activity	SmOD DCT - SF GeoSPARQL OWL DCAT RDFS vcard	
Dataset	DCAT DCT vcard RDF	
Sensor	SOSA - SSN - RDF RDFS GeoSPARQL SF OWL Geo - DCAT vcard	
Person	vcard	

The proposed mapping adopts technologies for the semantic web and follows international recommendations for the representation of environment-related entities. Most notably, frameworks and ontologies such as RDF, RDFS or SSN are recommended for the semantic web. Semantic Sensor Network and Sensor Observation Sampling Actuator ontologies are both strictly linked to the Open Geospatial Consortium (OGC) initiative, which has defined several ISO standards. The SmOD Environmental Monitoring Facility Vocabulary (developed under the Smart Open Data project) is an ontology, described in RDF and derived from the Environmental and Monitoring Facilities (EMF) schema developed for the European Commission INSPIRE Directive (Tasarova, T., Leadbetter, A. and Marine Institute 2017).
While the adoption of many ontologies to represent ROs may increase the difficulty of understanding, already by the amount of acronyms adopted, they are necessary to improve translating the complexity of the concepts to be expressed.
Table 3 is an extract of the information provided in the repository, with the following columns:

- “jsonPath” describes how to reach a specific element in the DEIMS-SDR API JSON document
- “JSON data item example” describes the chunk of JSON code containing the element
- “translation into …” contains the destination RDF representation
- “notes” report possible annotations.

Table 3 Excerpt of table defining the mapping of sensors.

jsonPath	JSON data item example	Translation into RDF
$.concat($.id.prefix,$.id.suffix)	"id":{"prefix":"https://deims.org/sensors/","suffix":"115e55f5-25c6-45ee-b490-393269e7bb06"}	<https://deims.org/sensors/115e55f5-25c6-45ee-b490-393269e7bb06> a sosa:Sensor; rdfs:seeAlso <{URL pointing SensorML of https://deims.org/sensors/115e55f5-25c6-45ee-b490-393269e7bb06}> .
$.title	"title":"Climate Station M4 Snow"	<https://deims.org/sensors/115e55f5-25c6-45ee-b490-393269e7bb06> rdfs:label "Climate Station M4 Snow"@en .

For the complete mapping definition and extensive examples, additional material is provided in (Alessandro Oggioni et al. 2022).

The current RDF representation of entities is under discussion in the eLTER PLUS project WP10 and WP11. A new version will be produced to increase interoperability with other initiatives, as other communities are considering RDF profiles exploiting other ontologies, in particular for tracking provenance information of research data. In this regard, the use of PROV ontology (PROV Ontology; Lebo, T., Sahoo, S., & McGuinness, D. 2013) is under discussion in order to link the different entities in a way to not only be compatible with European Union frameworks (e.g. under the INSPIRE framework like it has been done until now with EF), but also with other communities (e.g. with TERN, UK-NOCand SeaDataNet).

3.3 RI-wide search functionality

A recurring requirement formulated in different work packages of eLTER PLUS is the ability to search across the entire network. One option to create search functionality across the planned research infrastructure is to register all ROs in a search index and make it queryable through a centralised interface. There is a variety of options to enable such functionality, such as Solr or ElasticSearch. A prerequisite for this is to be able to programmatically fetch all available ROs. For this, APIs are needed that can query all available ROs. DEIMS-SDR already has such an API for sites and a beta version for an API for sensors.
A centralised search index would also enable working with additional data sources as long as there is enough harmonisation of the data models and a way to query that information automatically, i.e. an API. Furthermore, a centralised search interface enables additional functionality, such as providing graphical overviews of data availability further described in the following section.
Machine-learning algorithms could be used for matching keywords coming from different codes lists/thesauri to ensure interoperability of research objects and enabling user-friendly search facets.
Another way to realise this is to capitalise on the RDF representation of ROs as described in the previous section. SPARQL is a standard query language for RDF graphs. Representing the ROs in RDF would therefore allow setting up a SPARQL endpoint that enables queries across all ROs stored in the graph. While such an interface would require basic SPARQL knowledge to actually be able to work with and query the endpoint, it would be a comparatively fast and simple way to make ROs coming from different systems searchable and enable some of the functionality described in the following section.

