Project Title:	**ECOPOTENTIAL: IMPROVING FUTURE ECOSYSTEM BENEFITS THROUGH EARTH OBSERVATIONS**
Project number:	641762
Project Acronym:	ECOPOTENTIAL
Proposal full title:	DATA SETS ON THE HISTORIC EVOLUTION OF THE ECOSYSTEMS
Type:	Research and innovation actions
Work program topics addressed:	SC5-16-2014: "Making Earth Observation and Monitoring Data usable for ecosystem modelling and services"

Deliverable No: D5.6

HARMONISED DELIVERY OF DATA

Due date of deliverable:	31.05.2018
Actual submission date:	31.05.2018
Version:	1.0
Main Authors:	Barbara Magagna, Johannes Peterseil, Christoph Wohner, Sebastian Kirchhof & Samuel Bosch

This·project·has·received·funding·from·the·*European·Union's·Horizon·2020·research· and·innovation·programme*·under·grant·agreement·No·641762¶

Project ref. number	**641762**
Project title	**ECOPOTENTIAL: IMPROVING FUTURE ECOSYSTEM BENEFITS THROUGH EARTH OBSERVATIONS**

Deliverable title	HARMONISED DELIVERY OF DATA
Deliverable number	D5.6
Deliverable version	1.0
Contractual date of delivery	31.05.2018
Actual date of delivery	31.05.2018
Document status	Final
Document version	1.0
Online access	http://vocabs.ceh.ac.uk/evn/tbl/EnvThes.evn
Diffusion	public
Nature of deliverable	Report + online material
Workpackage	WP5
Partner responsible	EAA
Author(s)	Barbara Magagna, Johannes Peterseil, Christoph Wohner, Sebastian Kirchhof & Samuel Bosch
Editor	Carmela Marangi
Approved by	Carl Beierkuhnlein, Joan Maso
EC Project Officer	Gaëlle Le Bouler

Abstract	In order to assess the effects of global change on the environment and ecosystem data from observations as well as remote sensing are needed. A wide range of applications is used to conduct these analyses. Inter-alia in-situ data are used to calibrate models on different scales, validate Earth Observation products as well as calculate relevant indices to assess the magnitude and direction of change. The usage of a common semantics as well as clear documented data formats will ease the re-usability of the data. The report focuses on the implementation of EnvThes as well as on an overview of relevant data specifications which could be applied for protected areas. Task 5.6 aimed to contribute to enhance the interoperability of datasets with regard to the framework, protocols (methods, temporality) or geographic characteristics (projection / coordinates). This includes the harmonisation of data formats (syntactic interoperability), as well as the temporal and spatial resolution and thematic differences (semantic interoperability). The work resulted in the development of EnvThes a

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 641762

	common vocabulary for the annotation of dataset keywords and parameter names as well as in the evaluation of relevant data formats for the provision of data. Basic recommendations are given.
Keywords	ECOPOTENTIAL, in-situ, dataset, vocabulary, EnvThes, data format, SKOS, Environmental Monitoring Facility, standardisation

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 641762

Co-funded by the European Union

Table of Contents

This·project·has·received·funding·from·the·*European·Union's·Horizon·2020·research· and·innovation·programme*·under·grant·agreement·No·641762¶

List of Figures

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 641762

List of Tables

This·project·has·received·funding·from·the·*European·Union's·Horizon·2020·research·and·innovation·programme*·under·grant·agreement·No·641762¶

Executive summary

In order to assess the effects of global change on the environment and ecosystems a wide range of data is needed. This requires the availability of data and results from environmental observation and monitoring (e.g. in-situ monitoring campaigns, GEOSS, COPERNICUS, Earth Systems Data Cube) as well as, environmental and ecological modelling. Currently, a wide range of applications is used to conduct scientific analyses. Inter-alia in-situ data are used to calibrate models on different scales, validate Earth Observation products as well as calculate relevant indices to assess the magnitude and direction of change. A better integration of (spatial) data capture, modelling, management, sharing and access (big, open and linked data) workflows in thematic applications, systems and (cyber) geographic information infrastructures development framework is required.

In-situ observations resulting from a distributed network of observation sites are are characterised by a high heterogeneity. Data used in the analysis are often collected with a different focus (e.g. management of protection status). Even in long term monitoring networks like ILTER not only inter- but also intra-network heterogeneity results in a number of different interoperability issues which need to be addressed local as well as on global scale. In general this applies also for protected areas as a common coordination and harmonisation framework in terms of data collection is missing.

The usage of a common semantics as well as clear documented data formats will ease the re-usability of the data. The report focuses on the implementation of a common vocabulary as well as on an overview of relevant data specifications which could be applied for protected areas. The work builds on previous results and resulted in the extension of *EnvThes* (*Environmental Thesaurus*) a common vocabulary for the annotation of dataset keywords as well as in the evaluation of relevant data formats for the provision of data. Basic recommendations are given.

1 Introduction

In order to assess the effects of global change on the environment and ecosystems a wide range of data is needed. This requires the availability of data and results from environmental observation and monitoring (e.g. GEOSS, COPERNICUS, Earth Systems Data Cube or in-situ observation networks and RIs) as well as, environmental and ecological modelling. The better integration of (spatial) data capture, modelling, management, sharing and access (big, open and linked data) workflows included in thematic applications, systems and (cyber) geographic information infrastructures development framework is needed. This addresses not only the technical aspects but also information needed to assess the fitness for use of the data. The enhancement on data quality assessment, management processes, knowledge management, reinforcement of training opportunities/education, and individual/organizations capacity building needs, as well as improving communication and technical-political decision processes is one aspect in enhancing the data reusability (Alonso et al. 2017).

In-situ observations resulting from a distributed network of observation sites are are characterised by a high heterogeneity. Data used in the analysis are often collected with a different focus (e.g. management of protection status). Even long term monitoring networks like ILTER are characterised by inter- but also intra-network heterogeneity (see e.g. Vanderbilt et al. 2015) resulting in a number of different interoperability issues. These need to be solved on the local as well as on the global scale. In general this applies also for protected areas as a common coordination and harmonisation framework in terms of data collection is missing.

Currently a broad variety of data management solutions, applications and data formats exist. But also on the socio-cultural aspect of data integration and interoperability cultural differences and related implementations (e.g. data sharing rules) are important and differ largely across the participating PAs. These can be described by different interoperability issues ranging from technical to conceptual interoperability (Turnitsa and Tolk 2006). Whereas the technical interoperability is addressed by the software solutions taken, the other levels need the involvement at least of the information managers within the community. Nevertheless, not all levels can be easily addressed of its organisational complexity and also the limitations of funding for the global scale services. Focusing on the following interoperability issues will be needed:

The *syntactic interoperability* ensures that a common structure to exchange information, i.e., a common data format is applied. This applies also to services for data and data query, metadata, discovery, and workflow integrations which need to follow standardised norms and schemas.

The *semantic interoperability* aims to use a common information exchange model (e.g. common reference lists, vocabularies or ontologies) in order to ensure an unambiguous definition of the content shared. This also includes the definition of the thematic, spatial, and temporal references in the network.

The *legal interoperability* ensures that policies applied on different scales (e.g. national regulations) could hamper the sharing and reuse of data. When using data on a global scale, e.g. for the definition of EBV, data integration supported by 'machine-to-machine' interaction must be ensured. This is often difficult due to varying provenance of authorship and ownership of data and requires the identification of legal and policy bottlenecks (see Kissling et al. 2015).

In practice, standardisation is never perfect, and mediation of any of the above mentioned interoperability issues are usually required. It is likely to require a brokering service to aggregate metadata from all participants and to allow data access in a federated system of systems.

1.1 ECOPOTENTIAL project context

ECOPOTENTIAL is a large European-funded H2020 project that focuses its activities on a targeted set of internationally recognised Protected Areas, blending Earth Observations from remote sensing and field measurements, data analysis and modelling of current and future ecosystem conditions and services. 54 partners coming from Universities and Research Institutes and Protected Areas Management Bodies/Management authorities from 24 sites across Europe, South Africa and Caribbean Sea constitute the ECOPOTENTIAL community. A detailed description for each partner can be found in the official website of the project[1] bringing expertise on different fields ranging from *Earth Observation Analysis*, *Ecological Modelling*, *Environmental Modelling*, and *In-situ Scientist* (Ecologist/Biologist).

ECOPOTENTIAL include a number of Protected Areas which cover several climate zones and biogeographic regions (Fig. 1-1) in Europe and beyond and have been classified into three distinct main ecosystem types: a) mountain ecosystems, b) transitional zones between aquatic (including marine) and terrestrial ecosystems, and c) semi-arid and arid ecosystems.

Fig. 1.1 The protected areas that participate in ECOPOTENTIAL.

[1] See http://www.ECOPOTENTIAL-project.eu/partners

Also in terms of protections ECOPOTENTIAL covers the most important protected areas in Europe. Fig. 1.1 shows the protection status for each of the protected areas. The protection status ranges from National parks to Natura 2000 having protected areas which show more than one protection status.

With regard to metadata and data provision, all different groups act mainly as data provider as well as data user. The importance of metadata in order to make datasets discoverable, accessible and understandable is therefore for all of them an issue. Within the project DEIMS-SDR[2] has been adopted as main metadata editor for in-situ observation data resulting from the project or as well for legacy data which are used in the project context. The common metadata models adopted and extended in the project context are outlined in deliverable D5.2 provided by Poursanides et al. (2017).

ECOPOTENTIAL is including the main user roles ranging from data providers, data consumers and data producers and thus addressing the full data life cycle.

Experts on *Earth Observation Analysis* are focusing on the analysis of Earth Observation (EO) data of different spatial/temporal resolutions and satellite platforms. The derived data (e.g. land cover, NDVI, LST, albedo, soil moisture, DSM) are used to identify ecological processes and changes (e.g. on land cover) over broad spatial and temporal scales. In addition, these, data are also used as input layers to models. The provision of sufficient metadata is an important task in order to allow the discoverability and re-usability of the data.

The *In-situ scientists* provide data for calibration and validation of both the production of the variables from the analysis of Earth Observation data as well as the training of the Ecological Modelling/Environmental Modelling. The in-situ data provide the information on the ecosystem status and behaviour often over a longer time span. I.e. design and instrumentation of the observation are important information, which needs to be recorded. The provision of metadata in order to ensure the reusability of the data is an important task, also because in-situ observation often cannot be repeated or redone once the data are lost because of poor documentation of the datasets.

Experts on *Ecological Modelling and/or Environmental Modelling* use data and products that come from the in-situ or EO data providers in order to model environmental or ecosystem processes (nutrients cycle, natural hazards, climate change, biodiversity distribution, etc.) or to forecast their behaviour in future. Metadata describing the datasets used in the models are important to understand the context, resolution and quality of the data and their fitness to use for the problem addressed.

1.2 Approach

Within ECOPOTENTIAL WP5 is focusing on the documentation and provision of in-situ data. The metadata models are outlined in deliverable D5.2 (Poursanides et al. 2017) and provide an important input to the current work. Task 5.6 focuses (a) on the evaluation of the relevance of existing data models and (b) provide a semantic backbone for the annotation of data. The current report provides an introduction to controlled vocabularies and the implementation of EnvThes. In the second part relevant data models are described.

The chapter on **Thesauri and controlled vocabularies** focuses on the background concepts needed for the development of a common vocabulary as well as the steps conducted for the further development of

[2] See https://data.lter-europe.net/deims/

EnvThes. Within the ECOPOTENTIAL project EnvThes (Environmental Thesaurus) was adopted as the common vocabulary used for in-situ data.

The chapter on **Data formats and structures** focuses on relevant data models for the data provision for in-situ data providers as well as the implementation of a common harmonised specification of the Environmental Monitoring Facility as interchange format for research sites.

The **Annexes** provide supplementary material for the report.

2 Thesauri and controlled vocabularies

2.1 Introduction

In order to allow for common analysis of data resulting from different research and experimental sites, a common semantic framework needs to be established in order to (a) discover and (b) integrate the data from different sources. The development of a common semantic backbone – defined as the common language between the different data providers and generators – was identified as one of the core elements for data integration (Oggioni et al. 2012).

"*People can't share knowledge if they don't speak a common language*" (Davenport 1997). But a common vocabulary alone does not guarantee that people understand each other. It is well known that the human language is prone to misunderstandings, misinterpretation and thus of information loss. Disclosure and transfer of knowledge can only succeed if the communication between all the involved works. Semantic resources such as classifications and thesauri aim for solving these problems, first of all for achieving disambiguation.

2.2 Semantic resources – an overview

2.2.1 Classification of semantic resources

In the web one can find different approaches to describe semantic resources with more or less expressivity.

Often used is the term *controlled vocabulary* to refer to taxonomies, thesauri and ontologies. But also glossaries and simple controlled lists (also known as reference lists) can be classified as controlled vocabularies, as they are means to organize knowledge for subsequent retrieval[3]. They are carefully selected list of words and phrases, which are used to tag units of information (document or work) so that they may be more easily retrieved by a search. Controlled vocabulary schemes mandate the use of predefined, authorised terms that have been preselected by developers of the schemes, in contrast to natural language vocabularies, which have no such restriction.

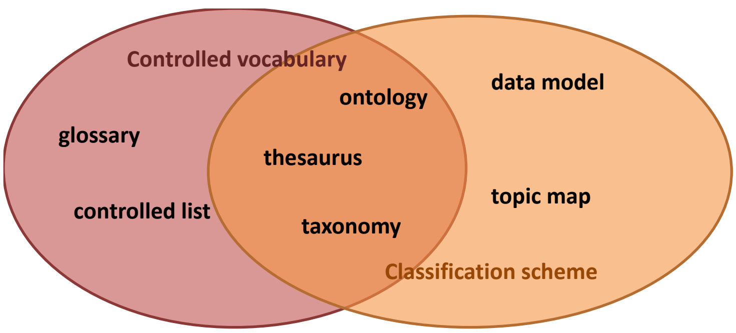


Fig. 2.1 Controlled vocabulary versus classification scheme

[3] https://en.wikipedia.org/wiki/Controlled_vocabulary, retrieved May 17th, 2018

They all have in common the focus on relationships between and among the concepts represented by the terms or names in a vocabulary. In library science also subject heading lists and synonym ring lists fall into this category (Harping 2010).

Another well-known concept that groups semantic resources is *classification scheme*. It arranges things into kinds of things (classes) or into groups of classes based on characteristics which the members have in common. Topic maps, data models as well as networks (in mathematics) are known as classification schemes. The resulting structures are crucial for metadata, often represented as a hierarchical structure and accompanied by descriptive information of the classes or groups. The ISO/IEC 11179 metadata registry standard uses classification schemes as a way to classify administered items, such as data elements, in a metadata registry.[4] According to quality criteria, such as whether there are overlaps between classes or whether polyhierarchy is allowed, it is possible to distinguish different types of classification schemes. While data models, topic maps and networks are clear types of this category, taxonomies, thesauri and ontologies have properties of both classification schemes and controlled vocabularies (see Fig. 2.1).

The term ontology has been used to describe models with different degree of structure known as *ontology spectrum*, which is somehow misleading as ontologies have also a very strict meaning explained in chapter 0. The concept here refers to the also often used term *semantic spectrum, semantic ladder* or *semantic precision* and is well captured by Fig. 2.17. It represents a series of increasingly precise, in other words, semantically expressive definitions of data elements in knowledge representations. There is always a trade-off between more precision and the ability to use tools to automatically integrate systems on the one hand and the cost of building and maintaining such systems.

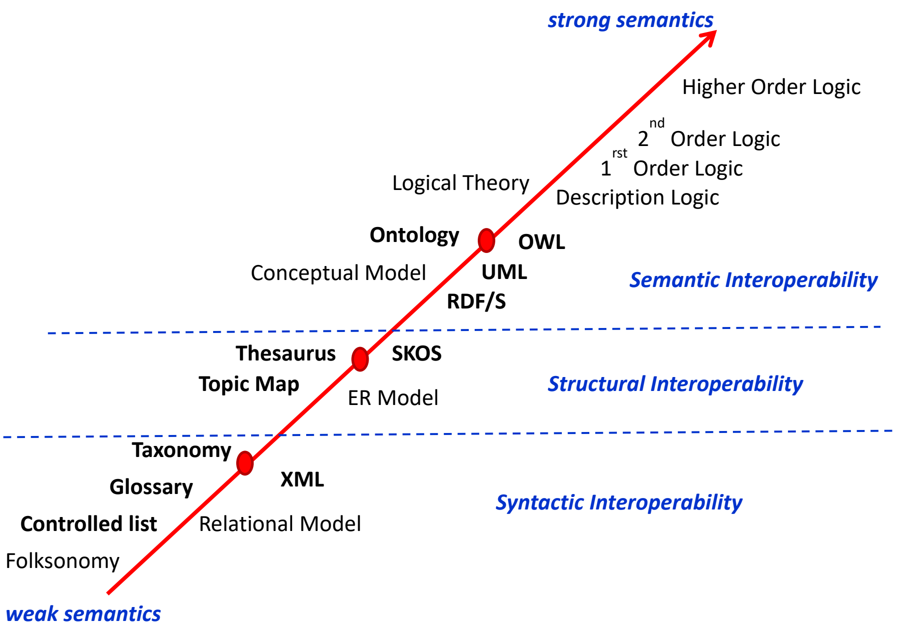


Fig. 2.2 Semantic spectrum (by Orbst 2006, altered)

[4] https://en.wikipedia.org/wiki/Comparison_and_contrast_of_classification_schemes_in_linguistics_and_metadata, retrieved May 17th, 2018

Syntactic interoperability, provided by for instance XML or the SQL standards, is a pre-requisite to semantic. It involves a common data format and common protocol to structure any data so that the manner of processing the information will be interpretable from the structure. It also allows detection of syntactic errors, which have to be resolved. However, where accurate translation of syntaxes is possible, systems using different syntaxes may also interoperate accurately.

Semantic interoperability[5] is the ability of computer systems to exchange data with unambiguous, shared meaning. Semantic interoperability is a requirement to enable machine computable logic, inference, knowledge discovery, and data federation between information systems.

Folksonomies: Folksonomy is the system in which users apply public tags to online items, typically to aid them in re-finding those items. This can give rise to a classification system based on those tags and their frequencies, in contrast to a taxonomic classification specified by the owners of the content when it is published. This practice is also known as collaborative tagging, social classification, social indexing, and social tagging. Folksonomy was originally "the result of personal free tagging of information [...] for one's own retrieval"[6], but online sharing and interaction expanded it into collaborative forms. Social tagging is the application of tags in an open online environment where the tags of other users are available to others. Collaborative tagging (also known as group tagging) is tagging performed by a group of users. This type of folksonomy is commonly used in cooperative and collaborative projects such as research, content repositories, and social bookmarking.

Controlled list: Controlled lists are also named as reference lists, reference data sets or code lists. A controlled list is a simple list of terms used to control terminology. In a list, each term is unique, also in meaning and the terms are all members of the same class. There are often employed in certain fields of a database where a short list of values is appropriate. But if they are codified in SKOS they can be also used to make other data meaningful and interpretable in an unambiguos way (e.g. metadata in catalogues). Translating from one reference list to another within the same domain is an essential need for ecologists.

Glossaries: A glossary, also known as a vocabulary or clavis, is an alphabetical list of a in a particular domain of knowledge with the definitions for those terms. Traditionally, a glossary appears at the end of a book and includes terms within that book that are either newly introduced, uncommon, or specialized. In a general sense, a glossary contains explanations of concepts relevant to a certain field of study or action.

Taxonomies: Taxonomy is the practice and science of classification of things or concepts, including the principles that underlie such classification. A taxonomy is the simplest variant as it contains only terms that are organized into a hierarchical structure.

---

5 https://en.wikipedia.org/wiki/Semantic_interoperability, retrieved May 17th, 2018

6 http://www.vanderwal.net/folksonomy.html, retrieved May 17th, 2018

Topic Maps: A topic map is a standard for the representation and interchange of knowledge, with an emphasis on the findability of information. Topic maps were originally developed in the late 1990s as a way to represent back-of-the-book index structures so that multiple indexes from different sources could be merged. However, the developers quickly realized that with a little additional generalization, they could create a meta-model with potentially far wider application[7]. The ISO standard is formally known as ISO/IEC 13250:2003.

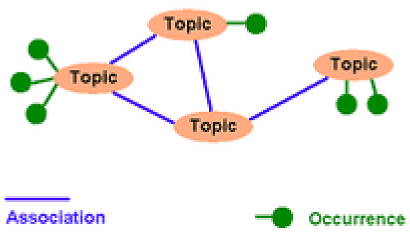


Fig. 2.3 Structure of the concepts using a Topic Map

A topic map represents information using: (a) topics, representing any concept, from people, countries, and organizations to software modules, individual files, and events, (b) associations, representing hypergraph relationships between topics, and (c) occurrences, representing information resources relevant to a particular topic.