3.4 Data availability reports and data visualisation

Once ROs are available in a central graph or search index, data could be grouped by sites and data type and show available data in a web user interface. The NEON data product catalogue (NEON 2023; Figure 7) exemplifies how such an overview can be presented.

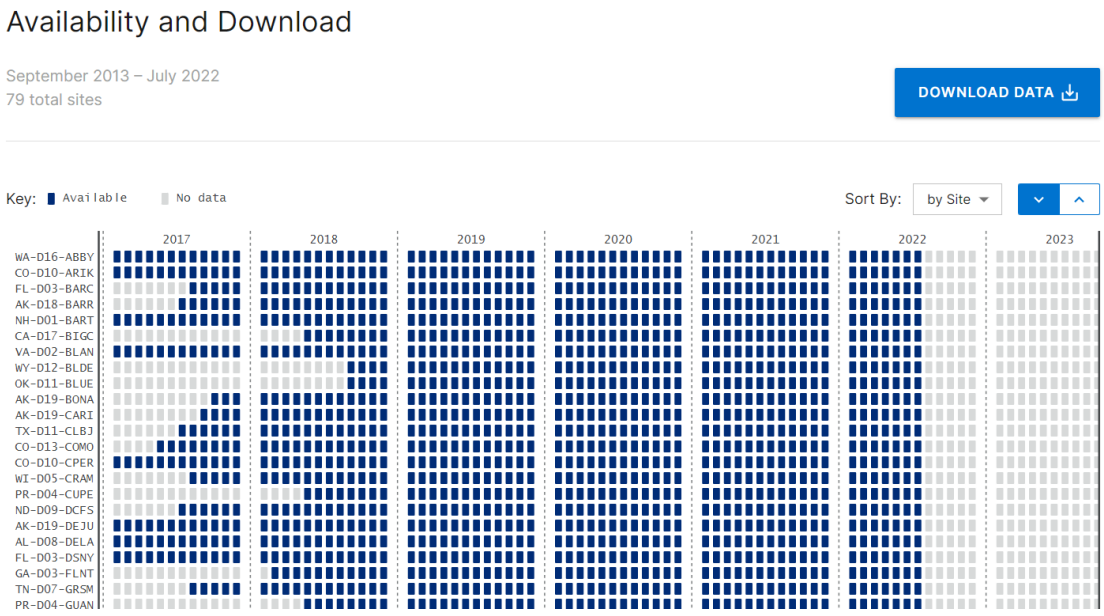

Figure 7 Data availability per site chart (NEON Data Portal, 2022)

Figure 7 shows the availability of a particular data product at different sites and time periods. Once a time-series dataset is selected, the data can be visualised and compared with data in the same period from another site as illustrated in Figure 8.

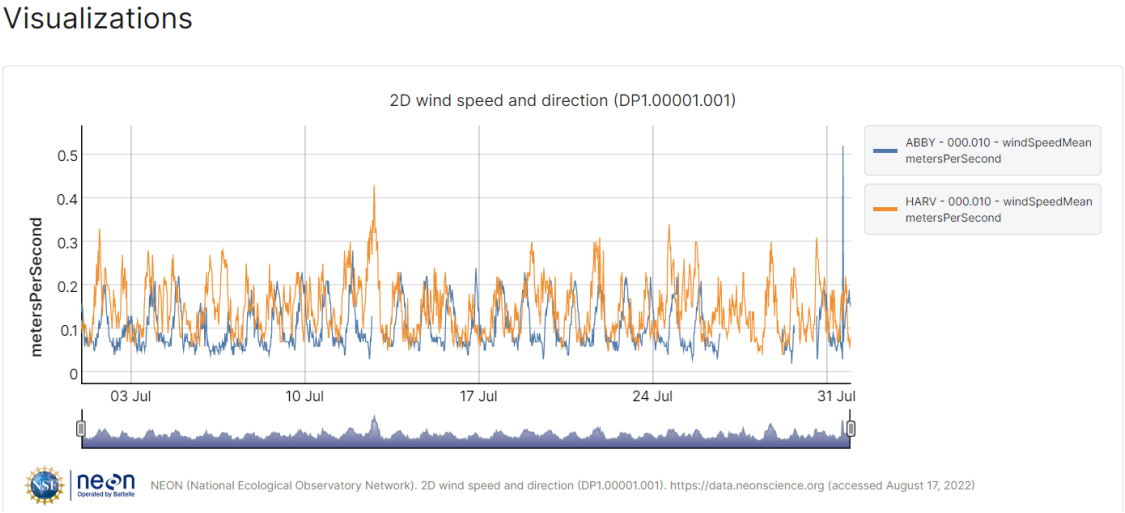

Figure 8 Data visualisation (NEON Data Portal, 2022)

An interface like this would also enable to search such an index based on particular values of thresholds, e.g. to list all sites where nitrogen deposition exceeds a defined critical threshold.

Push alerts could be tied to such thresholds to indicate if a critical load has been exceeded in a certain site.

3.5 Deriving provenance chains

As previously mentioned, it is required for an RO to have a persistent, unique and resolvable identifier. Such a PID can be used to query the documentation of each RO using an API. If a data product, e.g. a modelled GeoTIFF contains the information about all ROs that led to its creation, this enables creating an information chain detailing its provenance and all relevant processing steps. Another example for this are samples taken at various eLTER sites. Each sample would receive an identifier. Based on all samples, an analysis can be carried out using a particular method leading to a published dataset with a DOI. The used method as well as all sampled sites would already have an identifier. Now every step of the sampling and analysis process can be referenced thus enabling the visualisation of a provenance chain.
This chain would not only be available in a machine-readable manner but could also be graphically displayed to end users as exemplified in Figure 9.

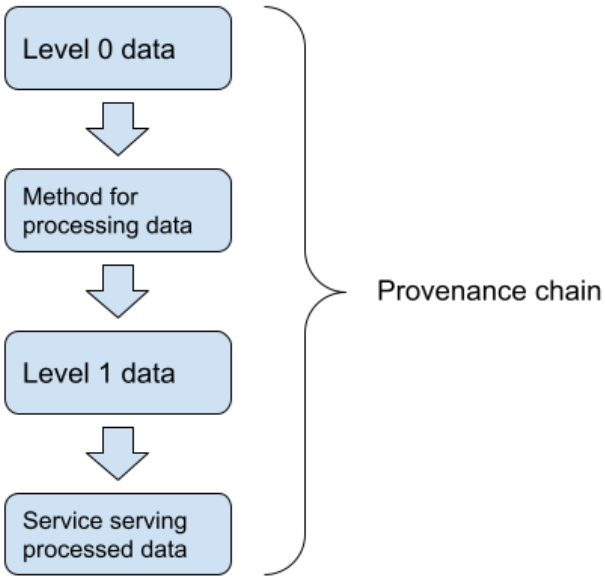

Figure 9 Schematic display of a provenance chain of ROs

This would greatly increase the intelligibility of datasets and their metadata and help to increase the reproducibility of scientific work as described in the introduction.

4 Discussion

In principle, it is possible to implement the RO concept in eLTER. In part this has already been done inherently by work carried out in eLTER PLUS and previous projects. Most of the relevant ROs such as sites, sensors, datasets, models and code can already be documented and assigned a PID using existing components, such as the DAR or DEIMS-SDR. Preliminary implementations capitalising on this have been demonstrated in the scope of this deliverable, e.g. by generating RDF representations of selected RO types. While some of the RO types have a more mature data model, such as sites while others, such as sensors, still need to be further refined.

Publishing software code in software development oriented repositories or services such as GitHub or Gitlab has many benefits: support for version management, test automation,

deployment pipelines and other software industry tools, and the possibility to benefit from the wider open source community. However, choosing a suitable code repository is not trivial. Features, costs and terms of use vary strongly between services. Setting up a dedicated code repository for eLTER would be an option to tackle some of these issues, if the benefits justify the required investment costs.
Code repositories also do not necessarily support the research oriented publication workflows and needs (research metadata, referencing to other ROs, citability, exposing related contributions as merits etc.), so it is worthwhile to separately document code/method elsewhere. Publishing a code "snapshot", like done for the code to generate RDF in section 3.2, in a research data repository is a self-contained option that enables citability of code and referencing it in provenance chains. While such an approach also has downsides, for instance snapshots becoming outdated, the benefits of provenance embedment outweigh this.