Thesauri: The philosophical groundings for thesauri are concept-oriented and deals with relations between concepts on one hand and between concepts and labels on the other hand. The main purpose is the representation of concepts in the right order. Through the ages the concept of thesaurus evolved, in the antiquity it was understood as lexis, in the mediaeval it had the meaning of dictionary and today it depends on the context in which this term is used. In general uses, it is a reference work that lists words grouped together according to similarity of meaning (containing synonyms and sometimes antonyms)[8]. In information science, a thesaurus is a form of controlled vocabulary which organizes information and provides terminology to catalogue and retrieve information. It serves to minimise semantic ambiguity by ensuring uniformity and consistency.

A thesaurus is a special kind of a controlled vocabulary as it consists of a collection of concepts. A concept in a thesaurus is more than merely a word, but it is a 'unit of thought', which can have several labels, i.e. different words that are associated with it. These might be synonyms, abbreviations, spelling variants, and other words that can be used to refer to that concept. All this information can be stored with a concept in a thesaurus. Beyond that a thesaurus is also structured, as it describes the relationships between its concepts.

[7] https://en.wikipedia.org/wiki/Topic_map, retrieved May 17th, 2018
[8] https://en.wikipedia.org/wiki/Thesaurus, retrieved May 17th, 2018

Ontologies: According to Sowa (2000) an ontology is a formal, explicit specification of a shared conceptualization which can be described as abstract model of relevant concepts that is used by a community. It is an arrangement of URI concepts that are related by various well defined kinds of relations. The arrangement can be visualized in a directed acyclic graph. Most of the ontologies are expressing in OWL.

The basic components of ontologies are:

- Classes that represent (abstract or specific) concepts,
- Relations that specify the different types of associations between classes
- Functions specifying arguments in triples in specified relations
- Axioms expressing constant propositions and
- Instances representing concrete elements and individual objects.

2.3 Background

2.3.1 The meaning triangle

The meaning triangle or semiotic triangle can be traced back to the 4th century BC in Aristotle's De Interpretatione book II, but was first published by Ogden & Richards (1923). The triangle is a model that describes the relationship between thought (*reference*), linguistic sign (or *representamen*) and a referent (the things they try to represent or refer to). It has been adopted by many researchers from diverse viewpoints and the three vertices of the triangle got many different denominations (see Esser, 2015).

Essentially it says that there is no direct relationship between a spoken word (or sign, or symbol) and a real object in the world. The spoken word has to be processed by an interpreter (human or machine), which results in a concept. This concept can then be related to a concrete thing in the world. Ideally there is a clear relationship between these three vertices.

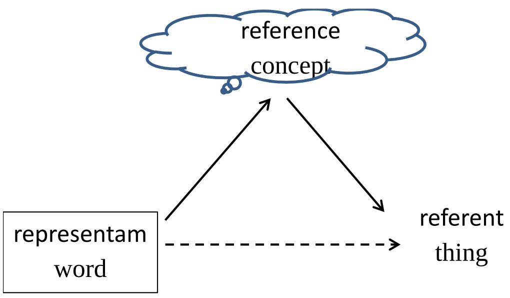


Fig. 2.4: The semiotic triangle

But human communication is not always straightforward. One word might evoke different interpretations. An example for this is the German word *Bank* which could mean *bank* or *bench* in English (Studer et al. 2001). Thus there are two different concepts for this term which refer to two distinct objects in the real world.

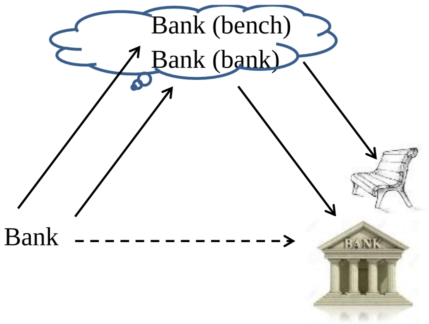


Fig. 2.5: Bank - ambiguous term

On the other hand one real object, e.g. the planet Venus, might be referred to in different contexts. The terms *morning star* and *evening star* are distinct terms that create different interpretations in the mind of the listener, namely a planet seen in the morning or in the evening (Frege cited in Sowa, 2000).

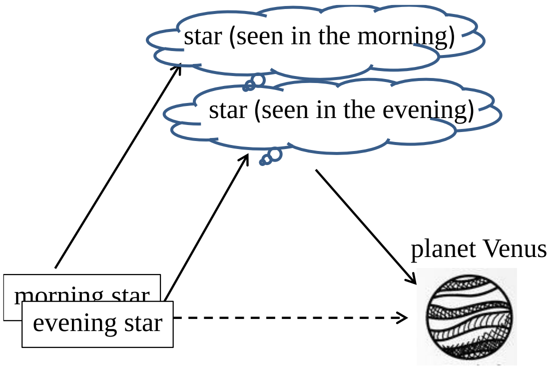


Fig. 2.6: The planet Venus, referred to in different contexts

Semantic approaches like thesauri and ontologies help to solve these linguistic transformation processes, by reducing the many possible representations of symbols to real objects optimally to one (Pickert 2011). In knowledge engineering these transformation processes are identified as encoding (*thing->concept->word*) and decoding (*word->concept->thing*).

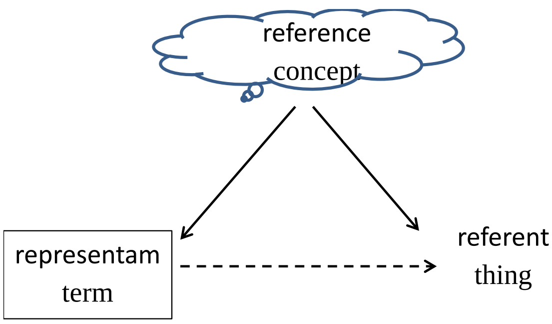


Fig. 2.7: The meaning triangle for thesauri

The philosophical groundings for thesauri are concept-oriented and deal with relations between concepts on one hand and between concepts and terms on the other hand. A concept in a thesaurus is more than merely a word, but it is a 'unit of thought', which can have several terms (labels), i.e. different words that are associated with it. Thus, for thesauri, the meaning triangle is characterized by following vertices depicted in Fig. 2.7.

2.3.2 Thesaurus tags as used in ISO 25964-1

The international norm for thesauri (ISO 25964-1) provides recommendations for the development and maintenance of thesauri intended for information retrieval applications. It is used in the following as backbone to define linguistic problems and the related elements and characteristics in thesauri. Tab. 2.1 gives, in parts, insights in the syntax used for thesauri according to the norm (for more details consult ISO 25964-1).

Tab. 2.1: ISO 25964-1 Tags

Type	ISO – Tag	Meaning
Reference	USE	Use; the preferred term that should be used in place of the non-preferred term
Equivalence relation	UF	Use for or Used for; alternative term (non-preferred term)
Root term	TT	Top term; the broadest concept in a hierarchy to which a specific concept belongs
Hierarchical relation	BT	Broader term; a concept having a wider meaning
Hierarchical relation (generic)	BTG	Broader term (generic)
Hierarchical relation (partitive)	BTP	Broader term (partitive)
Hierarchical relation (instantial)	BTI	Broader term (instantial)
Hiearchical Relation	NT	Narrower term; a concept with a more specific meaning
Hiearchical relation (generic)	NTG	Narrower term (generic)
Hierarchical relation (partitive)	NTP	Narrower term (partitive)
Hierarchical relation (instantial)	NTI	Narrower term (instantial)
Associatitve relation	RT	Related term; associated term, but not synonym, broader term or narrower term

2.3.3 Homonyms

A special problem for thesauri is the ambiguity of terms. In contrast to spoken language a thesaurus does not allow any vagueness in the use of terms and thus must resolve any ambiguities.

Co-funded by the
European Union

Homonyms are words which sound alike (homophones) or are spelled alike (homograph) but have different meanings[9]. For thesauri only homographs are of relevance. The example above, *Bank* is a typical homograph. A special case of homographs are polysemes. These are words with the same spelling and distinct but related meanings. Homonyms have different meanings and unrelated origins, whereas polysemes are usually considered to have multiple meanings (words such as *wood*, meaning either substance, or area covered with trees). A thesaurus has to solve disambiguation related to homographs and polysemes with qualifiers. Fig. 2.8 gives an overview of the different linguistic concepts.

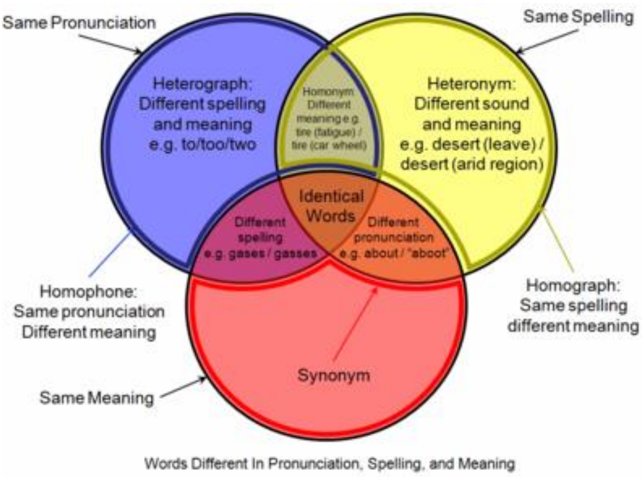


Fig. 2.8: Venn diagram showing the relationships between homographs (yellow) and related linguistic concepts[10]

2.3.4 Semantic relations

In any language two kinds of relationships are encountered between terms:

- Syntagmatic relationships: relationships that exist only because the concepts occur together in the context of a particular document. Ontologies, but not thesauri, are suitable to deal with this type of relationships.
- Paradigmatic relationships: relationships that are valid in almost all contexts and are intrinsic in their meaning. Thesauri are suitable to deal with this sort of relationships.

The difference between these relationships is made clear in Fig. 2.9.

---

[9] A more restrictive definition sees homonyms as words that are simultaneously homographs (https://en.wikipedia.org/wiki/Homonym)

[10] https://en.wikipedia.org/wiki/Homograph, retrieved Ma 7th, 2018

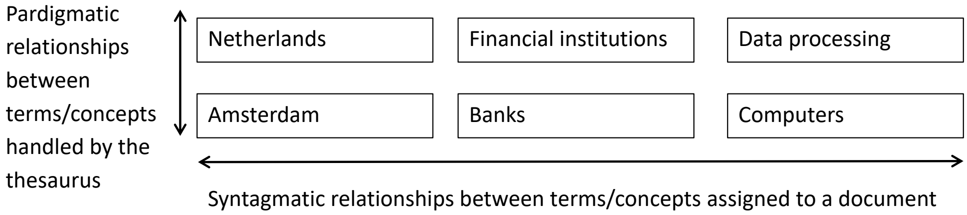


Fig. 2.9: Paradygmatic and syntagmatic relationships (ISO 25964-1)

Three types of paradigmatic relationship can be distinguished: (a) the equivalence relationship (= synonyms), (b) the hierachical relationship (= hierarchies), and (c) the associative relationship (= associations).

2.3.4.1 Synonyms

Synonyms are equivalent terms for the same concept. This is about term to term relationship which leads to the distinction of *preferred term* and *alternative terms*.

greenhouses
UF glasshouses -> alternative term
glasshouses
USE greenhouses -> preferred term

Fig. 2.10: Preferred versus alternative term

SP Spelling variant
MS MisSpelling
AB Abbreviation
FT Full form of the Term

Fig. 2.11: Tags for special types of synonyms

There exist different types of synonyms (compare ISO 25964-1; 8.2):

a) Terms of different linguistic origin (e.g. forest; wood)
b) Popular names and scientific names (e.g. baking soda; sodium bicarbonate)
c) Common names and trade names (e.g. chocolate creme; Nutella)
d) Variant names for emergent concepts (e.g. laptop computers; notebook computers)
e) Current terms versus deprecated terms (e.g. developing countries; underdeveloped countries)
f) Variant spellings (including stem variants), inverted word order and irregular plurals (e.g. groundwater; ground-water; ground water)
g) Terms originating from different cultures sharing a common languagd (e.g. flats, apartments)
h) Abbreviations or acronyms and full names (e.g. FAO; Food and Agriculture Organization)
i) Common nouns and slang terms (e.g. soluble coffee; instant coffee)

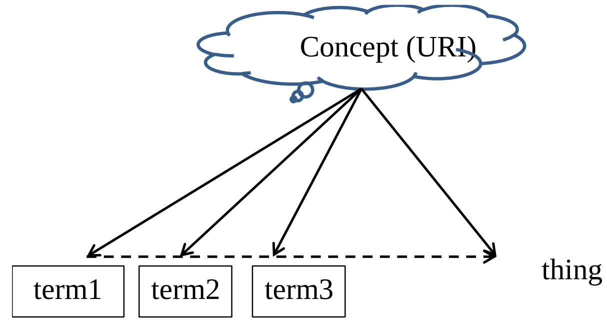


Fig. 2.12: Synonyms

In multilingual thesauri terms in different languages can be seen as synonyms in the broader sense.

2.3.4.2 Hierarchies

Hierarchies are relations between concepts when the scope of one of them falls completely within the scope of the other. The superordinated concept is defined as the broader term and the subordinated concept is defined as the narrower term.

animals
NT mammals -> narrower term
mammals
BT animals -> broader term

Fig. 2.13: Narrower versus broader term

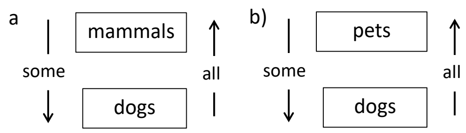


Fig. 2.14: "all-and- some-test": a) correct b) incorrect (some/some)

Three hierarchical relationships can be distinguished:

a) The generic relationship: this is the link between a class and its members. The ‘all and some test’ is used to check if a generic relationship between two concepts applies (used tags: BTG and NTG). The example b) in Fig. 2.14 does not satisfy this test because only some of the dogs are pets. In specific contexts, such as a pet hierarchy, this might however be considered generic, but it recommended to avoid these relationships if the thesaurus should stay interoperable with other semantic resources.
b) The whole-part relationship (also know as meronomy or partonomy) applies in situations where a part of an entity or system belongs uniquely to a particular possessing whole (used tags: BTP and NTP). The ISO 25964-1 limits this to four main classes of terms.
    - systems and organs of the body
    - geographical locations
    - disciplines or fields of discourse (see Fig. 2.15)
    - hierarchical social structures

Co-funded by the
European Union

Other cases can be excluded because the parts could belong to more than one whole, e.g. wheels coud be part not only of bicycles but also of other vehicles.

c) The instance relationship links a general concept, such as a class of things and an individual instance of that class, which is often represented by a proper name (used tags: BTI and NTI). In the example of Fig. 2.16 *Alps* and *Himalayas* are neither kinds nor parts of *mountain regions*, but represent individual instances.

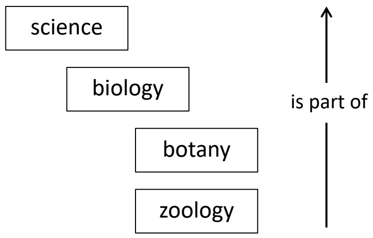


Fig. 2.15: Whole-part relationship

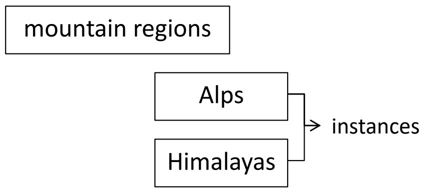


Fig. 2.16: Instance relationship

All three types of hierarchical relationships can be used together in one thesaurus.

2.3.4.3 Association

Associations are relationships between concepts that are semantically or conceptually associated to such an extent that this has to be made explicit in the thesaurus. It is recommended to use an associative relationship between concepts that overlap in scope, but not if they are siblings, because then they already share a common broader term.

birds
RT ornithology -> related term
ornithology
RT birds -> related term term

Fig. 2.17: related term

It makes sense to associate two terms if one is strongly implied by another. Here are typical examples:

a) A field of study and the objects studied (e.g. forestry/forests9
b) A process and its agents (e.g. temperature control/ thermostats)
c) An action and the product of the action (e.g. ploughing/furrows)
d) An action and its target (e.g. harvesting/crops)
e) Objects or materials and their defining properties (e.g. poisons/toxicity)
f) An artefact and its parts except those included in the meronomy (e.g. optical instruments/lenses)
g) Concepts linked by causal dependence (e.g. diseases/pathogens)
h) An object and its counter agent (e.g. plants/herbicides)

i) A concept and its unit of measurement (e.g. electric current/amperes)
j) A compound term and the noun that is its focus (e.g. model ships/ships)
k) An organism or substance derived from another (e.g. mules/donkeys)

2.3.4.4 Polyhierarchies/Monohierarchies

The hierarchical structure can have two different forms.

In the case of monohierarchies concepts belong only to a single broader class. However, there are concepts that can belong to more than one class at the same time, so a hierachical link should be created for all appropriate broader concepts. This is called polyhierarchy.

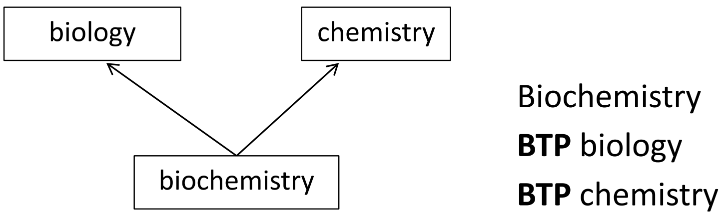


Fig. 2.18: Polyhierarchy

It is possible to have polyhierarchies based upon generic relationship to two broader terms, or based upon whole-part relationships or based upon both of the two hierchical types.

2.4 Semantic Standards

2.4.1 Semantic Web

The World Wide Web has become a giant accumulation of information in all languages and domains due to its decentralised structure and the enormous number of content producers. This indisputable success leads on the other hand to its most fundamental problem. The heterogeneity of the data formats and codifications in use hinders the interoperability between the different systems. Real information integration for a specific domain is not comprehensively achievable. Likewise problematic is the extraction of implicit knowledge from the available information. Online search engines like Google provide great support for research tasks but to exploit the full potential of the web semantic solutions are needed.

The Semantic Web is an extension of the World Wide Web through standards by the World Wide Web Consortium (W3C). The standards promote common data formats and exchange protocols on the Web[11]. The Semantic Web refers to the W3C's vision of the Web of linked data[12]. It provides a common framework that

[11] https://en.wikipedia.org/wiki/Semantic_Web, retrieved Ma 7th, 2018
[12] https://www.w3.org/standards/semanticweb/, retrieved Ma 7th, 2018

allows data to be shared and reused across application, enterprise, and community boundaries. Furthermore it should help to retrieve relevant information, to link information from diverse sources and to create implicit knowledge. The groundings of Semantic Web technologies are semantic standard formats and formal logic.

2.4.2 RDF

RDF is the standard model for the Semantic Web. It provides the foundation for publishing and linking data. It extends the linking structure of the Web forming a directed, labelled graph. The core structure is a set of triples, each consisting of a subject, a predicate and an object. In an RDF graph each triple is represented as a node-arc-node link.

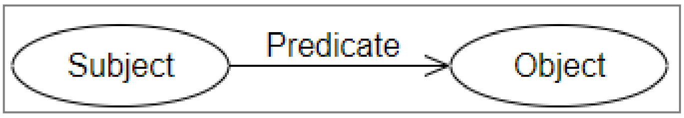


Fig. 2.19: An RDF graph with two nodes (Subject and Object) and a triple connecting them (Predicate)

While XML works with hierarchical tree structures, which are difficult to connect to each other, RDF offers an approach which provides more possibilities for interoperability and the usage of decentralised data structures. The object of the triple can act as a subject in other triples which allows to extent the graph nearly infinitely (see Fig. 2.20).

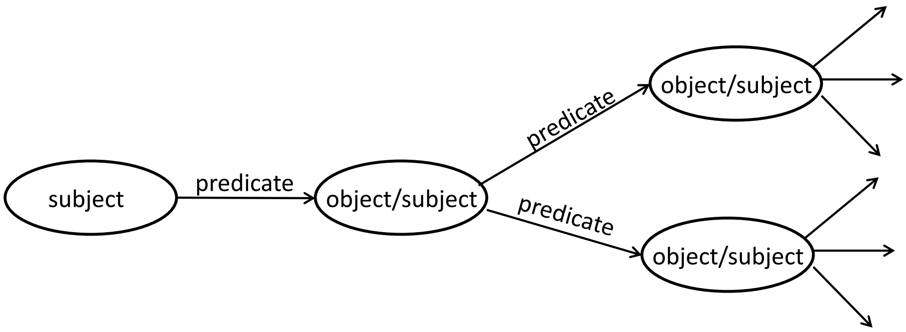


Fig. 2.20: subject and objects in RDF-triples

Linking content is only possible if problems of homonymy and equivalences are solved. For each resource of the RDF triple (for subject, object as well as for predicate) URIs (Uniform Resource Indentifier) are used to uniquely identify which supports at the same time disambiguation. Related to the semiotic triangle URIs are the symbol (*rapresentam*) of a resource (see Fig. 2.21).