However, even with in parts comprehensive possibilities to document particular RO types, it is currently not possible to document the entire provenance chain of a generated dataset. There is a gap in regards to the documentation of methods, samples, specimens or digital evidence at the moment. To fill this gap, either international registries should be capitalised on or an existing system, for example the DAR, should be extended to enable documenting these RO types before linking them to evolving identifiers. Regardless of this restriction, one of the main conclusions drawn from this deliverable is that the immediate next steps should be meeting the requirements of the ROs concept and increasing its usage.

To increase the usage and distribution of ROs, established frameworks and formats should be stronger capitalised on, especially FDOF and RO-Crates. Following the basic guidelines of FDOF is in line with the design of RO-Crates and therefore both concepts can be pursued simultaneously. Since RO-Crates relies on JSON-LD and JSON-LD is designed around the concept of a "context" to provide additional mappings from JSON to an RDF model, both RO-Crate and FDOF can be aligned with relative ease. It is therefore less about which format is superior and should be pursued but instead the main goal has to be that all requirements of these frameworks should be met. Once this is done, the implementation of the actual encoding in each format is a comparatively simple and straightforward process which can be done quickly as soon as requirements are met. Therefore, existing systems should be modified or extended to meet those requirements and implement the corresponding data formats afterwards. Most importantly, this includes the issuing and usage of PIDs for every RO in eLTER as well as the persistent storage of the RO itself. For storing ROs, external services such as EUDAT B2SHARE or Zenodo might also be used. For the documentation of data lineage/provenance, it is necessary that all relevant steps that led to the creation of a dataset are documented and that this documentation is openly available and can be accessed by external users as well.
This means that not only PIDs need to be issued for all relevant ROs but that APIs need to either be built or enhanced to serve RO information using the PIDs. For people, such an identifier could be the ORCID, which is already well established or other of the discussed PIDs.

A possible workflow for creating provenance chains could be that for every data product generated in eLTER not only the data product itself but also the list of ROs used for the creation of the RO should be provided. This list could consist of the type of RO (e.g. level 0 data, used methods, people responsible for the processing) as well as their PID. EcoSense could then visualise the data product and also supply information on how the data product was generated.

Next steps to be done are therefore to raise awareness in the community of the use of ROs to provide better description of research activities so that they can be more easily discovered and acknowledged (e.g. through referencing particular source data or methods as well as via citation of the final paper output). To enable a move toward better documentation and citation of the different elements for the research workflow using ROs, ICT tools need to be developed

and promoted with user communities such as eLTER. The current ICT tools in eLTER can be used to promote these working practices and evolve to accommodate new international standards in how supporting materials are referenced (e.g. via PID registries). This will be a key topic for ICT development plans and eLTER training courses within existing projects and when the operational ESFRI is formed.

Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 871128 (eLTER PLUS).

References

Agrosens (2022). Available online at https://agrosens.rs/#/app-h/welcome.

Alessandro Oggioni; Paolo Tagliolato; CristianoFugazza (2022): IREA-CNR-MI/eLTER-RI_RDF: eLTER-RI2RDF: Zenodo.

Bechhofer, Sean; Roure, David de; Gamble, Matthew; Goble, Carole; Buchan, Iain (2010): Research Objects: Towards Exchange and Reuse of Digital Knowledge. In *Nat Prec*. DOI: 10.1038/npre.2010.4626.1.

Belhajjame, Khalid; Zhao, Jun; Garijo, Daniel; Gamble, Matthew; Hettne, Kristina; Palma, Raul et al. (2015): Using a suite of ontologies for preserving workflow-centric research objects. In *Journal of Web Semantics* 32, pp. 16–42. DOI: 10.1016/j.websem.2015.01.003.

CDN (2022). Available online at https://cdn.lter-europe.net, checked on January 2023.

CESSDA Data Catalogue (2023). Available online at https://datacatalogue.cessda.eu/, checked on January 2023.

Copernicus Climate Change Service (2020): E-OBS daily gridded meteorological data for Europe from 1950 to present derived from in-situ observations.

Copernicus DEM (2016). Available online at https://land.copernicus.eu/imagery-in-situ/eu-dem/eu-dem-v1.1, checked on January 2023.

CORINE Land Cover (2018). Available online at https://land.copernicus.eu/pan-european/corine-land-cover, checked on January 2023.

Costello, Mark J.; Michener, William K.; Gahegan, Mark; Zhang, Zhi-Qiang; Bourne, Philip E. (2013): Biodiversity data should be published, cited, and peer reviewed. In *Trends in ecology & evolution* 28 (8), pp. 454–461. DOI: 10.1016/j.tree.2013.05.002.

CrossRef (2023). Available online at https://www.crossref.org/services/funder-registry/, checked on January 2023.

Darwin Core Task Group (2009): Darwin Core. Biodiversity Information Standards (TDWG). Available online at http://www.tdwg.org/standards/450.

DataCite (2023a): IGSN. Available online at https://support.datacite.org/docs/about-igsn-ids-for-material-samples, checked on January 2023.

DataCite (2023b): Tracking metadata provenance. Available online at https://support.datacite.org/docs/tracking-provenance, checked on January 2023.

DataCite Metadata Working Group (2019): DataCite Metadata Schema Documentation for the Publication and Citation of Research Data v4.3. With assistance of Madeleine de Smaele, Robin Dasler, Jan Ashton, Sophie Roy, Martin Fenner, Mark Jacobson et al.

DCAT (2020). Available online at https://www.w3.org/TR/vocab-dcat-1/, checked on Jan 2023.

DCMI Usage Board (2020): Dublin core DCMI Metadata Terms. Available online at https://www.dublincore.org/specifications/dublin-core/dcmi-terms/, checked on January 2023.

DIP (2023). Available online at https://dip.lter-europe.net, checked on January 2023.

EcoSense (2023). Available online at https://ecosense.biosense.rs, updated on January 2023.

eLTER Cookie Cutter (2023). Available online at https://github.com/eLTER-RI/spatial-data-processor, checked on January 2023.

eLTER RDF GitHub (2022). Available online at https://github.com/IREA-CNR-MI/eLTER-RI_RDF, checked on January 2023.

EnvThes (2023). Available online at https://vocabs.lter-europe.net/envthes/en/, checked on January 2023.

e-shape (2019), checked on January 2023.

European Commission, Directorate-General for Research and Innovation, Corcho, O., Eriksson, M., Kurowski, K., et al., EOSC interoperability framework (2021): Report from the EOSC Executive Board Working Groups FAIR and Architecture. Available online at https://data.europa.eu/doi/10.2777/620649, checked on January 2023.

European Commission, Directorate-General for Research and Innovation, Hellström, M., Heughebaert, A., Kotarski, R., et al. (2022): A Persistent Identifier (PID) policy for the European Open Science Cloud (EOSC). Available online at https://data.europa.eu/doi/10.2777/926037, checked on January 2023.

FDOF (2022). Available online at https://fairdigitalobjectframework.org/, checked on January 2023.

Hanson, Brooks; Sugden, Andrew; Alberts, Bruce (2011): Making data maximally available. In *Science (New York, N.Y.)* 331 (6018), p. 649. DOI: 10.1126/science.1203354.

ICOS (2023): ICOS Data Products. Available online at https://www.icos-cp.eu/data-products, checked on January 2023.

Knibbe, F., Herring, J., Abhayaratna, J., Bonduel, M., Perry, M., Car, N.J., Cox, S.J.D., & Homburg, T. (2021): GeoSPARQL Ontology. Available online at https://opengeospatial.github.io/ogc-geosparql/geosparql11/index.html, checked on January 2023.

Kobler, Johannes; Pröll, Gisela; Dirnböck, Thomas (2021): LTER Zöbelboden, Austria, Intensive Monitoring Plots, Forest inventory and tree biomass data 2003 - 2019.

Koivula, H., et al. (2022): Concept and extension of central data node capabilities. Deliverable D11.2 EU Horizon 2020 eLTER PLUS Project, Grant agreement No. 871128.

Krahl, Rolf; Darroch, Louise; Huber, Robert; Devaraju, Anusuriya; Klump, Jens; Habermann, Ted et al. (2021): Metadata Schema for the Persistent Identification of Instruments.

Lebo, T., Sahoo, S., & McGuinness, D. (2013): PROV-O: The PROV Ontology. Available online at https://www.w3.org/TR/prov-o/, checked on January 2023.

Matthew B. Jones; Margaret O'Brien; Bryce Mecum; Carl Boettiger; Mark Schildhauer; Mitchell Maier et al. (2019): Ecological Metadata Language version 2.2.0: KNB Data Repository.