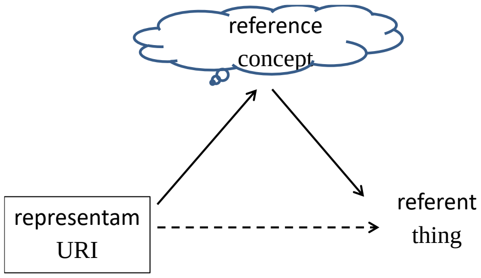


Fig. 2.21: RDF-resources and the semiotic triangle

2.4.3 OWL

RDF is the fundament of the Semantic Web, but is limited in its application areas and the extraction of implicit knowledge. Complex information needs more accurate representation languages such as OWL (Web Ontology Language). OWL is built upon RDF, but extends its vocabulary. The OWL languages are characterized by formal semantics including axioms and inferences. OWL is used for authoring ontologies. The W3C-endorsed OWL specification includes the definition of three variants of OWL, with different levels of expressiveness. These are OWL Lite, OWL DL and OWL Full (ordered by increasing expressiveness).

2.4.4 Linked Open Data

Linked Data is about using the Web to connect related data that wasn't previously linked, or using the Web to lower the barriers to linking data currently linked using other methods. More specifically, Wikipedia defines Linked Data as "a term used to describe a recommended best practice for exposing, sharing, and connecting pieces of data, information, and knowledge on the Semantic Web using URIs and RDF."[13] Tim Berners-Lee coined the term and specified four principles:

a) Use URIs to name things
b) Use HTTP URIs so that these things can be dereferenced
c) Provide useful information about what a name identifies when it's looked up, using open standards such as RDF, SPARQL, etc.
d) Refer to other things using their HTTP URI-based names when publishing data on the Web.

Linked open data is linked data that is open data. Large linked open data sets include DBpedia and Wikidata. The Linked Open Data Project is a W3C community project with the goal to extend the Web with a data commons by publishing various open data sets as RDF on the Web and by setting links between dats items from different sources. It makes use of the properties of the Semantic Web, which are:

- With RDF technology it is possible for everybody to publish its own data.
- The Web of Data is self-descriptive

[13] http://linkeddata.org/, retrieved May 7th, 2018

- Applications can follow the data graph to find new information sources

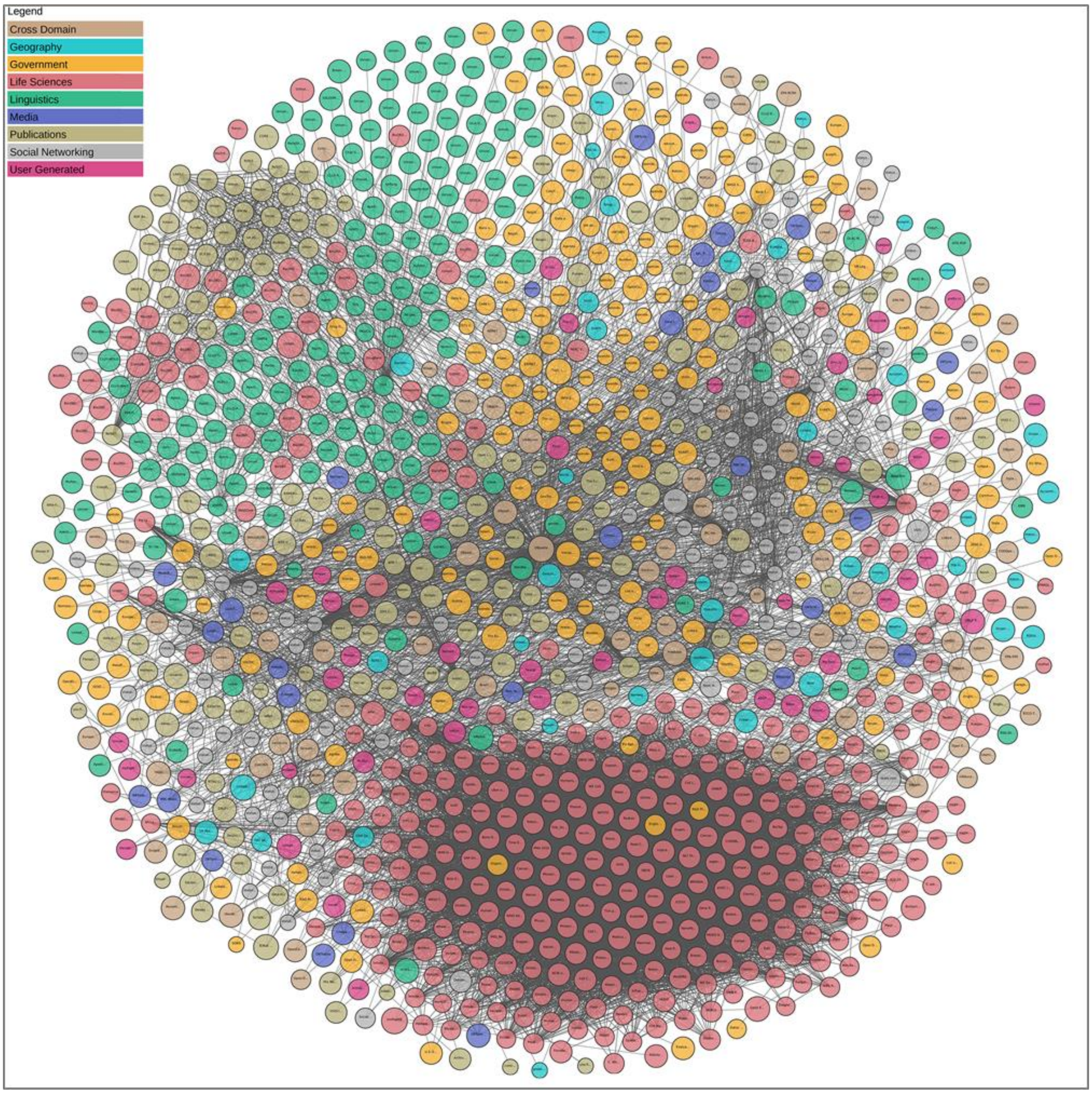


Fig. 2.22: Linked Open Data Cloud[14]

With this approach more than thousand data repositories have been established within the Linked Open Data Cloud, and more than 50% of these datasets link to other resources. More than 140 repositories are SKOS-based thesauri. It allows anyone to expose their datasets in a big collection of external resources. This gives the opportunity to enrich its own data with external linked content and to query across the cloud individually, far beyond traditional query engines.

[14] http://lod-cloud.net, retrieved May 17th, 2018

2.4.5 SKOS

ISO 25964-1 lists four exchange standard formats for thesauri: MARC (Machine-Readable-Cataloguing for bibliographic information), Zthes (XML based), DD 8723-5 (from the British Standards Institution) and SKOS (Simple Knowledge Organization System).

The latter was endorsed as a W3C recommendation[15] in 2009 for sharing and linking knowledge organization systems in the semantic web. In contrast to the other three mentioned standards, SKOS was developed for the specific aim to serve as a thesaurus standard. In accordance to ISO 25964-1 it is concept-based, which means that is built upon concepts and their representing terms.

SKOS is based on RDF, which is one of the fundamental Semantic Web specifications. RDF uses graphs as its data structure. This means that each concept is a node and edges connect these nodes to create a graph. But SKOS is not an ontology in the classical sense, which are descriptions of parts of the real world based on axioms and facts. It rather serves as a guideline for a specific domain.

SKOS concepts are identified with unique URIs and described by natural language labels. These are linked via relations to other concepts enabling a hierarchical and associative organization of them within the system. In addition to associations within a thesaurus it is possible to map concepts also across vocabularies. This SKOS functionality helps to semantically enrich each concept and thus also the whole vocabulary.

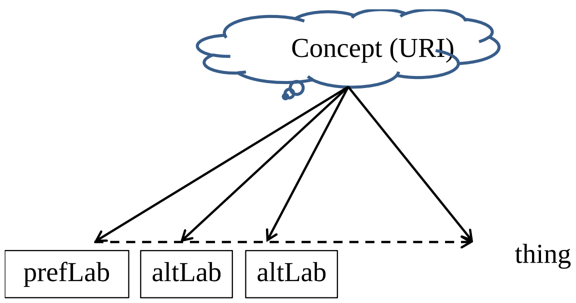


Fig. 2.23 SKOS concepts in the semiotic triangle

The SKOS vocabulary is a set of URI, which is described in detail in the SKOS reference. Tab. 2.2 lists a set of SKOS URIs with their characteristics.

[15] https://www.w3.org/TR/skos-reference/

Tab. 2.2: Overview of SKOS URIs

Type	Description	URI	Thesaurus-syntax[16]
Concept Class	A SKOS concept is described as a unit of thought identified by a URI and is the fundamental element in a thesaurus.	skos:Concept	
Concept Schemes	A Concept scheme allows to keep together a number of concepts that are thematically related. It corresponds to the notion of an individual thesaurus or classification scheme. It is recommended that no two concepts have the same preferred lexical label in a given language when they belong to the same concept scheme.	skos:ConceptScheme skos:inScheme skos:hasTopConcept skos:topConceptOf	 TT
Lexical Labels	The lexical labels correspond to the equivalence relationships described in chapter 2.3.4.1. A label is a literal with a language tag (e.g. 'en' for English). With skos:prefLabel a preferred lexical label to a resource is assigned. Skos:altLabel makes it possible to assign an alternative label. Hidden labels may be used to include misspelled variants for text-based indexing and search operations. For each concept only one prefLabel per language tag is allowed.	skos:altLabel skos:hiddenLabel skos:prefLabel	UF MS USE
Notations	Notations are symbols which are natural-language independent used as the primary means to access the concepts like in the Agrovoc-Thesaurus.	skos:notation	
Documentation properties	Documentation properties are human-readable (informal) documentation. Skos:note is a property for general documentation purposes. All the others are specializations e.g. skos:skopeNote is used to indicate the use of a concept.	skos:note skos:definition skos:editorialNote skos:example skos:historyNote skos:changeNote skos:scopeNote	
Semantic relations	Semantic relations are used to define hierarchical and assiciative relations (skos:related) as described in chapter 2.3.4.2. skos:broader and skos:narrower are each others inverse and are used to specify the relationship between one genre and its specific species. There can be more than one broader relationship per concept (polyhierarchies).	skos:broader skos:narrower skos:related	BT NT RT
Concept collections	Collections give the opportunity to gather similar concepts in one group. It is not possible to relate concept collections to each other, this has to be defined on concept level.	skos:Collection skos:member skos:memberList	
Mapping relations	These relations are used to map across vocabularies to enrich semantically each of them. The most popular relationship is skos:exactMatch.	skos:broadMatch skos:closeMatch skos:exactMatch skos:relatedMatch	

[16] As used in ISO 25964-1

2.4.6 SKOS-XL

SKOS-XL is an optional extension of the SKOS standards. The SKOS-XL vocabulary includes following set of URIs: *skosxl: label, skosxl:literalForm, skosxl:prefLabel, skosxl:altLabel, skosxl:hiddenLabel, skosxl:labelRelation*

An instance of the class skosxl:Label is a resource and may be named with a URI. An instance of the class skosxl:Label has a single literal form[17].

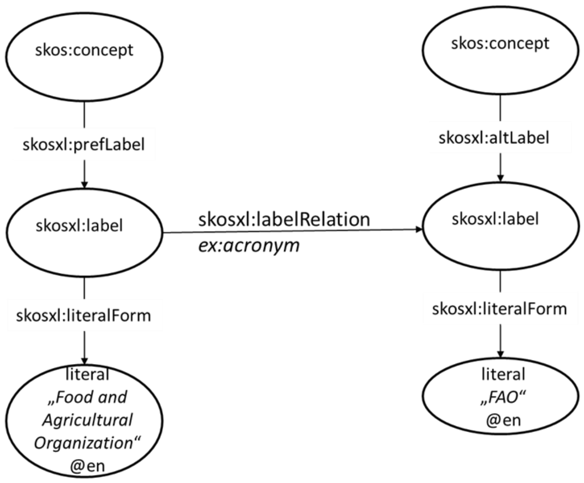


Fig. 2.24 Specialisations of relations between labels in SKOS-XL

This allows to define specialisations of relations between two labels (acronym of Food and Agricultural Organization is FAO) by defining a sub-property of skos:labelRelation (e.g. ex:acronym) as shown in Fig. 2.24, which corresponds to the AB tag as used in ISO 25964-1. Only with the SKOS vocabulary defining an abbreviation of a specific label would not be possible. Another use case would be to link two multilingual thesauri, where a German *prefLabel* of a concept corresponds to an English *altLabel*. The drawback of this approach is that it might lead to incompatibilities with other vocabularies not offering these extensions.

2.4.7 SPARQL

SPARQL (SPARQL Protocol and RDF Query Language) is an RDF query language. It is a semantic query language for databases, able to retrieve and manipulate data stored in Resource Description Framework (RDF) format. It was made a standard by the RDF Data Access Working Group (DAWG) of the World Wide Web Consortium, and is recognized as one of the key technologies of the semantic web.[18]

A SPARQL endpoint is a conformant SPARQL protocol service as defined in the SPROT (SPARQL protocol for RDF) specification. It enables users (human or other) to query a knowledge base via the SPARQL language.

[17] http://www.w3.org/TR/skos-reference/#ref-SKOS-PRIMER
[18] http://www.w3.org/TR/rdf-sparql-query/

Results are typically returned in a pre-selected machine-processable format (e.g. XML, HTML, Simple JSON). Therefore, a SPARQL endpoint is mostly conceived as a machine-friendly interface towards a knowledge base and it is recommended that the formulation of the queries should be implemented by the calling software.[19]

```
select distinct ?concept (str(?prefLab) as ?label) (str(?def) as ?definition) (str(?parent) as ?parent)
where {?concept <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
<http://www.w3.org/2004/02/skos/core#Concept>.
?concept <http://www.w3.org/2004/02/skos/core#prefLabel> ?prefLab.
?concept <http://www.w3.org/2004/02/skos/core#definition> ?def.
?concept <http://www.w3.org/2004/02/skos/core#parent> ?parent.}
```

Fig. 2.25 SPARQL example query

Fig. 2.25 shows the syntax of SPARQL with an example query formulating the question "*give me identifier, preferred label and definition of all concepts, that have as well a preferred label as a definition*".

2.5 EnvThes – the Environmental Thesaurus

In order to share information across the different components of the network common standards on the metadata and the underlying semantics are needed. Networks like ILTER is characterised by a high inter- but also intra-network heterogeneity (Vanderbilt et al. 2015). This is also true e.g. for protected areas which also can be seen as loosely organised network of data collectors in terms of data management. In order to overcome these barriers of data integration the adoption of commonly used standards and semantics is one of the options. This led e.g. to the recommendation to adopt EML as common metadata language and EnvThes as common semantic backbone (Vanderbilt et al. 2010).

In order to provide a stable semantic backbone the ***Environmental Thesaurus*** (EnvThes, see http://vocabs.ceh.ac.uk/evn/tbl/EnvThes.evn) was developed to provide terms annotating data resulting from long term ecosystem research and monitoring (Schentz et al. 2013). EnvThes is built on US LTER Controlled Vocabulary (Porter 2010) extending it with links to other vocabularies of domain interest like EUROVOC, GEMET to name just a few.

The vocabulary is open and based on current semantic web standards (SKOS and SPARQL) and supports multilinguality. EnvThes is the core vocabulary used in the DEIMS-SDR to annotate keywords, research topics, observed parameters and various other metadata elements. In addition first tests for the use of EnvThes as multilingual thesaurus for annotation and discovery were made (Vanderbilt et al. 2010, Vanderbilt et al. 2017).

Within the ECOPOTENTIAL project a revision of the current EnvThes version 1 was done and an extension of the concepts with regard to the integration of remote sensing variables was developed. This ensures the wider applicability of EnvThes.

[19] http://semanticweb.org/wiki/SPARQL_endpoint.html

2.5.1 Scope and requirements

The integration and (also semantic) interpretation of data coming from different resources is an important step in the collection and aggregation of data to address large scale scientific questions. Especially in modelling approaches data from a wide range of data providers need to be integrated in order to cover either the thematic, spatial or temporal gradients needed for the analysis. This is often hampered by the use of different terminologies and references when describing the data but also when providing information on the parameters (or measures) observed. This can be e.g. due to different languages used or different scientific approaches or domains.

Using a common language describing the contents – a common semantic – is needed to overcome these semantic barriers. This can happen either implicitly by the interpretation of the data analyst or explicitly by the use of a common controlled vocabulary or shared reference lists. Fig. 2.26 shows the schematic links between the data and common semantic. These links can either be made by simple links or by the use of annotations.

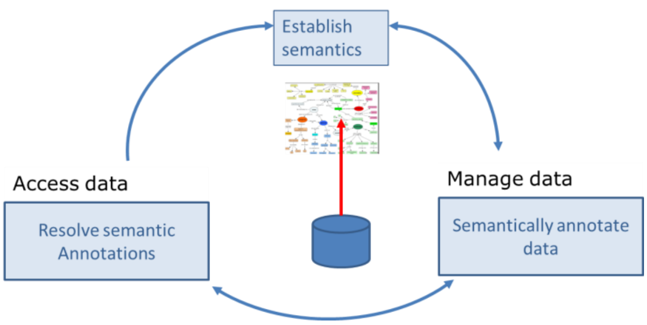


Fig. 2.26: Conceptual schema on the use of controlled vocabulary for semantic annotation

Within the current project EnvThes was used as common semantic basis. This led three main usages and the underlying workflows which are described in more detail in this chapter.

Tab. 2.3: Use cases for EnvThes exploitation

UC 1	Use the controlled vocabulary as source for keywords used in the process of data documentation (metadata) in order to describe the thematic, spatial and temporal content of the datasets, data services, and research sites
UC 2	Use the controlled vocabulary and shared reference lists to specify and define the terms used in reference lists for a given parameter in the data (e.g. infrastructure type installed at a research site)
UC 3	Use the controlled vocabulary to annotate of data files and their contents for later analysis

These basic use cases drive the development, structure and content of the controlled vocabulary and lead to the definition of the basic development principles for EnvThes.

2.5.1.1 Metadata keywords

A keyword is generally defined as a term or phrase that is a topic of significance. In the field of digital information management keywords are used to classify or organise digital content or facilitate online search for information. The user uses these terms to select relevant data items from a collection of available information. A keyword section is mostly part of metadata specifications (e.g. ISO19115/19139 or EML) not only defining the scope of the keyword section but also the source of the terms. Therefore, to allow for systematic keyword tagging, controlled vocabularies are used to avoid diverging naming of the same things, diverging spellings of the same term or typographical errors. In this sense controlled vocabularies and thesauri, e.g. like EnvThes, are used to provide “select lists” the user can choose options from.

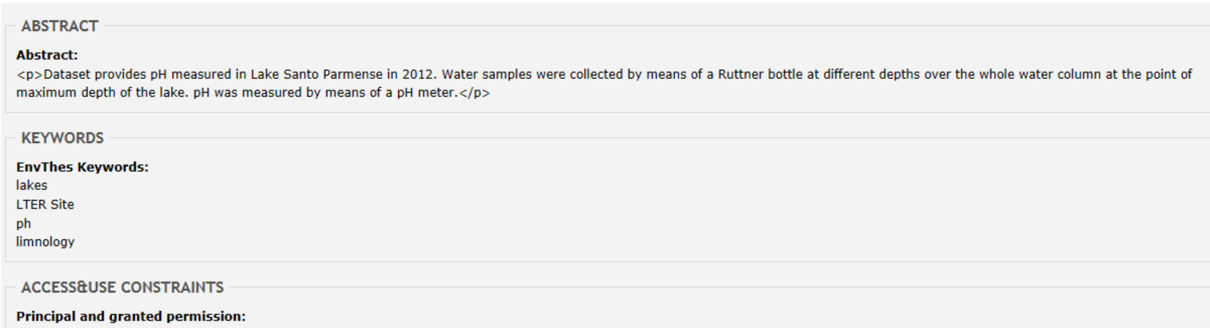


Fig. 2.27: Example of the keyword section in DEIMS-SDR

This keyword tagging could either be done by the data provider directly in the workflow of providing meta information. Other options are to use annotation services providing a user support for the selection of the appropriate keywords. Currently EnvThes serves as source for the keywords of metadata managed by the metadata platform DEIMS-SDR, thus allowing a cross dataset search.