McKinney, J. & Iannella, R. (2014): vCard Ontology - for describing People and Organizations. Available online at https://www.w3.org/TR/vcard-rdf/, checked on January 2023.

NEON (2023): NEON Data Portal. Available online at https://data.neonscience.org/data-products/explore, checked on January 2023.

OGC (2012): SOS. Available online at https://www.ogc.org/standards/sos, checked on January 2023.

OWL - Web Ontology Language (2004). Available online at http://www.w3.org/TR/owl-ref.

Parland-von Essen, Jessica; Fält, Katja; Maalick, Zubair; Alonen, Miika; Gonzalez, Eduardo (2018): Supporting FAIR data: categorization of research data as a tool in data management. In *INF* 37 (4). DOI: 10.23978/inf.77419.

Peterseil et al. (2022): Workflows for retrieval and harmonisation of legacy data. Deliverable D4.1 EU Horizon 2020 eLTER PLUS Project, Grant agreement No. 871128. 62pp + Annexes (1-5).

Rauber, Andreas; Gößwein, Bernhard; Zwölf, Carlo Maria; Schubert, Chris; Wörister, Florian; Duncan, James et al. (2021): Issue 3.4, Fall 2021. In *Harvard Data Science Review* 3 (4). DOI: 10.1162/99608f92.be565013.

RDF Working Group (2014a): RDF 1.1 Concepts and Abstract Syntax; W3C Recommendation. Available online at https://www.w3.org/TR/rdf11-concepts/, checked on January 2023.

RDF Working Group (2014b): RDF Schema 1.1. W3C Recommendation. Available online at https://www.w3.org/TR/rdf-schema/, checked on January 2023.

Research Data Alliance FAIR Data Maturity Model Working Group (2020): FAIR Data Maturity Model: specification and guidelines - draft.

rocrate Python Package (2022), updated on May 2022, checked on January 2023.

Sefton, Peter; Ó Carragáin, Eoghan; Soiland-Reyes, Stian; Corcho, Oscar; Garijo, Daniel; Palma, Raul et al. (2022): RO-Crate Metadata Specification 1.1.2.

Site Catalogue Github (2023). Available online at https://github.com/stopopol/site_catalogue, checked on January 2023.

Soiland-Reyes, Stian; Sefton, Peter; Crosas, Mercè; Castro, Leyla Jael; Coppens, Frederik; Fernández, José M. et al. (2022): Packaging research artefacts with RO-Crate. In *DS* 5 (2), pp. 97–138. DOI: 10.3233/ds-210053.

SoilGrids (2023). Available online at https://soilgrids.org/, checked on January 2023.

Stocker, Markus; Darroch, Louise; Krahl, Rolf; Habermann, Ted; Devaraju, Anusuriya; Schwardmann, Ulrich et al. (2020): Persistent Identification of Instruments. In *Data Science Journal* 19, Article 18. DOI: 10.5334/dsj-2020-018.

Tasarova, T., Leadbetter, A. and Marine Institute (2017): SmOD Environmental Monitoring Facility Vocabulary. Available online at https://www.w3.org/2015/03/inspire/ef, checked on January 2023.

W3C (2013): PROV-DM. Available online at https://www.w3.org/TR/prov-dm, checked on January 2023.

Wikipedia (2021): Research Object. Available online at https://en.wikipedia.org/wiki/Research_Object, updated on May 2021, checked on January 2023.

Wilkinson, Mark D.; Dumontier, Michel; Aalbersberg, I. Jsbrand Jan; Appleton, Gabrielle; Axton, Myles; Baak, Arie et al. (2016): The FAIR Guiding Principles for scientific data management and stewardship. In *Scientific data* 3, p. 160018. DOI: 10.1038/sdata.2016.18.

Wohner, Christoph; Peterseil, Johannes; Klug, Hermann (2022): Designing and implementing a data model for describing environmental monitoring and research sites. In *Ecological informatics* 70. DOI: 10.1016/j.ecoinf.2022.101708.

Zacharias, S. et al. (2021): Discussion paper on eLTER Standard Observations (eLTER SOs). Deliverable D3.1 EU Horizon 2020 eLTER PLUS Project, Grant agreement No. 871128. Available online at https://elter-ri.eu/storage/app/uploads/public/62c/ea2/a00/62cea2a002845239798196.pdf, checked on January 2023.