2.5.1.2 Data annotation

Exactly defined controlled vocabularies for the exchange of scientific information, as e.g. the table of elements, SI system of units, works of reference for anatomy, physiology, pathology, species lists, lists of soil types and many others have been established long before the first computer was built. Especially in the domain of libraries the concept of controlled vocabularies was used.

Nowadays controlled vocabularies are not only used within science, they are already state of the art in our everyday procedures seamlessly implemented in a number of workflows, e.g. spellcheckers of word processing software and other office software, the autocomplete functions in SMS- Editors or for discovery fields.

This also reflects the need to harmonise the description and naming of data elements, like e.g. reference list or vocabularies used. When trying to combine datasets from different sources harmonised elements (e.g. parameter naming) are needed. So for the field of data and information management for the long term observation and experiment communities this can be defined as basic use case. This use case is defined by using controlled terms in data documents. This can either be done by looking up a concept in a vocabulary and transcribing it manually or using a copy/ paste procedure for the preferred label, entering the term into a document, spreadsheet, database or what so ever.

This procedure involves a certain level of uncertainty, as the person, reading the document, evaluating the spreadsheet or data of the database, not necessarily knows where the term was taken from. If the resource, e.g. the controlled term, is not available online, information like definition or linked terms cannot be checked. To avoid a certain level of uncertainty the source controlled vocabulary, e.g. species taxonomy or reference vocabulary used, needs to be recorded in addition to the term itself. If this is not provided, most of the ambiguity of the term remains to the user and normally cannot be solved easily.

2.5.1.3 Machine readable references

In addition to the need of unambiguous terms, it is also necessary to have the possibility to lookup definitions, translations and relations to other terms, when working with data – e.g. data discovery, data access and analysis. When approaching the details of the dataset and its elements the user "drills into" the data element retrieving not only the data values itself but also related meta information, as e.g. the meaning of terms.

This can of course be done by manually looking up those definitions in controlled vocabularies, but an easier, more exact way is the establishment of links from the data, metadata, documents to the concepts within the vocabulary. To enable such links the use of identifiers for the terms is indispensable. Using SKOS/RDF and a Linked Data interface, where the identifier is a URL, this URL can be entered into the document/ spreadsheet, database and the user can follow the hyperlink to the concept within the thesauri. Normally, tools like pdf-editors and readers, Microsoft Office Tools, openOffice tools, WIKI tools, content management systems (CMS) allow for following these links in the same way as provided for a hyperlink. Thus the related concept or term can be accessed.

As this linked information is a machine readable way, it also allows software to follow the link, e.g. by using SPARQL queries. If the data of the spreadsheet or database is exposed via a SPARQL endpoint, a joined SPARQL query can be done combining e.g. the data of the database with definitions or translations for the concept in the thesaurus.

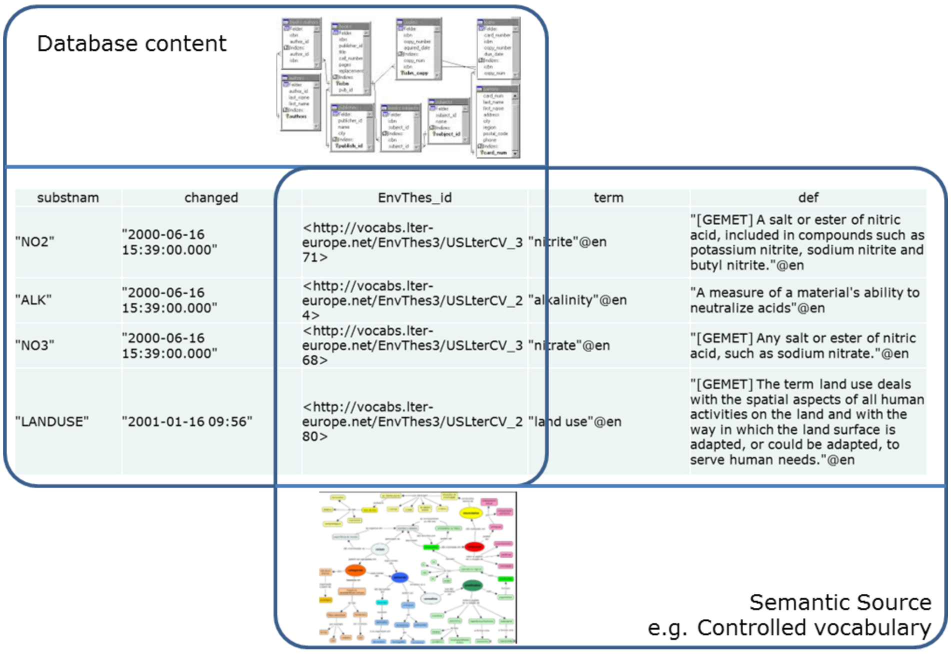


Fig. 2.28 Linking the database content with definitions coming from a controlled vocabulary

If there are multiple SPARQL endpoints, exposing clear references (URLs) to the controlled vocabulary these references can be used to integrate the data from these different resources.

2.5.2 EnvThes Content

2.5.2.1 Concept schemes

The EnvThes vocabulary includes two concept schemes, (a) EnvThes, and (b) Reference lists.

The concept schema ***EnvThes*** is the main corpus of the thesaurus containing the concepts and strcutre created by an editor team consisting of ECOPOTENTIAL and LTER exeperts. These concepts are considered to be central source for the use as keywords and metadata annotation.

The concept schema ***Reference Lists*** is a container for relevant controlled lists often used, e.g. in DEIMS-SDR, which are not elsewhere available on the web as referable objects. This includes e.g. habitat lists (like the EUNIS habitat classification[20]) or map legend entries or lists used in documents (like CORINE land cover[21]). By converting them into SKOS concepts (with stable URIs) they can be reused also from external users.

[20] https://www.eea.europa.eu/data-and-maps/data/eunis-habitat-classification

[21] http://www.umweltbundesamt.at/fileadmin/site/umweltthemen/raumplanung/1_flaechennutzung/corine/CORINE_Nomenklatur.pdf

2.5.2.2 EnvThes top concepts

Research focus: is defined as the main field of scientific interest. This addresses not only the research topics as the thematic focus, but also the scale as the spatial and temporal focus.

Infrastructure: is defined as the basic physical and organizational structures and facilities (e.g. buildings, roads, power supplies and devices) that are needed for the operation of the observation facilities (e.g. sites or protected areas). This top concept encompasses all terms needed to describe the phases from the collection, management and provision of data.

Method: A way of proceeding or doing something, especially a systematic or regular one.

Parameter: A parameter is a compound concept, with the preferred label composed by the <*property observed*> and the <*object of interest*>, e.g. 'diameter tree stem'. Parameter concepts can have multiple parent concepts; e.g. concentration of carbon dioxide can be at the same time used as an ecosystem, a soil or a water parameter.

Object of Interest: This group of concepts comprises events, processes and entities which represent the object of interest that is being observed or measured. Entities embrace substances, organisms, organizational units and matrix.

Property: The property of the object of interest that is observed during the act of observation. The concepts of properties are elements of "Measurements on properties of objects of interest" expression. Thus, one property can be associated with several measures on several objects of interest.

Constraint: A constraint controls another concept by keeping it within a particular domain or limit. Constraints can be used in the context of properties and object of interest which help to keep these lists slim and manageable. Grammatically they correspond to adjectives characterizing nouns and thus specialising concepts (e.g. *organic* for specializing matter)

Deprecated concept: is defined as a concept which is no longer in use. Concepts which are not used anymore or are replaced by another updated terms are moved to deprecated and maintain their link to the following term. By this all concepts are accessible throughout the lifetime of the controlled vocabulary.

Except for parameters polyhierarchies are avoided in EnvThes to provide a clear and explicit structure.

2.5.3 EnvThes Governance

2.5.3.1 EnvThes development

EnvThes has been established building on existing vocabularies of the ecological community and extending them where considered relevant for the domain of long term ecosystem observation. This vocabulary work started with the Life+ project ENVeurope (2010-2014) and was continued in ExpeER (2010-2015). The US-

LTER vocabulary[22] developed by J. Porter (2011) served as backbone vocabulary for EnvThes. It was extended constantly to accommodate specific needs of the above mentioned projects (specifically in the definition of parameters and methods). Also concepts of the ANAEE community were included. Numerous mapping relations were defined towards foundational vocabularies such as GEMET, the official thesaurus of the EEA, EUROVOC, the thesaurus of the European Commission and AGROVOC, the vocabulary of the FAO with translations in multiple languages.

The current focus of the advanced development is to align EnvThes to **basic elements of O&M** (ISO 19156) in order to make use of the controlled vocabulary together with data structured according to O&M easier. Another objective is to guarantee **semantic interoperability** with similar parameter descriptions from related initiatives operating semantic systems and databases (such as PANGAEA, SeaDataNet, AnaEE). Instead of incorporating external SKOS vocabularies in EnvThes, the concepts should refer to the matching terms by referring to the URI link.

The present most important aim is, however, to prepare the **next version release (2)**, which should be free from structural inconsistencies and syntactic errors. This comprises also an enhancement in the clarity of the choice of the top concepts and the reduction of complexity in the hierarchy. Also in terms of completeness regarding definitions and semantic mappings this version should show considerable improvements.

With a stable version 2 the focus can then be put in the provision of translations of the terms in multiple languages.

2.5.3.2 Design principles

Considering ISO 25964-1 as basic guideline for the design principles for the development of EnvThes the following decisions about syntax and structure were taken:

- The main language of EnvThes is English. Thus wherever a definition of a concept is provided in another language, the English definition should also be added
- Each label needs to have a language tag.
- Concept labels should be spelled in singular and written in small letters. Exceptions are organisms because in scientific literature the singular is never used. Likewise acronyms are written as generally used in capital letters.
- Per concept and language only one preferred label should be defined.
- Per concept only one definition is allowed as a vocabulary should provide an unambiguous reference
- The source of definition if not self-provided should be placed at the beginning of the definition in square brackets, e.g. definition: *[GEMET] The space occupied…* (see Fig. 2.31)
- Scope notes should be utilized to constrain the use of the concept (e.g. *DEIMS keywords*)
- Deprecated terms which are replaced should indicate the new term to be used instead via the *close match* property.

[22] http://vocab.lternet.edu/vocab/vocab/sobre.php

- It should not be attempted to implement complex relations because these would impede a proper interoperability with other semantic resources. Ontologies could store these additional relations instead.
- The completeness of definition of a concept is given when it has specified following properties:
    - A preferred label in English (and if possible in a second language)
    - At least one alternative label in English (and in a second language) where appropriate
    - A definition given in English indicating the source
    - Exact match relation defined if possible
    - A scope note to indicate the envisaged usage
    - The name of the person who entered the concept indicated as creator

2.5.3.3 Governance structure

EnvThes is a community effort which needs input from different members of the community. The organizational structure of EnvThes comprises following members:

System manager: is the manager responsible for the technical hosting of EnvThes, which is located at CEH. His main task is to guarantee a smooth performance of the TopBraid EVN and TopBraid Explorer servers. He establishes the accounts for editors in TopBraid.

Technical coordinator: has the task to support the facilitator in technical questions and to enable the technical interoperability between EnvThes and DEIMS-SDR.

Facilitator: supports and supervises the Editor Team in the development of the vocabulary in technical and procedural terms and is responsible for the structural and semantic accuracy. Additionally the facilitator establishes cooperation with other communities curating semantic resources of relevance for the LTER community.

Editor team: takes decisions on the content and structure of the thesaurus and is responsible for the scientific correctness of the vocabulary. The core team is composed by three stable members from the terrestrial domain and supported by three additional scientists from diverse domains (remote sensing, freshwater domain, landscape ecology) contributing from time to time. The core members meet more or less regularly in organized teleconferences and face to face workshops steered by the facilitator.

Collaboration Team: will play a bigger role once EnvThes has a stable version to which additional editors can contribute. The main contribution will be the translation of terms in different languages. The members will be recruited from scientists involved in ECOPOTENTIAL and eLTER.

2.5.3.4 EnvThes editor – TopBraid

During the lifetime of EnvThes several thesauri editors where tried out such as IQvoc[23], VocBench[24] and PoolParty[25]. The actual software in use is TopBraid Enterprise Vocabulary Net (EVN)[26] which has been recognized to be the most suitable editor for EnvThes.

TopBraid EVN is a flexible, web-based system for managing semantic information models. TopBraid EVN managers can publish vocabularies for exploration and use in the browsing environment (TopBraid Explorer). CEH is the host server for thesauri managed in TopBraid for different communities like LTER Europe.

EnvThes can be accessed openly by users without editing rights via http://vocabs.ceh.ac.uk/evn/tbl/EnvThes.evn.

Key features of TopBraid are the intuitive graphical user interface with auto-completion, drag and drop and richt text editing. The interface has three sections. The concept hierachy is displayed at the left side. In the middle is placed the main window with detailed description and properties of one selected concept. On the right side one can find a search interface with simple lookups and advanced search functions. Additional term search field is placed under the concept hierarchy.

[23] http://iqvoc.net/

[24] http://vocbench.uniroma2.it/

[25] https://www.poolparty.biz/

[26] https://www.topquadrant.com/products/topbraid-enterprise-vocabulary-net/

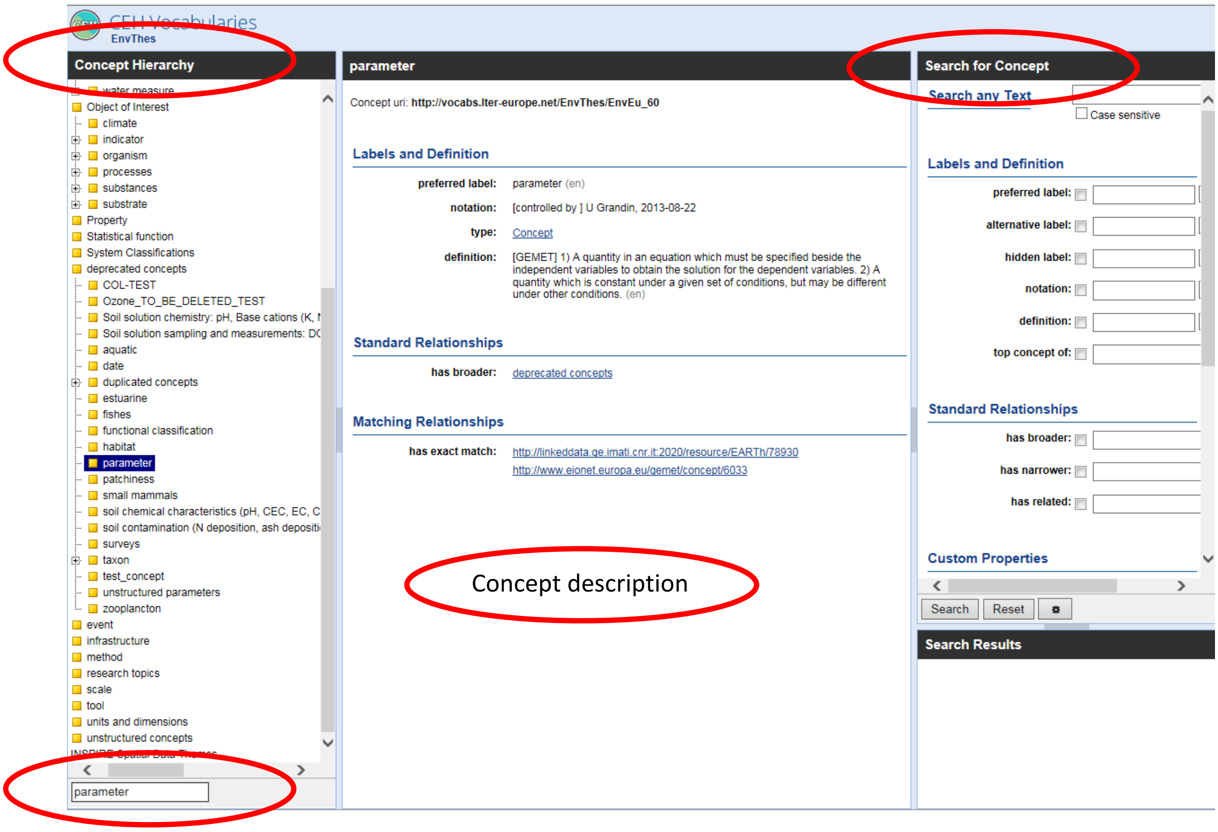


Fig. 2.29: Open access view of EnvThes in TopBraid

TopBraid has an inbuilt access control management with assigned roles. The managers can assign a password controlled account to editors. The URL used by editors is the following:

http://evn.ceh.ac.uk/

In TopBraid EVN every change is logged and time stamped, change history can be searched, changes can be rolled back if desired. TopBraid has import and export features for RDBMs, spreadsheets, XML, SPARQL endpoints, RDF and OWL. It supports W3C standards such as SKOS, SPARQL and OWL.

The user interface for editors is the same as for viewers plus some additional features at the bottom of the screen (see Fig. 2.30).

The first button on the right enables to create a new concept, the second a new concept scheme. The star button offers actions for different processes such as exporting the hierarchy to a spreadsheet and delete. In the central part of the interface the edit button enables to change all properties of a concept and at the very right it is possible to comment the concept and to see the whole history of comments.

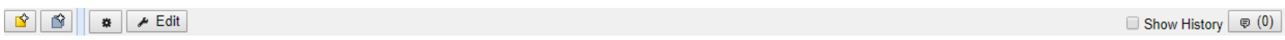


Fig. 2.30: Editor features at the bottom of the TopBraid user interface.

The main possible editing's can be summed up as follows:

- Create a new concept with its properties
- Change properties of an existing concept
- Delete an existing concept
- Modify the position of an existing concept in the hierarchy

The properties of a concept that can be edited comprise all SKOS elements as described in chapter 2.4.5. For each label also a language tag should be added. Per language only one preferred label should be defined, but more than one alternative labels are possible. Unfortunately TopBraid doesn't control this SKOS rule and it lies in the responsibility of the editor (or of the facilitator) to provide this information accordingly.

A concept should also possess a definition at least in English. While this is for sure a desirable aim it is a quite time-consuming work to provide proper definitions for all concepts within a thesaurus and can take years till its completion. Definition alignments as described in chapter 2.5.4.3 are a good approach to accomplish this task.

Where ever possible concepts should be enriched by pointers to other semantic resources using mapping relations. Because of an unsolved bug the link provided in match relation is not resolved to the destination URI, but points back to an internal page of the thesaurus. Version 6 of TopBraid EVN seems to be bugless in this respect and it is planned to port EnvThes to the newer version within 2018.

For specifying the status of a concept TopBraid EVN provides an additional (non SKOS based) element named *deprecated* which can be set true or false. Concepts should never be deleted, because they may be referenced by metadata or data. The end user is alerted by a red box highlighting the issue and should think to use another concept instead. If the editor deprecated the concept to be replaced by a new one in the same concept scheme, this should be made clear by linking to the new concept using the association *has close match.*

Concept uri: **http://vocabs.lter-europe.net/EnvThes/msa0347**

View as ttl rdf/xml

Labels and Definition

preferred label: ecological niche (en)

hidden label: ecological niches (en)

type: Concept

definition: [GEMET] 1) The space occupied by a species, which includes both the physical space as well as the functional role of the species. 2) Ecological niche refers to the characteristics of an environment that provides all the essential food and protection for the continued survival of a particular species of flora or fauna. In addition to food and shelter, there is no long-term threat to existence in that place from potential predators, parasites and competitors. The concept of the ecological niche goes a long way beyond the idea of the species habitat. (en)

Standard Relationships

has broader: ecosystem

has related: habitat preference
habitat quality
habitat suitability
habitat use

Matching Relationships

has exact match: http://www.eionet.europa.eu/gemet/concept/2456

Custom Properties

Created: 2016-08-03T00:00:00Z

Creator: barbara.magagna@umweltbundesamt.at

Modified: 2018-05-18T18:45:18.166+01:00

Notes

scope note: MS Academic Keywords (en)

Fig. 2.31: Concept description

abundance of annelida (Concept)

Concept uri: **http://vocabs.lter-europe.net/EnvThes/10078**

Last edited by doron.goldfarb on Jan 26, 2017 9:47:59 AM

THIS CONCEPT HAS BEEN DEPRECATED

Labels and Definition

preferred label: abundance of annelida (en)

type: Concept

definition: measured abundance of annelid species within a habitat or ecosystem. (en)

Status

deprecated: true

Standard relationships

has broader: deprecated concept

has related: abundance
annelids
species abundance

Custom Properties

Created: 2014-12-16T14:27:21Z

Creator: jww

Modified: 2016-11-25T12:51:43.864+00:00

Fig. 2.32: Deprecated concept

While extensions of SKOS are allowed these might be lost when transferred to other SKOS editors or when the thesaurus is displayed through semantic repository. For EnvThes it was decided that no more used concepts should additional be placed under the top concept *deprecated concept* which alleviates the above-mentioned problem.

The most beneficial aspect that TopBraid offers is the multiuser collaborative environment for vocabulary developers. Virtual work-in-progress copies of the vocabulary, so called working copies, allow parallel development of versions and enable controlled publishing, review and approval workflow.

Working copies are copies of the complete published thesaurus. They can be created by the manager which has often also a facilitation role within an editor team. In a working copy an editor makes changes, adds concepts, definitions or translations without affecting the thesaurus as it is seen by the public. The facilitator reviews all the changes by checking the change history log. Only after the acceptance by the facilitator the working copy is converted into a production copy which than can be published to be seen by external users (see Fig. 2.33).

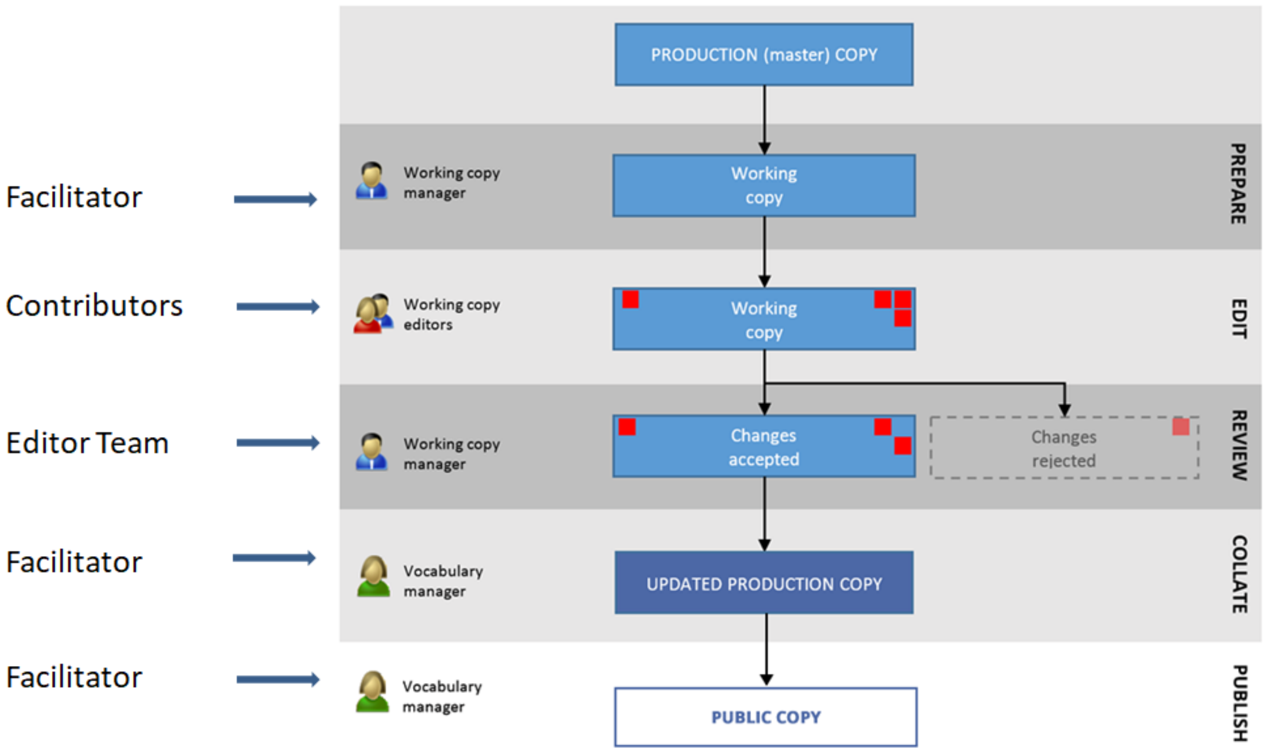


Fig. 2.33: Working copy management in TopBraid

Working copies help to organize contemporaneous contributions of different editors. However, to avoid affecting the work being done by others it is highly recommended to focus on independent branches of the vocabulary especially when the structure is being changed. The impact of restructuring the same hierarchy by more than one editor could create conflicting situations which are not easy to solve. Practical use cases for using working copies are translations in different languages which do not affect the hierarchy structure at all.

2.5.3.5 Adopted procedures outside TopBraid

Although TopBraid EVN provides advanced means for collaborative development, not all sorts of issues can be treated efficiently in the editor. The EnvThes SPARQL endpoint (**http://vocabs.ceh.ac.uk/evn/tbl/sparql**) was used to query specific triple lists which we then exported in excel sheets to further analyse them. Another useful method is to use the export function within the TopBraid editor to copy the hierarchy below a certain concept into an Excel sheet. This excel sheets are used to subdivide the work between editors and to provide customized overviews for specific top concepts. They are also very useful when combined with external sources. Excel sheets can also be used for voting upon different options. The drawback is on the other hand, that all needed updates according to the Excel sheet results have to be done manually. ThesauForm (Laporte et al 2012) has a voting/weighting system integrated in the editor to support the validation stage, a feature which is missing in TopBraid.

2.5.3.6 EnvThes versioning management

During its creation and first development stage a clear versioning practice was not applied for EnvThes as mainly the editor team and the facilitator were concerned with the management of the vocabulary. A versioning management will become more important once EnvThes is becoming a consolidated stable reference, ready to be extended by a bigger interested user community.

All EnvThes concepts are in the process of getting new harmonized URIs after having been restructured and quality-checked and enriched as described in the chapter 2.5.4. After this process EnvThes version 2 (after assuming version 1, although not officially, had been launched with ExpeER in 2015) will be ready to be published. It is planned to integrate the vocabulary in an ontology repository like AgroPortal[27] or any upcoming semantic portal with biodiversity focus (Fiore et al. 2017). Metadata about the vocabulary together with the version number are important features requested by such libraries, as it must be clear which version users are referring to when integrating a semantic resource together with other data or content.

The versioning policy intended to be applied is based on following reasoning:

1. Only substantial changes, such as restructuring in the hierarchy and deletion/additions of top concepts, or the import of the whole vocabulary to a new editing system require a version change. The same will occur according to upcoming needs controlled by the editor team and not after a prefixed period. The version change should be pre-announced in the community to alert users for possible interferences.
2. Any concept enrichments like addition of definitions, mapping relations, language extensions as well as the creation and deletion of concepts below the top concept level are not considered as

[27] http://agroportal.lirmm.fr/

substantial change and will not trigger the publication of a new version. These changes can be done continuously in the vocabulary.

3. Snapshots of main versions will be kept internally in order to allow comparisions.

2.5.3.7 EnvThes governance in the maintenance phase

After the publication of version 2, EnvThes is considered to be in a consolidated status, which has to be maintained and extended according to requests from the wider community. To allow this sort of wider public participation at the further development of the vocabulary it is planned to establish a community friendly platform using freely available solutions like github which is demonstrated to be extremely valuable for community driven discussions around issues for extending ENVO[28].

2.5.4 EnvThes enhancements

Between 2015 and 2018 a number of enhancements were performed on the vocabulary:

- The overall semantics was aligned to O&M model and to the need for semantic harmonization with other parameter vocabularies which lead to a redefinition of the top concepts accordingly
- The structure was refined with a simpler hierarchy reducing the amount of levels
- All concepts were reviewed and repositioned according to the adopted design model
- All concepts were quality checked according to the design principles in chapter 2.5.3.2
- The vocabulary was extended with more than 500 concepts extracted from MS academic keywords
- Concepts were enriched with definitions and mapping relations using alignment techniques

2.5.4.1 Quality enhancement

In order to enhance the quality of EnvThes basic quality checks (Mader et al. 2012) had been conducted and resolved using SPARQL queries by the SPARQL endpoint:

- Omitted or invalid language: Literals should be tagged consequently. Check if all labels have a language tags (en)
- Label conflicts: Check if there are concepts having the same preferred label or the same alternative label
- Orphan concepts: check if there are concepts without any associative or hierachical relationships
- More than one definition for one concept
- Consistent usage of notation
- Valueless associative relations: are removed

[28] https://github.com/EnvironmentOntology/envo/issues/602

2.5.4.2 Extensions

The editor team decided in 2016 to enrich EnvThes with concepts based on the most popular keywords used in Microsoft Research publications[29] related to environmental sciences. From the 24.000 keywords the first 10.000 keywords ranked by popularity were chosen to use as the study basis. They were imported in Excel sheets to be evaluated separately by each of the core members in terms of relevance for EnvThes. 618 terms were approved by the editor team to be integrated as EnvThes concepts. These concepts were further analysed to be enriched with definitions stemming from GEMET, Henderson's Dictionary of Biologic Terms (1995) and DBPedia[30] (where GEMET was considered the most reliable source and DBPedia the weakest source), which was possible for 551 terms. For terms enriched with DBPedia definitions, which are numerous for each concept and more descriptions than classical definitions, a further refinement is still necessary but postponed to be tackled in the next phase (after version 2 publication).

2.5.4.3 Definition alignments

One objective of task 5.6 was to further develop the EnvThes Thesaurus towards production readiness. An important aspect of this work was to seek for possibilities to enrich gaps in its present concept metadata via external resources. The underlying rationale was the assumption that many of the EnvThes concepts would be present in other openly available semantic resources (i.e. taxonomies, thesauri or ontologies) as well, whose metadata potentially covered additional aspects not present in EnvThes and thus invited to be cited and re-used there.

Before the content of such resources can be re-used, they need to be identified first. In recent years, dedicated Web platforms, referred to as Semantic Repositories in this deliverable, have been established to support potential users with finding them. Such repositories usually host tens to hundreds of different resources with, taken together, millions of concepts. Different repositories follow different approaches, one major distinction being whether users can upload resources themselves or their collection rather managed by dedicated curators, which has a strong influence on the overall content quality of the hosted resources. An overview on such different aspects is provided by d'Aquin and Noy (2012). Common to most repositories are means to search across their hosted content and most of them also offer dedicated APIs for machine access. The availability of the latter raises the possibility to perform bulk lookups for potential matches between EnvThes concepts and those from the hosted resources.

Accordingly, this section describes the process and results of a related experiment targeted towards enriching EnvThes concepts lacking explicit descriptions/definitions by looking up potential candidates hosted in Semantic Repositories accessible via the Web. It made use of existing infrastructure, co-developed by EAA throughout the EUDAT 2020 project, for harvesting semantic concepts from their APIs. The so-called "Semantic Lookup Service Aggregator" (SLS-Aggregator), whose underlying motivation is described by Goldfarb & Le Franc (2017), provides a common code framework for harvesting content from Semantic

---

[29] http://academic.research.microsoft.com/, retrieved in March 2016

[30] http://wiki.dbpedia.org/, retrieved in March 2016

Repository APIs which has been used to harvest 13M+ concepts from three well known repositories BioPortal[31], AgroPortal[32] and EBI-OLS[33] as proof of concept in this regard.

Alignment between EnvThes and aggregated concepts

Tab. 2.4 lists four concepts harvested from the three different repositories as an example. The main attributes of interest for the alignment exercise are shown in the different columns, comprised of the concept URI serving as unique identifier and usually resolving to the available concept metadata, the preferred label, alternative labels, description, resource id and repository name. EnvThes concepts were represented in a similar way. Their alignment with the aggregated concepts was subsequently performed via exact string matches between the preferred and/or the alternative labels, all of them converted to lower-case. For each match it was noted if it was based on agreeing preferred, agreeing alternative labels or the respective combinations of them.

The label based alignment resulted in 50,152 potential matches between 2,317 EnvThes concepts and 21,061 of the aggregated concepts. Since only matches between the subset of EnvThes concepts without descriptions and the subset of aggregated ones having this information were of interest, the number of potential matches was reduced to 3,266 between 531 EnvThes concepts and 2,775 from the aggregation. Being based on an automated alignment, it was standing to reason that only a subset of the match candidates was correct and they therefore had to be evaluated by domain experts in this regard.

Tab. 2.4: Exemplary concepts collected from different repositories via the SLS-Aggregator

Concept URI	Pref. Label	Alternat. Label	Description	Res. ID	Repo
http://purl.bioontology.org/ontology/CSP/2341-3000	genetically modified plant	transgenic plant	plants whose genetic material has been altered by genetic manipulation, or their progeny.	CRISP	BioPortal
http://purl.bioontology.org/ontology/MESH/D010103	Oxygen Isotopes	Isotopes, Oxygen	Stable oxygen atoms that have the same atomic number as the element oxygen, but differ in atomic weight. O-17 and 18 are stable oxygen isotopes.	MESH	BioPortal
http://purl.obolibrary.org/obo/ENVO_00002149	sea water	seawater ocean water	Water which has physicochemical properties that have been determined	ENVO	EBI-OLS

31 https://bioportal.bioontology.org/

32 http://agroportal.lirmm.fr/

33 https://www.ebi.ac.uk/ols/index

			by the processes occurring in a sea or ocean.		
http://www.cropontology.org/rdf/CO_336:0000096	Seed quality	Seed appearance	Overall quality of the seed based on defective seed coat, greenish or diseased seeds	CO_336	AgroPortal

Accordingly, the list of 3,266 match candidates was sent to four domain experts for evaluation. They were asked to note if a match was generally correct and, moreover, especially when this was the case for multiple matches, were asked to note if/which match was the best. Additionally, they were provided with optional fields to note comments and predefined reasons for mismatches, especially whether a match was incorrect due to being too narrow/too broad. Due to different availabilities, not all of the experts were able to go through the full list and it was thus decided to split it in a way that each concept was at least evaluated by two of them. Fig. 2.34 shows a Venn diagram on how the task was divided amongst the experts.

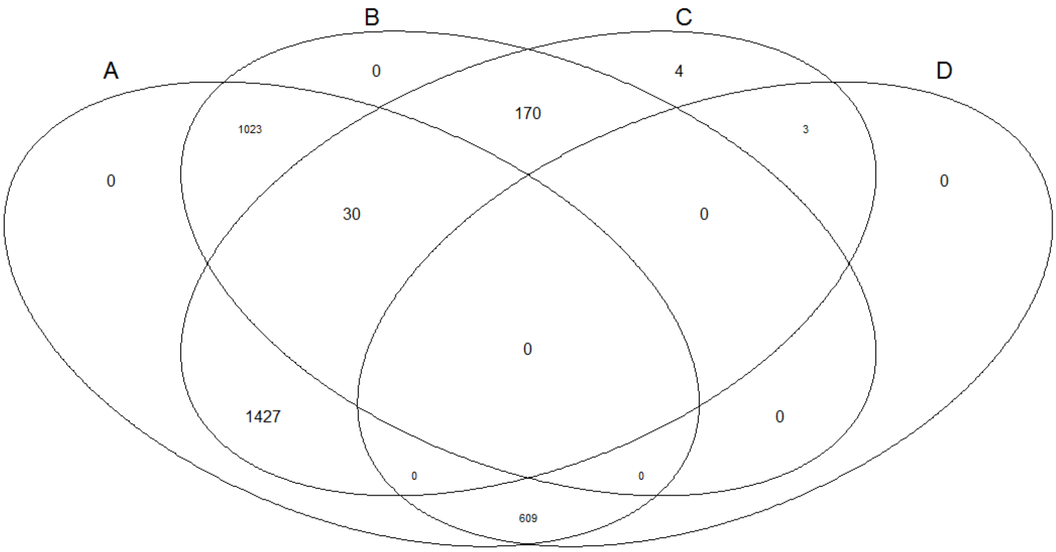


Fig. 2.34: Concept evaluation workload shared between domain experts

In order to quantify the correctness of the matches, it was distinguished between two scenarios. The “soft” scenario required at least one of the evaluators to state the correctness of a match, while the “strict” scenario required their agreement in this regard. The analysis of the outcome of the evaluation revealed 1,539 “soft” correct matches with 449 (84.56%) out of the 531 EnvThes concepts without descriptions, reduced to 975 “strict” correct matches for 364 (68.55%) of them. Moreover, 630 “soft” best matches were found for 430 (80.98%) and only 220 “strict” best matches for 205 (38.61%) of the EnvThes concepts.

The results revealed a number of interesting implications. One relevant observation was the significantly increased correctness for matches found between concepts having the same preferred label compared to those where matches only occurred when alternative labels were involved. As shown in Tab. 2.5 for each of the different scenarios, matches between EnvThes concepts and aggregated ones found via their preferred labels (rightmost column) in all cases yielded more correct results than all the other combinations together. As far as matches tagged as incorrect were concerned, however, a different picture appeared: While matches agreed to be incorrect by both evaluators (“Correct Soft” : “no”) exhibited an inverse behaviour having significantly more incorrect matches based on agreements between only alternative labels, this was not the case for the other scenarios, where also matches between only preferred labels were often tagged as not appropriate, highlighting the sometimes strong disagreement between the evaluators on the quality of a match. Based on the relatively sparsely provided information about reasons for perceived mismatches, one possible reason for this was suggested by the occurrence of cases where one evaluator found the match correct but the other one considered it too narrow or too broad.

Tab. 2.5: Distribution of correct/incorrect matches based on prefered/alternative label agreement

		Alternative - Alternative	Alternative-Preferred	Preferred-Alternative	Preferred-Preferred
Correct Soft	**yes**	45	83	284	1127
	no	1052	141	286	248
Correct Strict	**yes**	27	38	136	774
	no	1070	185	434	601
Best soft	**yes**	25	27	108	470
	no	1072	197	462	903
Best strict	**yes**	7	7	36	170
	no	1090	216	533	1196

Concept matches tagged as correct by at least one evaluator were considered as potential candidates for citing their descriptions in the corresponding EnvThes concept. Obviously, agreement on "best" matches indicated immediate reusability, while the presence of disagreement lowered the confidence accordingly. The available concept matches were thus separated, whenever a best strict match for an EnvThes concept was encountered, all the other "less confident" match candidates for the same EnvThes concept were eliminated. This procedure was repeated for the matches having the next highest confidence, etc. The "hierarchy of confidence" was established as follows: Best matches agreed by two were ranked highest, those tagged as correct by two but as best only by one came next, followed by those tagged as correct and best only by one. Matches agreed to be correct but not best by two evaluators and those only found correct by one of them came last.

Tab. 2.6 shows the results of this procedure. As it becomes visible, the identified expert disagreements lead to multiple conflicting match candidates for all but the "best strict" cases. At this point it thus became necessary to consult a third opinion for these cases, the results are shown in the "Accepted" column in the table. In total, concept matches providing definitions for 371 (69.87% of 531) unique EnvThes concepts were accepted. Moreover, 6 additional "best strict" matches were added in form of "sameAs" links.

Tab. 2.6: Concept matches ranked by confidence

Confidence		Number of match candidates	Number of unique EnvThes concepts	Accepted
	best strict	220	205	205
corr strict	best soft	238	152	127
corr soft	best soft	496	73	35
corr strict		95	6	4
corr soft		14	13	1

Encoding different levels of match confidence via skos:extactMatch and skos:closeMatch

Besides using accepted concept matches for citing their definitions, another result of the alignment were potential links between EnvThes concepts and those found to correspond with them. Explicitly encoding such links within EnvThes would support future data integration and contribute to the "five-star" rating[34] of open data as envisioned by Tim Berners Lee. Supporting such endeavours, the SKOS[35] definition provides a number of options for that, amongst them the properties "exactMatch" and "closeMatch". As their names imply, the former is used to connect different concepts which unambiguously describe one and the same entity while the latter describes a looser relationship in this regard.

This separation corresponded well with the different levels of confidence applied throughout the alignment process, inviting to encode accepted, strict and soft matches accordingly. While "ExactMatch" would directly translate to "accepted matches", it must yet be decided which would be the minimum level of confidence for considering match candidates to be a "closeMatch".

Varying coverage of EnvThes concepts in different repositories and resources

Having obtained correctly matching reusable concepts for about 70% of the EnvThes concepts for which potential candidates were identified suggested that the experiment represented a success. Although a fully automated alignment via label matches only was demonstrated to be too error-prone without human supervision, the involved experts agreed that the effort to control potential matches for correctness was significantly below the effort it would have taken to look up individual definitions for EnvThes concepts on the Web. Since potential match candidates were found for only about one third of the EnvThes concepts without descriptions, it is expected that the identification and inclusion of additional repositories will increase this ratio.

As far as the matching concepts are concerned, it was of interest to check for the repository/resource with the highest overlap with the EnvThes concepts. Considering repositories, Tab. 2.7 shows the counts for match candidates and correct matches in the different evaluation scenarios. In any case, BioPortal by far contained the highest number – about 88.5% – of match candidates, which also propagated to the highest number of

34 http://5stardata.info/en/

35 http://www.w3.org/2009/08/skos-reference/skos.html

matching concepts. As shown in the table, about 22.3% of all match candidates and 25.7% of all accepted matches moreover came from resources hosted in two or more repositories and BioPortal was always involved in such combinations.

Switching from absolute to relative counts allowed a different view on the coverage, highlighting those repositories having higher proportions of matching concepts. Resources hosted in all three repositories for example had a much higher ratio of correctly tagged matches, 31.25% of their match candidates were accepted for citation. A closer look at the respective resources revealed that all of them came from the OBO foundry[36] family of ontologies, a highly curated and thus reliable resource for scientific terminology, including resources such as the “Environment Ontology” (ENVO), the “Phenotype and Trait Ontology” (PATO) or the “Chemical Entities of Biological Interest” (CHEBI).

[36] http://www.obofoundry.org/, retrieved May 2nd, 2018

Co-funded by the European Union

Tab. 2.7: Repositories yielding the highest overlap with EnvThes

Repository	full	Corr soft	Corr soft relative	Corr strict	Corr strict relative	Best soft	Best soft relative	Best strict	Best strict relative	Accepted	Accepted relative
agroportal	138	99	0.7173913	54	0.3913043	40	0.2898551	17	0.12318841	22	0.15942029
Agroportal bioportal	238	94	0.3949580	52	0.2184874	33	0.1386555	11	0.04621849	13	0.05462185
Agroportal Bioportal ebi-ols	128	**110**	**0.8593750**	**82**	**0.6406250**	**60**	**0.4687500**	**28**	**0.21875000**	**40**	**0.31250000**
bioportal	2160	934	0.4324074	592	0.2740741	389	0.1800926	126	0.05833333	238	0.11018519
Bioportal ebi-ols	360	197	0.5472222	148	0.4111111	76	0.2111111	24	0.06666667	44	0.12222222
ebi-ols	237	103	0.4345992	47	0.1983122	32	0.1350211	14	0.05907173	20	0.08438819

The observations on resource level triggered interest in finding out which existing resource, regardless of the hosting repository, had the highest concept overlap with the EnvThes. Tab. 2.8 lists the resources found to overlap with the EnvThes by at least five concepts, ranked by relative proportions of accepted terms. The five resources with the highest absolute accepted overlap are highlighted in bold.

Considering the resources with highest concept overlap revealed “expected” similarities with resources such as ENVO, ECSO or OM, on the other hand suggested significant correspondence with some resources from the medical domain, most dominantly the “National Cancer Institute Thesaurus” and the “Medical Subject Headings”. Looking into the accepted matches for these resources revealed that they included very generic terms such as “standard deviation”, “potential”, “organic”, “percentile”, “saturated” or “time”, but also more specialized ones such as “molecular marker”, “population biology”, “cholorophyta” or “host-parasite interaction”. This observation suggested that such well-established resources provide high quality descriptions which can be reused in different fields and, moreover, at least as far as the used terminology is concerned, highlighted the potential of integrating data from these different domains for cross-disciplinary research.

Tab. 2.8: Resource overlap with EnvThes

Resource Name	full	Corr soft	Corr soft relative	Corr strict	Corr strict relative	Best soft	Best soft relative	Best strict	Best strict relative	Accepted	Accepted relative
OM - Ontology of units of Measure (OM)	**26**	**26**	**1.00**	**25**	**0.96**	**21**	**0.81**	**5**	**0.19**	**17**	**0.65**
Cerrado concepts and plant community dynamics (CCON)	10	10	1.00	7	0.70	6	0.60	5	0.50	5	0.50
Medical Subject Headings (MESH)	**73**	**65**	**0.89**	**46**	**0.63**	**44**	**0.60**	**19**	**0.26**	**33**	**0.45**
The Ecosystem Ontology (ECSO)	**40**	**34**	**0.85**	**19**	**0.48**	**25**	**0.63**	**9**	**0.23**	**18**	**0.45**
Environment Ontology for Livestock (EOL)	14	11	0.79	6	0.43	8	0.57	4	0.29	6	0.43
The Extensible Observation Ontology (OBOE)	17	17	1.00	16	0.94	9	0.53	4	0.24	7	0.41
Trait ontology (TO)	33	24	0.73	17	0.52	18	0.55	8	0.24	13	0.39
Computer Retrieval of Information on Scientific Projects Thesaurus (CRISP)	39	28	0.72	26	0.67	17	0.44	8	0.21	14	0.36
Ontology of Biological and Clinical Statistics (OBCS)	18	17	0.94	14	0.78	7	0.39	3	0.17	6	0.33
Statistics Ontology (SO)	18	17	0.94	15	0.83	5	0.28	3	0.17	5	0.28
Environment Ontology (ENVO)	**154**	**95**	**0.62**	**59**	**0.38**	**48**	**0.31**	**22**	**0.14**	**33**	**0.21**
Semanticscience Integrated Ontology (SIO)	30	28	0.93	24	0.80	9	0.30	2	0.07	6	0.20
Gene Ontology (GO)	42	29	0.69	16	0.38	20	0.48	6	0.14	7	0.17
NanoParticle Ontology (NPO)	30	26	0.87	17	0.57	9	0.30	3	0.10	5	0.17
Experimental Factor Ontology (EFO)	31	20	0.65	14	0.45	7	0.23	3	0.10	5	0.16
National Cancer Institute Thesaurus (NCIT)	**476**	**224**	**0.47**	**135**	**0.28**	**107**	**0.22**	**34**	**0.07**	**68**	**0.14**
Bilingual Ontology of Alzheimer's Disease and Related Diseases (ONTOAD)	44	32	0.73	19	0.43	9	0.20	2	0.05	5	0.11
Chemical Entities of Biological Interest Ontology (CHEBI)	94	30	0.32	25	0.27	11	0.12	4	0.04	9	0.10
The Phenotype And Trait Ontology (PATO)	68	54	0.79	32	0.47	21	0.31	6	0.09	6	0.09

2.5.4.4 Multilingualism

In order to provide a stable semantic backbone the Environmental Thesaurus (EnvThes, see http://vocabs.ceh.ac.uk/evn/tbl/EnvThes.evn) was developed to provide terms annotating data resulting from long term ecosystem research and monitoring (Schentz et al. 2013). EnvThes is built on US LTER Controlled Vocabulary (Porter 2010) extending it with links to other vocabularies of domain interest like EUROVOC, GEMET to name just a few. The vocabulary is open and based on current semantic web standards (SKOS and SPARQL) and supports multilingualism. EnvThes is the core vocabulary used in the DEIMS-SDR to annotate keywords, research topics, observed parameters and various other metadata elements. In addition first tests for the use of EnvThes as multilingual thesaurus for annotation and discovery were made Vanderbilt et al. 2010, Vanderbilt et al. 2017).

Within ECOPOTENTIAL the translation into different languages supported by the project consortium is planned for the last year of the project. This will be targeted to the user needs, e.g. keywords for the discovery. The mapping of terms to GEMET as the main European multilingual thesaurus will be an important first step.

2.5.4.5 Establishment of interlinks to other controlled vocabularies

In the long term perspective it will be important to provide an integrated access to information overcoming the isolated silos of conceptual worlds provided and developed by different communities. The establishment of a unified single vocabulary covering all needs across the scientific domains in a single step is still an illusion. Nevertheless, steps towards the integration of those differing conceptual worlds can be performed, e.g. by searching matching concepts across different vocabularies. The degree of matching, “exact match”, “broader term” or “narrower term” can be defined and the link can be established as human and machine readable link. This links, if exposed as Linked Data SPARQL endpoint, can be accessed and queried through the data integration process. The enrichment with mapping relations will be further refined based on the alignment procedure presented in chapter 2.5.4.3 after version 2 publication, in the last year of the ECOPOTENTIAL project.

3 Data formats and structures

3.1 Introduction

Many initiatives in the different domains aim to standardise the data flows and are working on the harmonisation of data products. The most prominent implementation of these efforts can be seen in the marine domain where harmonisation efforts started already early. Focusing on biodiversity and environmental data infrastructures like ***OBIS*** and ***SeaDataNet*** are one of the examples. Within the long term monitoring network projects like ***ALTER-Net*** and ***EnvEurope*** or networks like ***ILTER*** (e.g. Vanderbilt et al. 2015) and ***LTER Europe*** (e.g. Schentz et al. 2011) worked on the standardisation of the data reporting.

One of the recent developments starting in 2007 is the implementation of the ***INSPIRE Directive***[37] (Directive 2007/2/EC) which aims to create a European Union spatial data infrastructure for the purposes of EU environmental policies and policies or activities which may have an impact on the environment. This European Spatial Data Infrastructure will enable the sharing of environmental spatial information among public sector organisations, facilitate public access to spatial information across Europe and assist in policy-making across boundaries. INSPIRE is based on the infrastructures for spatial information established and operated by the Member States of the European Union. The Directive addresses 34 spatial data themes needed for environmental applications. The Directive came into force on 15 May 2007 and will be implemented in various stages, with full implementation required by 2021.

INSPIRE is based on a number of common principles:

- ***Findability*** - 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.
- ***Accessibility*** - 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.
- ***Interoperability*** - It should be possible to combine seamless spatial information from different sources across Europe and share it with many users and applications.
- ***Re-useability*** - Geographic information needed for good governance at all levels should be readily and transparently available.
- ***Reduce data redundancy*** - Data should be collected only once and kept where it can be maintained most effectively.

On a global scale the ***FAIR principles*** (Wilkinson et al. 2016) provide principles to ensure the findability (F), accessibility (A), interoperability (I) and re-useability (R) of data. Laying out the general guidelines[38] a number of networks and projects are working towards its implementation (e.g.GO-FAIR[39]). The aim of these initiatives is to make data interoperable and enable automatic workflows. These activities focus not only on the standardisation of metadata but also on the data formats.

---

[37] See http://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32007L0002

[38] See https://www.force11.org/group/fairgroup/fairprinciples

[39] See https://www.go-fair.org/

Tab. 3.1 FAIR Data Principles

(F) FINDABLE	F1. (meta)data are assigned a globally unique and eternally persistent identifier. F2. data are described with rich metadata. F3. (meta)data are registered or indexed in a searchable resource. F4. metadata specify the data identifier.
ACCESSIBLE	A1 (meta)data are retrievable by their identifier using a standardized communications protocol. A1.1 the protocol is open, free, and universally implementable. A1.2 the protocol allows for an authentication and authorization procedure, where necessary. A2 metadata are accessible, even when the data are no longer available.
INTEROPERABL E	I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation. I2. (meta)data use vocabularies that follow FAIR principles. I3. (meta)data include qualified references to other (meta)data.
RE-USABLE	R1. meta(data) have a plurality of accurate and relevant attributes. R1.1. (meta)data are released with a clear and accessible data usage license. R1.2. (meta)data are associated with their provenance. R1.3. (meta)data meet domain-relevant community standards.

Taking these principles into account ECOPOTENTIAL will work towards fostering the discoverability and interoperability of data. Based on the LTER metadata community profile implemented in DEIMS-SDR a common set of metadata elements was defined in order to describe and document resulting from in-situ measurement campaigns (see Poursanides et al. 2017). This also included the reference to data quality elements as important information to assess the fitness for use of the data (see Alonso et al. 2017).

An evaluation of the data products documented on DEIMS-SDR provides an overview of the possible data formats provided. Whereas most of the data can be provided as maps (vector or raster format) tabular data for observation are also very frequent. Fig. 3.1 provides an overview on the distribution. The categories on the y axis represent the data product types and the x axis represents the number of possible data formats. Each documented data product counts. In total 103 data products were taken into account.

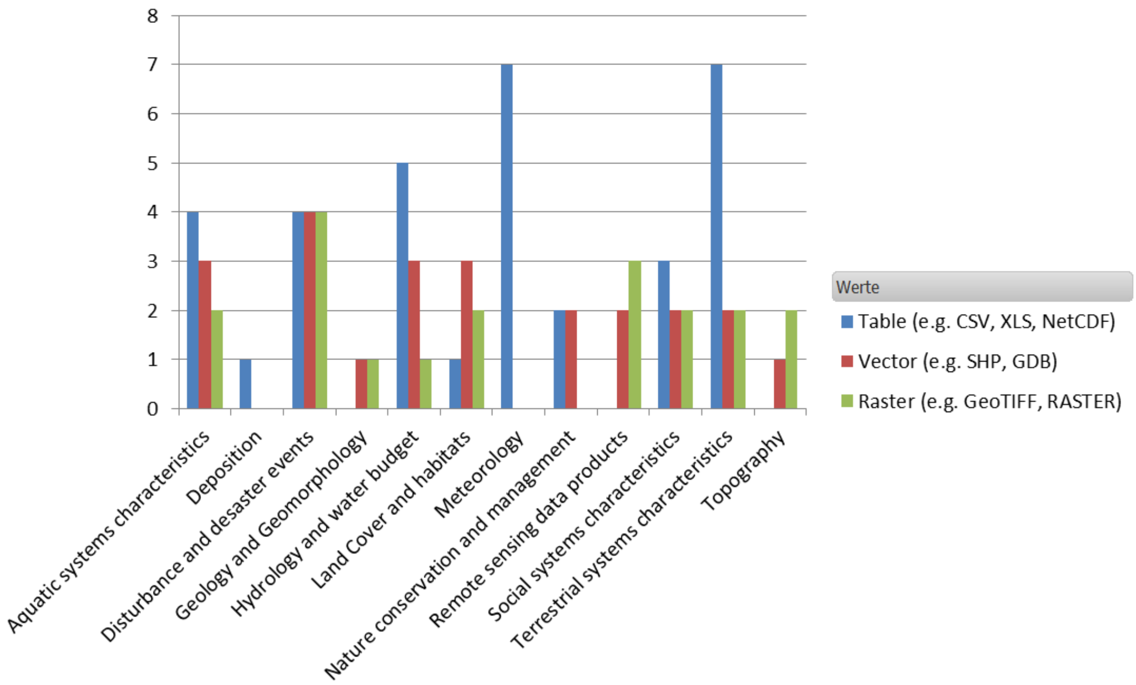


Fig. 3.1 Overview on data formats for the different data products (n=103; Source: DEIMS-SDR)

Tab. 3.2 Overview of documented data products (n=103; Source DEIMS-SDR)

Data product type	n	Table (e.g. CSV, XLS, NetCDF)	Vector (e.g. SHP, GDB)	Raster (e.g. GeoTIFF, raster)
Topography	3		x	x
Meteorology	13	x		
Deposition	1	x		
Geology and Geomorphology	2		x	x
Hydrology and water budget	7	x	x	x
Land Cover and habitats	9		x	x
Terrestrial systems characteristics	27	x		
Aquatic systems characteristics	11	x	x	x
Social systems characteristics	7	x	x	x
Nature conservation and management	9		x	x
Disturbance and desaster events	6	x	x	x
Remote sensing data products	8		x	x
TOTAL	**103**			

Tab. 3.2 provides a summarised view on the results. Typical spatial data are topography, land cover or EO data products being provided mostly in vector (e.g. shp or gdb) or raster (e.g. GeoTIFF, raster or netCDF) format. Field observation data are mostly provided in tabular format (csv, txt, xls, netCDF). The importance of data services, e.g. OGC WxS for spatial data or OGC SOS for time series, are of minor importance currently in the data provision. Nevertheless data services will play an important role for the harmonisation of the data formats and data access.

3.2 Methods and challenges

Based on the evaluation of the documented data products and the data provided in the project context an assessment of relevant data formats was done. This was based on previous work (e.g. EU-BON) or relevant research and data infrastructures (e.g. OBIS, eLTER). The most appropriate data formats were listed and described.

As no formal data reporting process and therefore a strict harmonisation can be applied top down for the protected areas a more collaborative approach was taken. In addition the usage of the data differs greatly between the different applications (e.g. models or EO data validation).

The work resulted in (a) general recommendations for the provision of data and (b) a list of relevant data formats which should be supported by the data provides. As stated earlier the implementation of data services and the focus on syntactic harmonisation by this should be fostered in the near future.

ECOPOTENTIAL did not aim to establish another harmonised data store, which would be difficult to be maintained after the runtime of the project, but focused on the integration of metadata on datasets and data services.

3.3 General recommendations

A vast majority of the data shared in the project context was provided as data files which are a convenient means of data transportation. This allows an the data provider to easily maintain and update the information provided. In order to ensure long term usability of the data a number of general recommendation should be considered.

Recommendation 1 All protected areas need to be **registered as observation facilities** (sites) in the DEIMS Dataset and Site Registry in order to allow for linking of the datasets and services and provide the link the observation context.

The documentation of the observation and experimentation facilities allows to provide information on the setup and organisational context of the observation. The details for the documentation are laid out in deliverable D5.2 (Poursanides et al. 2017). Data collected in the same context can be linked by the observation and experimentation facility in DEIMS-SDR. DEIMS-SDR provides an easy and light weight registration for the sites.

For the DEIMS-SDR:Site a basic set of information needs to be provided for the registration which includes the name, the location and a contact point. Additional information like research context, keywords, parameters observed or the infrastructure implemented are recommended. Each observation and experimentation facility is identified by a unique SITE_UUID.

As many in-situ field observations are conducted within the protected areas and a common purpose of the observations is given (e.g. management of the PA) the PAs can be seen as observation and experimentation facilities in a wider sense. Many of the PAs are already directly involved in the network of long term monitoring sites (e.g. LTER Europe, CZO) in Europe and on a global scale (e.g. ILTER).

Whenever datasets are provided from a PA the corresponding SITE_UUID should be provided in the metadata documentation for the dataset or data service.

Recommendation 2 All in-situ datasets and services need to be **documented by metadata** following the dataset community profile for LTER and ECOPOTENTIAL. This allows the discoverability of the data. DEIMS-SDR or any other ISO 19115 compliant metadata editor can be used. This also should include the documentation of the methods applied and the parameters provided (see Poursanides et al. 2017)

In order to ensure discoverability and re-useability of data the provision of good quality metadata is needed. The metadata should not only allow to discover the data but also describe on how the data were generated. For ECOPOTENTIAL the eLTER metadata model was adopted and extended to the needs for the data flows. The basic models are outlined in the deliverable D5.2 (Poursanides et al. 2017). If appropriate metadata editors and catalogues are missing DEIMS-SDR can be used to document and share file based data. For any data file shared via B2SHARE a PID using an EPIC handle is issued. Alternative metadata editors supporting ISO19115/ISO19139 as basic metadata format can be used. This also ensures the compatibility with the INSPIRE metadata regulation and enables the integration of the metadata into distributed metadata catalogues.

Metadata should be open and available even if the data are restricted. An important information in the metadata is the legal constraint and the information on how to access the data. This can be directly via file download or data services or indirect via email.

Recommendation 3 For data provision the most **fitting community standard** should be used. The structure of the data (file) needs to be described in the metadata or in an accompanying file (e.g. data_structure.csv)

Many data harmonisation efforts exist in different domains making it difficult to apply an universal standard across different domains. Especially protected areas are serving to different communities and applying different reporting obligations. In order to reduce the burden of additional data harmonisation work the focus on relevant and well introduced community standards is recommended. If no community standard is applied currently chapter 3.4 provides an overview on relevant standards recommended for ECOPOTENTIAL.

The data reported should include the basic information on the observation context. This encompasses the information on the where, when, what and how the observation was made. A general model is provided by

OGC Observation and Measurement[40] and ISO19156:2011 defining the main elements like Feature of Interest, Observed phenomenon, etc. Any data standard should follow the main elements of the observation described by the metadata on dataset level or record level.

In addition the naming of the parameter should follow the basic model of O&M. The work on EnvThes is prepared to develop an ontology allowing the linkage of 'real-world' parameter names to the atomic concepts used in O&M. This also allows the standardisation and harmonisation of parameters.

Recommendation 4 If list of values and reference lists are used **standard vocabularies or registered reference lists** should be used referencing to the URN of the term and list. If this is not possible the reference list need to be provided in an accompanying file (e.g. reference_list.csv). This is especially true for habitat lists and species lists.

An important part for the usability of the data is the provision of the reference lists used for the different fields. Especially for categorical variables reference lists are needed. These can either be based on global and open lists (e.g. identified by a URN or PID) or local lists. The usage of common and open accessible reference lists is recommended as the content is machine readable and accessible. One of the examples is the INSPIRE Code List Registry[41] on European scale. This allows the referencing to a web resource. If possible standards for controlled vocabularies like SKOS should be adopted.

If referencing is not possible the reference lists need to be provided together with the data. This should include the information on the identifier applied, the abbreviation, the name, the definition and the source of the reference list. This information should be stored with an accompanying CSV file to the dataset.

For biodiversity and habitat data wherever possible common agreed reference lists should be used. This includes:

- **EU-NOMEN**

The EU-Nomen portal enables the correct use of species names and their classification, to more accurately manage information on animals and plants. This is the first all-taxa inventory for European species. (Source: EU-Nomen website). The portal is one of the outcomes of the PESI (Pan-European Species Directories Infrastructure) project. The objective of this project was to integrate and secure taxonomically authoritative species name registers that underpin the management of biodiversity in Europe. PESI will integrate the three main all-taxon registers in Europe, namely the European Register of Marine Species, Fauna Europaea, and Euro+Med PlantBase in coordination with EU based nomenclatures and the network of EU based Global Species Databases. (Source: PESI website)

- **EUNIS**

EUNIS data are collected and maintained by the European Topic Centre on Biological Diversity for the European Environment Agency and the European Environmental Information Observation Network to be

[40] See http://www.opengeospatial.org/standards/om

[41] See http://inspire.ec.europa.eu/registry/

used for environmental reporting and for assistance to the NATURA2000 process (EU Birds and Habitats Directives) and coordinated to the related EMERALD Network of the Bern Convention. The Species part of EUNIS contains information about more than 275 000 taxa occurring in Europe. However, the amount of information collected on each species varies in accordance with the potential use of the data. (Source: EEA EUNIS website)

- **NATURA2000**

Natura 2000 is a European network of important ecological sites under the Birds Directive and Habitats Directive and has the aim of conserving biodiversity on land and at sea by protecting the most seriously threatened habitats and species across Europe. This legislation is called the Habitats Directive (adopted in 1992) and complements the Birds Directive adopted in 1979. Within the legislation special attention is paid to two groups of species. The first consists of fauna species listed in Annex II to the Habitats Directive. These include a number of marine mammals and certain fish. Secondly, various sea birds are also very important to the Natura 2000 network. These are protected under the Birds Directive, and their prevalence, population size and distribution are criteria for the nomination of Special Protection Areas (SPAs) that form part of Natura 2000.

Recommendation 5 Parameter names should follow **community standards** and should be taken from standard vocabulary (e.g. EnvThes, CF convention)

The parameter names describing the phenomenon observed should be taken from a standard vocabulary adopted for the communities. A general list for the mainly terrestrial observations is provided by *EnvThes:Measures* also used in DEIMS-SDR. For the climate change community the CF convention[42] provide an important source. Research infrastructure like SeaDataNet or OBIS are maintaining their own parameter lists.

In addition the parameter names from core vocabulary (e.g. EnvThes) should be used for the annotation of datasets with keywords (discovery) and assignment of common parameter names.

Recommendation 6 **Non-proprietary and uncompressed file formats** should be preferred in order to ensure the long term usability of the data formats and to reduce the efforts for data transformation.

When selecting file formats[43] for publishing and archiving, the formats should ideally be: (a) Non-proprietary, Unencrypted and Uncompressed, (b) in common usage by the research community and (c) adherent to an open, documented standard. This should include aspects like

- the interoperability among diverse platforms and applications,
- the formats are fully published and available royalty-free

[42] See http://cfconventions.org/standard-names.html

[43] See https://library.stanford.edu/research/data-management-services/data-best-practices/best-practices-file-formats

- the formats are fully and independently implementable by multiple software providers on multiple platforms without any intellectual property restrictions for necessary technology,
- the formats are developed and maintained by an open standards organization with a well-defined inclusive process for evolution of the standard.

For long term archiving in addition strategies to transform and migrate data formats need to be taken into consideration. This is especially true when a data format is not supported any more (e.g. mdb). Examples for recommended data formats are:

- Containers: TAR, GZIP, ZIP
- Databases: XML, CSV
- Geospatial: SHP, DBF, GeoTIFF, NetCDF
- Moving images: MOV, MPEG, AVI, MXF
- Still images: TIFF, JPEG 2000, PDF, PNG, GIF, BMP
- Sounds: WAVE, AIFF, MP3, MXF
- Statistics: ASCII, DTA, POR, SAS, SAV, R
- Tabular data: CSV
- Text: XML, PDF/A, HTML, ASCII, UTF-8

Recommendation 7 **Naming conventions** should be followed in order to ensure the unique idenfication and the reusability for humans.

In order to ensure the usability for humans the file names[44] should allow the identification of the content from the name. Naming convention should ensure the consistent application. This could include e.g information on the protected area, the project or observation campaign (name or acronym), the location/spatial coordinates, the researcher name/initials, the date or date range of observation, the type of data, the conditions of use or the version number of file. A three-letter file extension for application-specific files (e.g. CSV).

The provision of a readme.txt file that explains your naming format along with any abbreviations or codes used is recommended.

Other tips for file naming

- A good format for date designations is YYYYMMDD or YYMMDD. This format makes sure all of your files stay in chronological order, even over the span of many years.
- Try not to make file names too long, since long file names do not work well with all types of software.
- Special characters such as ~ ! @ # $ % ^ & * ( ) ` ; < > ? , [ ] { } ' " and | should be avoided.
- When using a sequential numbering system, using leading zeros for clarity and to make sure files sort in sequential order. For example, use "001, 002, ...010, 011 ... 100, 101, etc." instead of "1, 2, ...10, 11 ... 100, 101, etc."

[44] See https://library.stanford.edu/research/data-management-services/data-best-practices/best-practices-file-naming

- Do not use spaces. Some software will not recognize file names with spaces, and file names with spaces must be enclosed in quotes when using the command line. Other options include:
- Underscores, e.g. file_name.xxx
- Dashes, e.g. file-name.xxx
- No separation, e.g. filename.xxx
- Camel case, where the first letter of each section of text is capitalized, e.g. FileName.xxx

3.4 Relevant data models for ECOPOTENTIAL

Within ECOPOTENTIAL relevant data models were evaluated on the usability for the data flows. This resulted in a range of different relevant data models which are referenced in the following chapter. ECPOTENTIAL was building on the expertise and work of EU-BON as relevant project in the biodiversity domain (see Saarenmaa et al. 2014).

As the applications used in the project context showed a wide variety a light weight approach for data harmonisation was chosen in order to reduce the barriers for data provision for the data providers.

3.4.1 Biodiversity data

In order to facilitate the exchange of biodiversity data, the TDWG community (Biodiversity Information Standards, formerly known as Taxonomic Databases Working Group), which is affiliated with the International Union of Biological Sciences, develops, adopts and promotes standards and guidelines for the recording and exchange of data about organisms and related data. The most commonly applied data formats are listed below.

ABCD (Access to Biological Collection Databases)[45] was developed by the ABCD task group, is a ratified TDWG standard with the purpose to document biological collection and observation data. ABCD is an XML-Schema (Extensible Markup Language) developed to facilitate the acquisition, storage and sharing of incidental point records to inform about the geographic distributions of organisms and/or their representation in collections (physical or digital). One of the largest global aggregator of biodiversity data in the form of occurrence records – the Global Biodiversity Information Facility GBIF (https://www.gbif.org/) – draws its data together through the use of the ABCD standard. ***ABCD-EFG***[46] is an extension of the ABCD standard for use with paleontological, mineralogical and geological digitalized collection data (Petersen et al. 2018).

45 See http://www.tdwg.org/standards/115

46 See https://terms.tdwg.org/wiki/ABCD_EFG

Darwin Core[47] is a body of standard by TDWG designed to facilitate the exchange of information about the geographic occurrence of species and the existence of specimens in collections. The Darwin Core Standard can be viewed as an extension for biodiversity data of the metadata standards developed by the Dublin Core Metadata Initiative (www.dublincore.org). The Darwin Core is based on taxa and their occurrence data as documented by observations, specimens and samples in collections (physical or digital), as well as additional taxon-related information. It contains a glossary of terms to provide stable semantic definitions with the goal of facilitating the sharing of information about biological diversity by providing reference definitions, examples, and commentaries. The resource to understand the term definitions and their relationships to each other (the RDF normative; http://rs.tdwg.org/dwc/rdf/dwctermshistory.rdf) is written in the Resource Description Framework (RDF). Besides using ABCD, the Global Biodiversity Information Facility GBIF (https://www.gbif.org/) also draws its data together through the use of the Darwin Core standard.

GGBN Data Standard[48] is a set of vocabularies designed to represent tissue, DNA or RNA samples associated to voucher specimens, tissue samples and collections. It is developed by the Global Genome Biodiversity Network (GGBN) based on the ABCD DNA extension.

Humboldt Core (Guralnick et al. 2017). With increasing recognition for the necessity of detailed, harmonized biodiversity data e.g. for global science-policy efforts and evaluations for biodiversity conservation (e.g., IPBES, GEO BON, etc.), new conceptual informatics frameworks have been developed to cover a wider spectrum of biodiversity evidence. The Humboldt Core addresses the so far widely lacking standard for assessing and describing species communities using taxonomic inventories. The main differences of inventory datasets as compared to incidental point records are that inventories cover a defined area of land/water/air space for a particular period of data collection, focussing mostly on multiple biological entities (e.g. species), often including some form of abundance or biomass estimate.

Audubon Core[49] is a set of vocabularies for describing biodiversity-related multimedia resources. Multimedia Resources are digital or physical artefacts which comprise more than text, including pictures, artwork, sound, video, animations, etc. The Audobon Core standard addresses the management of the respective media and collection, descriptions of the content, the taxonomic, geographic, and temporal coverage, and the appropriate way to retrieve, attribute and reproduce them.

Ecological Metadata Language (EML, Michener et al. 1997***)***[50] is a metadata standard developed for earth, environmental and ecological sciences which aims at providing a high-quality metadata specification for describing ecological data (https://knb.ecoinformatics.org/#external//emlparser/docs/index.html). It is an XML-based standard, implemented as a series of XML document types that can be used in a modular manner

[47] See http://rs.tdwg.org/dwc/

[48] See https://terms.tdwg.org/wiki/GGBN_Data_Standard

[49] See https://terms.tdwg.org/wiki/Audubon_Core

[50] See https://knb.ecoinformatics.org/#external//emlparser/docs/index.html

to document the data. Each EML module is designed to describe one logical part of the total metadata that should be included with any ecological dataset.

3.4.2 Environmental data

OBIS-ENV[51] Recently, the Ocean Biogeographic Information System (OBIS), launched the OBIS-ENV-DATA project which developed a standard for managing datasets that combine biological, physical and chemical measurements as well as details regarding the sampling or observation methods, equipment and effort (De Pooter et al., 2017). A new standard and guidelines were proposed in order to ensure that these combined data sources can remain together allowing for new scientific research whereby these biological and environmental parameters are analysed together. The new standard builds on the Darwin Core Archive (DwC-A), which is already used in OBIS and GBIF, and governed by the Executive Committee of the Biodiversity Information Standards, also known as the Taxonomic Databases Working Group (TDWG). It consists of a DwC Event Core, a DwV Occurrence Extension and a new DwC MeasurementOrFact Extension. This new structure enables the linkage of measurements or facts - quantitative and qualitative properties - to both sampling events and species occurrences, and includes additional fields for property standardization. Finally, the use of the new parentEventID DwC term, enables the creation of a sampling event hierarchy. In conclusion, the new standard allows for the structured exchange of biogeographic data with additional data such as sampling methodology, animal tracking and telemetry data, biological measurements (e.g., body length, percentage live cover, ...) as well as environmental measurements such as nutrient concentrations, sediment characteristics or other abiotic parameters measured during sampling. This project was initially launched in the context of marine datasets but can equally be used for terrestrial datasets combining biogeographic and environmental data into one Darwin Core Archive.

[51] See https://www.seadatanet.org/Standards/Data-Transport-Formats

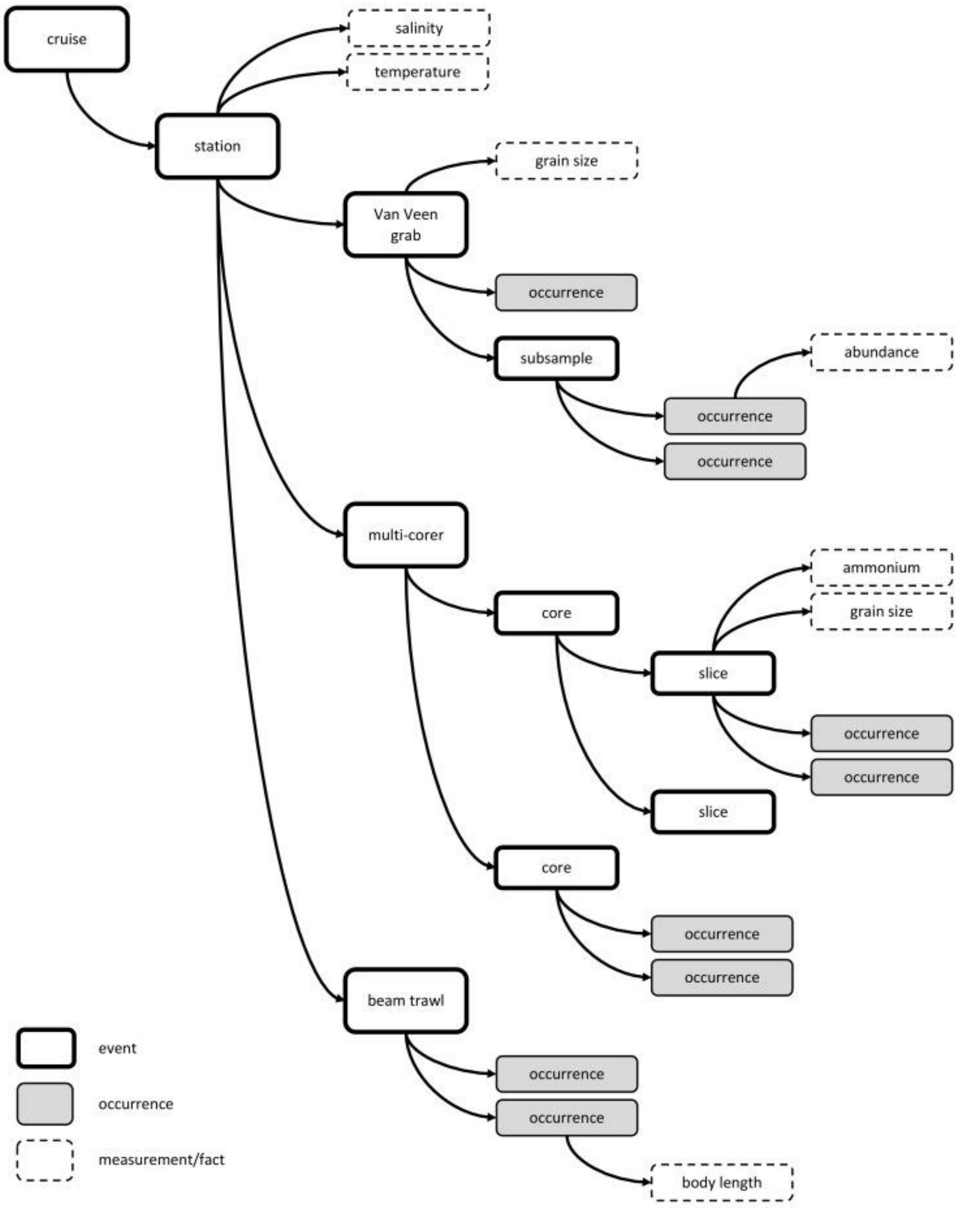


Fig. 3.2 A hypothetical example based on a complicated sequence of sampling events at a given sampling location (from De Pooter et al. 2017).

In the example (see Fig. 3.2) the bold rectangles are sampling events, the dashed rectangles measurements or facts, the grey rectangles are occurrences. The arrows between the rectangles illustrate the (hierarchical) relations between the different sampling events and between events and their associated occurrences and measurements. The example shows data sampled using a Van Veen grab, a beam trawl, and a multi-corer. The macrobenthos analysis was based on the complete Van Veen grab sample, while the meiofauna analysis was based on subsamples. The multicore sample was divided into different depth slices. Likewise, an abiotic measurement can refer to the entire sample or to a subsample.

SeaDataNet[52], the pan-European infrastructure to ease the access to marine data measured by the countries bordering the European seas, supports three different data exchange formats: Ocean Data View (ODV), MEDATLAS and CFPOINT (CF NetCDF). The SeaDataNet ODV file standard is a spreadsheet like format with three different types of columns: metadata, primary variable data and data. It can be used to map data from different sources (profiles, time series and trajectories) on to a SeaDataNet compatible data model. The file structure consist of first a set of SeaDataNet specific comment rows which are used for mapping the terms used in the (meta-)data to controlled vocabularies and for linking to external resources. Followed by a column header row with a set of required metadata columns and finally the data rows. As part of SeaDataNet II project, a set of guidelines and recommendations for including biological data in the framework of SeaDataNet have been published. For this a specific variant of the ODV files for biology data was made available. It uses specific additional mandatory data parameter columns including minimum and maximum depth of observation, sample identifier, sampling effort, scientific name and identifier, sex, life stage and observed individual count. Additionally, some fields are mandatory in some predefined cases e.g. event information in order to group different samples from a single deployment of an instrument. Next to purely biological information, additional environmental information such has habitat characteristics and ecotoxicology can be provided.

Datafile formats: ODV, MEDATLAS, NETCDF https://www.seadatanet.org/Standards/Data-Transport-Formats

Format documentation: Recommended data transport for biological data in the framework of SeaDataNet https://www.seadatanet.org/Standards/Data-Transport-Formats

eLTER Data Reporting Format (eLTER VA Format, see eLTER deliverable D3.3 (in prep.)) provides a flexible extensible data format to share biological as well as bio-geochemical data in a file based manner. The eLTER Data Reporting format is based on the common specification of the ICP Integrated Monitoring programme[53] and was extended due to the specific needs. Starting with the EnvEurope project[54] the format was further developed in the eLTER H2020 project[55] in order to create a common basis for the provision of data across different scientific domains. The data specification provides a common vocabulary of 57 terms for the description of the data ranging from the observation location (e.g. SCODE) to the observation itself (e.g. VALUE). Links to the observation facilities (e.g. LTER site or protected area) as well as to the observation stations (e.g. plots or sensors) are core elements of the data specification.

The observation records are organised in lines (see Tab. 3.3) which allow providing additional metadata, e.g. data quality flag or detection limit, for each value. The basic observation is encoded using a combination of WHERE (SCODE), WHAT (SUBST), WHEN (TIME) and VALUE. In case of biodiversity observation, e.g.

52 See https://www.seadatanet.org/Standards/Data-Transport-Formats

53 See http://www.syke.fi/en-US/Research__Development/Ecosystem_services/Monitoring/Integrated_Monitoring/Manual_for_Integrated_Monitoring

54 See http://www.enveurope.eu/ (ENVeurope 2010-2014 Life Enviroment Project LIFE08 ENV/IT/000399)

55 See http://www.lter-europe.net/lter-europe/projects/eLTER

vegetation releveés additional columns, e.g. TAXA are added, but following the basic structure of the reporting format. In order to simplify the data provision a basic and extend data format are supported.

Tab. 3.3 Example file for eLTER Data Reporting Format, data.csv

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
…	…	…	…	…	…	…	…

In addition to the observation values a detailed documentation of the observation locations (e.g. stations), the parameter names, the methods as well as the used reference lists (e.g. species lists) need to be provided as separate tables. This can either be encapsulated in a single XLS file (not recommended) or a series of CSV files using UTF-8 encoding.

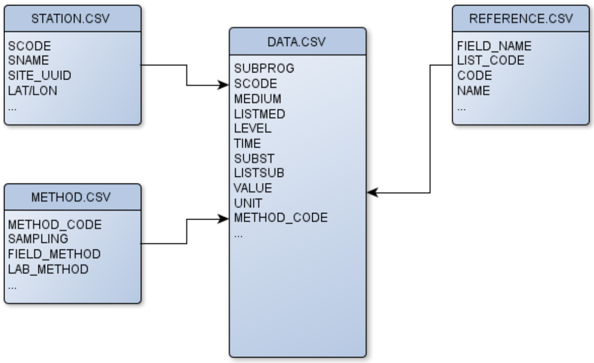


Fig. 3.3 Structure of the eLTER Data Reporting

An example can be downloaded from https://data.lter-europe.net/deims/dataset/f619990c-8ba9-441b-90f3-468c8313886b and http://hdl.handle.net/11304/8428ccc1-5fcc-4005-a379-cccd76c1ce31. For each of the data files uploaded to B2SHARE in addition to the DATASET_UUID as unique identifier a persistent identifier is issued.

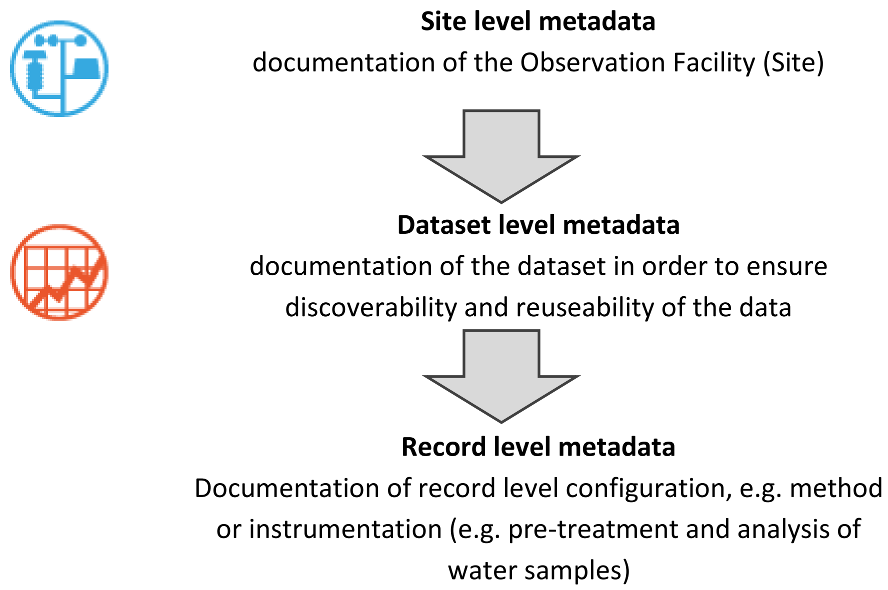


Fig. 3.4 Metadata levels for the eLTER Data Reporting Format

The integration of record level metadata allows the provision of full documentation of a single observation value from dataset level metadata to record level metadata. Using DEIMS-SDR as registry of the broader observation facilities additional information on the observation context, e.g. research topics or basic setup, can be provided. The observation facility can be shared using the INSPIRE EF application schema implementation for DEIMS-SDR. Fig. 3.4 provides an overview on the different levels.

3.4.3 Spatial data

Within the INSPIRE directive of the European Parliament and of the Council (Directive 2007/2/EC) the establishment of a spatial data infrastructure for Europe in order to support the Community environmental policies is defined. This results in a number of technical implementation rules for the 34 identified relevant data themes which enable the interoperable provision and use of spatial datasets and services.

For the provision of interoperable and harmonised datasets and services from Protected Areas the following data specifications and recommendations are of general interest: (a) Protected areas, (b) Species distribution, (c) Habitats and Biotopes, and (d) Environmental Monitoring Facilities.

3.4.3.1 Protected sites[56]

The INSPIRE Directive defines a Protected Site as an "*Area designated or managed within a framework of international, Community and Member States' legislation to achieve specific conservation objectives*" [Directive 2007/2/EC]. According to the International Union for the Conservation of Nature (IUCN) a

[56] See https://inspire.ec.europa.eu/Themes/117/2892

Protected Site is an "*area of land and/or sea especially dedicated to the protection and maintenance of biological diversity, and of natural and associated cultural resources, and managed through legal or other effective means*". This includes all kinds of protected areas being in terrestrial, aquatic and/or marine environments, and may be under either public or private ownership. Protected Sites may differ greatly in their reasons for protection, their designation and their management including protected areas according to the Habitats Directive (1992) (Directive 92/43/EC), the Birds Directive (1979) (Directive 79/409/EC), the Water Framework Directive (2000) (Directive 2000/60/EEC), the World Heritage Convention (1975), the Ramsar Convention (1971), the Barcelona Convention (1976), the Helsinki Convention (1974), the OSPAR Convention (1992) and according to national laws of each European country and EU and international sector policies (for example, relating to nature conservation, forests or fisheries).

A simple application schema defines the boundaries and designation value (protection status) of the protected area. This includes a very limited set of fundamental attributes, including geometry, identifier, name, legal foundation (designation status), date, and document reference. Only current Protected Sites are included. The specification allows extensions thus additional application schemas specific to other purposes may be created by Member States or other European organisations and added to the model. Specific application schemas may be updated as a result of decisions and agreements at European level.

Data based on the specification can be downloaded from the European Environmental Agency[57]. In addition to the Protected Sites the Environmental Monitoring Facility schema can be used to share information on the observation context and the linked datasets provided.

3.4.3.2 Species distribution[58]

The INSPIRE Directive defines **Species Distribution** as geographical distribution of occurrence of animal and plant species aggregated by grid, region, administrative unit or other analytical unit (Directive 2007/2/EC). The definition refers to a distribution of occurrence of a given species and is not intended to cover the raw field observation data which are linked to the Environmental Monitoring Facilities then. The definition interprets occurrence as the spatial representation of a species at a specific location and a specific time period, rather than being equivalent to an observation. This includes also an extended model were linking to the original observations is used as source for the calculation of aggregations. The definition refers to a "distribution of occurrence" and as such does not intend to cover the actual raw field observations. These are considered as the input for the calculation of any given species distribution. Original field observations should be covered under the Environmental Monitoring Facilities theme, even if not further specified at this place. However, sometimes the distinction between the raw and aggregated data can be vague and not clear. For example if the distribution represents the outputs of a grid based sample programme at a single point in time. Nevertheless, the specification for Species Distribution aims to allow for the inclusion of raw observation data.

The "Species Distribution" category of spatial data defined in the INSPIRE Directive is one of several themes in a wider grouping of biological organisms and biological communities - biodiversity. Species Distribution

---

57 See https://www.eea.europa.eu/data-and-maps/data/natura-9#tab-metadata

58 See https://inspire.ec.europa.eu/Themes/133/2892

includes species occurrence as points, grid cells at different scales or polygons of specific defined areas. In order to achieve harmonization across different data sources the usage of EU-Nomen as the main species reference is recommended. Additionally references to species names using the European Nature Information System (EUNIS) and Natura2000 (priority species) could be used. In addition to the global names a mapping to local species names (as e.g. occurring in the original data) can be provided. In order to ensure uambiguaty and machine readability the referencing should be done via persistent identifier to globally and online available species lists.

The *SpeciesDistributionDataSet* represents a collection of instances of the species distributions (*SpeciesDistributionUnit*) integrating species distribution information aggregated over grid cells or areas of any other analytical unit with geometry (like e.g. administrative areas) and over periods of time where occurences have been recorded. The species distribution units carry information on the identity of a species occurring and information about life cycle of the data object. Furthermore, the *DistributionInfoType* provides information about the occurrence category or population size, the residency status and information on the sensitiveness i.e. whether the combination of location and species in a specific case is sensitive for nature protection purposes.

An extended Species distribution schema allows to provide a linking to observation data specified within the *Environmental monitoring facilities* specification (Annex III: EF) and in addition includes a *darwinCoreTriple* attribute which allows connection to original observational data that can be accessed from GBIF data providers.

3.4.3.3 Habitats and Biotopes[59]

The INSPIRE Directive defines Habitats and Biotopes as geographical areas characterised by specific ecological conditions, processes, structure, and (life support) functions that physically support the organisms that living there. They include terrestrial, fresh water and marine areas distinguished by geographical, abiotic and biotic features, whether entirely natural or semi-natural. Common to all spatial data that fall under this category is a characterisation of the distribution of geographical areas as functional areas for living organisms: biotopes are the spatial environment of a biotic community; habitats are the spatial environment of specific species. In order to cover the different use cases two different application schemas are provided: one on the characteristics (mapping) and one on the distributions of habitats and biotopes (based on reference areas, e.g. grid or country).

Similar to the variety of available species lists, different countries or communities developed their own habitat classification systems (e.g. GHC). This is often due to the fact that habitat mapping focuses on specific purposes or specific constraints (e.g. financial). As such, there may be difficulties in mapping accurately certain habitat classes between national nomenclatures and also between national and European nomenclatures. Harmonisation needs to take into account local, national and international habitat classification systems. Harmonisation can be achieved, if there is one habitat classification system, which serves as “first among equals” to which all other classification systems can be mapped. The EUNIS habitat classification system serves this purpose. In addition a set of habitat types has been drawn up for the Marine

[59] See https://inspire.ec.europa.eu/Themes/146/2892

Strategy Framework Directive. Local or national habitat classification can be used as well as long as a link is provided to these references.

As a result, all habitat features will have one or more habitat type encodings, obligatory one(s) from, most frequently, the EUNIS habitat classification code list and optional one(s) from a registered code list related to an international, national or local habitat classification system.

3.4.3.4 Land cover[60]

The INSPIRE Land Cover theme is focused on the physical and biological cover of the earth's surface including artificial surfaces, agricultural areas, forests, (semi-)natural areas, wetlands, water bodies (Directive 2007/2/EC). Land cover is defined as the physical or biological description of the earth surface and by this different from the land use (see INSPIRE Annex III, theme number 4) which is defined as the description of the use of the Earth surface. A land cover data set consists of a collection of land cover units. These units may be points, polygons or raster cells (resulting in two core models, one for vector data and one for raster data). The land cover data set is also linked to a code list (e.g. the CORINE Land Cover code list). CORINE Land Cover as well as most regional and national land cover data sets, can be represented using one of the core models. Land cover information used in monitoring linked to EU agricultural policy (IACS), in carbon monitoring (LULUCF) and used in land and ecosystem accounting based on CORINE Land Cover (LEAC).

The common, conceptual core model for land cover data has the following structure: A land cover data set consists of a collection of land cover units. These units may be points, polygons or raster cells (resulting in two core models, one for vector data and one for raster data). The land cover data set is also linked to a code list (e.g. the CORINE Land Cover code list). The code list is a nomenclature of land cover classes where each class is represented by a code and a name. At each land cover unit, the land cover has been observed on one or more observation dates. The multiplicity of observation dates is introduced in order to be able to describe land cover change. For each observation date attached to a land cover unit, the observation is represented by one or more codes from the code list (representing land cover classes). Several codes are allowed in order to allow the use of mosaics. It is also possible to add a percentage showing the relative presence of each class within the land cover unit.

The core model proposed for the INSPIRE implementing rules represents a land cover data set consisting of a collection of land cover units. The land cover unit can be a point, a polygon or a raster cell. The land cover data set is also associated with a code list with legal land cover codes and their names (e.g. the CORINE Land Cover code list). A land cover code from the code list is assigned to each land cover unit. The core model furthermore allows several codes to be assigned to each land cover unit (in order to represent mosaics). It is also, in this case, possible to attach a *Covered percentage* to each code in the mosaic. Finally, the core model allows the observation to be attached to an observation date, and several observation dates to be attached to each land cover unit. The observation date is included because it provides important metadata at the observation level and also because it allows representation of land cover change.

[60] See https://inspire.ec.europa.eu/Themes/123/2892

The data specification does not prescribe or recommend any particular land cover nomenclature for use in INSPIRE. There is a multitude of different ways to describe land cover. This is partly due to the wide range of aspects of the environment embraced by land cover, but also due to the many different uses of land cover data. There is only one "real world" but many different descriptions of this world depending on the aims, methodology and terminology of the observer. It is therefore a misguided approach to enforce a single classification system as the common classification system for Europe.

3.4.3.5 Environmental Monitoring Facilities[61]

Environmental Monitoring Facilities (EF) is a cross-cutting data specification dealing in principal with any thematic area related to environment observations. The scope could reach from in-situ surveys and sampling campaigns to satellite based data. In this respect the '**environmental monitoring facility**' can be defined as "*… the linking element between spatial data themes as defined by the INSPIRE Directive and observations and measurements on specific aspects of the environment (e.g. air quality, atmospheric conditions, water quality)*". The data specification tries to address both dimensions. The data specification addresses both aspects in scope for EF; on the one hand the description of a monitoring facility and on the other hand the link to observations and measurements (based on ISO 19156 Observations and Measurements (O&M)). This includes also the link to information describing aggregations/collections of monitoring facilities and their thematic or organisational grouping and background.

The application schema for *Environmental Monitoring Facilities* contains 4 spatial object types:

(a) ***Environmental Monitoring Programme (EMP)*** which is defined as "*… an Environmental Monitoring Programme (EMP) is a policy relevant description defining the target of a collection of observations and/or the deployment of Environmental Monitoring Facilities in the field. Usually an Environmental Monitoring Programme has a long-term perspective over at least a few years. An Environmental Monitoring Programme covers an area of interest (e.g. a region) and is based on environmental legislation*"[62]

(b) ***Environmental Monitoring Network (EMN)*** which is defined as the physical network of the observation locations (=environmental monitoring facilities) implemented according to the specification of the EMP;

(c) ***Environmental Monitoring Activity (EMA)*** which is defined as the specific observation campaign conducted on a number of observation facilities, and the

(d) ***Environmental Monitoring Facility (EMF)*** which is defined as the observation location (e.g. sensor or sensor cluster). The EMF can either be fixed, mobile or attached to another one. As environmental monitoring facilities can be described at various levels of detail, the model provides a recursive hierarchical link between Environmental Monitoring Facilities. This reflects the fact that a station can have various parts or a platform can host a number of sensors or measuring equipment.

The schema follows a generic approach which should enable thematic communities to use this structure across domains. The environmental monitoring facilities enable the linking to observations and

[61] See https://inspire.ec.europa.eu/Themes/120/2892

[62] See https://inspire.ec.europa.eu/id/document/tg/ef

measurements taken at a specific location and thus the provision of real time observation values using OGC SOS standard.

Examples: https://data.lter-europe.net/deims/node/8642/emf

Relevance of ECOPOTENTIAL: the protected areas and the related information on instrumentation and resulting data can be shared via DEIMS-SDR using the INSPIRE EF application schema (see Chapter 3.5.1). This enables the establishment of a linked site catalogue and access to datasets related to a single observation location. The observation facility can be the total or a part of the protected area. The protection status is one attribute stored together with the observation facility in DEIMS-SDR[63]

3.5 Sharing Information on Protected Areas Observations

In order share information on the observations from the different protected areas documented in DEIMS-SDR and allowing to develop applications on top of this data the INSPIRE Environmental Monitoring Facility application schema was chosen to generate a generic and harmonised export. In DEIMS-SDR information on the observation facility as well as on the resulting in-situ datasets is collected (see Poursanides et al. 2017).

The work was conducted in ECOPOTENTIAL in order serve a number of use cases in the harmonised data exchange missing in the analysed IT environments. The main use cases identified were:

- Sharing of site information (e.g. on protected areas) in a standardised format in order to generate the basis for the discovery and reuse in a distributed environment
- Providing standardised exchange of site boundaries for any observation and experimentation facilities including also non-protected areas
- Enabling the implementation of the INSPIRE regulation for any observation and experimentation facilities by implementing a full INSPIRE compliant EMF dataset record based on information provided in DEIMS-SDR.

For the implementation the following definitions were applied:

- **DEIMS-SDR SITE** is defined as any observation and experimentation facility (site) which characterised by an area where the long term observations or experiments are carried out. A site can be described by a series of datasets (characteristics) generated at this location and the linking to a monitoring network (e.g. LTER Europe) defining the context of the activities.
- **EXCHANGE FORMAT** is a syntactically and semantically given data specification to ensure interoperable exchange of information between information and communication systems (ICT) and to allow for a machine-to-machine communication.
- **INSPIRE ENVIRONMENTAL MONITORING FACILITIES (EF)** Location and operation of environmental monitoring facilities includes observation and measurement of emissions, of the state of environmental media and of other ecosystem parameters (biodiversity, ecological conditions of vegetation, etc.) by or on behalf of public authorities.

[63] See ECOPOTENTIAL Report D5.2 Metadata specification (Poursanides et al. 2017)

Within the ECOPOTENTIAL project the mapping of the site documentation to INSPIRE EF schema was conducted as well as an interactive map viewer to visualise the information on the sites were implemented. Both activities contribute to the development of DEIMS-SDR as central registry of observation facilities. The resulting software is open source and shared via github.

An example of the resulting INSPIRE EF recored for the Nationalpark Kalkalpen (AT, https://data.lter-europe.net/deims/site/49515dda-1198-4013-8f43-c33e107af081) can be accessed under https://data.lter-europe.net/deims/node/8642/emf. Annex 3 provides a full example for the implementation of iNSPIRE EF for a Protected Area.

3.5.1 Mapping of site information to INSPIRE EF

Information provided in DEIMS-SDR on the observation and experimentation facilities was mapped to the INSPIRE EF model distinguishing between the Environmental Monitoring Facility [EMF], the Environmental Monitoring Acitivity [EMA], and the Environmental Monitoring Network [EMN]. The relations are modelled using the patterns defined in the application schema.

Fig. 3.5 Linking of Observation Facility [EMF] to data product [EMA] and observation

Fig. 3.5 provides an overview of the relations between the different concepts implemented in the INSPIRE EF application schema. A detailed mapping is given in ***Annex 2*** *- Mapping of Site and Related Documentation to INSPIRE EMF Data Model*

3.5.1.1 Environmental Monitoring Facility

The **Environmental Montioring Facility [EMF]** is mapped to the specific **DEIMS-SDR:Site** (observation and experimentation facility) registered and documented in DEIMS-SDR. This includes all Protected Areas documented on DEIMS-SDR. The specific information on the site (e.g. parameters observed or location) is mapped to the different fields in the INSPIRE EF specification. A full documentation can be found in Annex 2.

Inter-alia, an important information about the sites is the linking to the observation which are carried out at the sites. This is done using the relations 'Observation capability' and 'Has Observation'.

- ***Observation capability*** linking to the observed parameters defined for the observation and experimentation facility. The INSPIRE optional element *observingCapability* is complex object where only the property ***observedProperty*** has been mapped to the Site parameter element[64], which is linked to the EnvThes. The remaining mandatory sub elements are rendered empty with the nilReason attribute with value missing as following example:

```
<ef:observingCapability>
	<ef:ObservingCapability gml:id="ObservingCapability_LTER_EU_AT_003_58775cf204d78">
		<ef:observingTime xsi:nil="true" nilReason="missing"/>
		<ef:processType xsi:nil="true" nilReason="missing"/>
		<ef:resultNature xsi:nil="true" nilReason="missing"/>
		<ef:procedure nilReason="missing"/>
		<ef:featureOfInterest nilReason="missing"/>
		<ef:observedProperty
		xlink:href="http://vocabs.ceh.ac.uk/evn/tbl/sparql?default-graph-uri=urn:x-
		evn-
		pub:EnvThes&format=text/json&query=SELECT%20%3Fresult%0AWHERE%20%7B%0A%09GR
		APH%20%3Curn%3Ax-evn-
		pub%3AEnvThes%3E%20%7B%0A%09%09%3Fresult%20a%20%3Chttp%3A%2F%2Fwww.w3.org%2
		F2004%2F02%2Fskos%2Fcore%23Concept%3E%20.%0A%20%20%20%20%20%20%20%20%20%20%
		20%20%20%20%20%20FILTER%20EXISTS%20%7B%0A%20%20%20%20%20%20%20%20%20%20%20%
		3Fresult%20%3FanyProperty%20%3FanyValue%20.%0A%20%20%20%20%20%20%20%20%20%2
		0%20FILTER%20(isLiteral(%3FanyValue)%20%26%26%20regex(LCASE(str(%3FanyValue
		))%2C%20%22(%3F%3D.*ecosystem
		measure*)%22))%20.%0A%20%20%20%20%20%20%20%7D%20.%0A%09%7D%0ABIND%20(%3Chtt
		p%3A%2F%2Fwww.w3.org%2F2004%2F02%2Fskos%2Fcore%23prefLabel%3E(%3Fresult)%20
		AS%20%3Flabel)%20.%0A%7D%20ORDER%20BY%20(LCASE(%3Flabel))"/>
	</ef:ObservingCapability>
</ef:observingCapability>
```

  The value of the gml identifier for observingCapability object is created as a concatenation of the string ObservingCapability, Site Code and PHP uniqid function, which generates a unique ID based on the microtime (current time in microseconds).

- ***Has observation*** linking to a single dataset or data service specified in the dataset metadata. The linking is done via the **Site_UUID** which referes to the DEIMS-SDR:Site. The INSPIRE element *hasObservations* (multiple) has been mapped to dataset metadata linked to the site and provides xlink to either ISO19139 XML representation generated by iso19139 DEIMS-SDR module or if provided in the dataset metadata it provides link to the online distribution link (e.g. OGC SOS service). Example of both ways is provided in the following XML code fragment:

```
<ef:hasObservation xlink:title="IT_SI001230_Lake Scuro_Chla_20130228"
xlink:href="http://sk.ise.cnr.it/observations/sos/kvp?service=SOS&version=2.0.0&request=Ge
tObservation&offering=offering:http://sp7.irea.cnr.it/sensors/sk.ise.cnr.it/procedure/noMa
nufacturerDeclared/noModelDeclared/noSerialNumberDeclared/20150428125100001/observations&o
bservedProperty=http://vocab.nerc.ac.uk/collection/P02/current/CPWC/&procedure=http://sp7.
irea.cnr.it/sensors/sk.ise.cnr.it/procedure/noManufacturerDeclared/noModelDeclared/noSeria
lNumberDeclared/20150428125100001&featureOfInterest=http://sp7.irea.cnr.it/sensors/sk.ise.
cnr.it/foi/SSF/SP/EPSG:4326/44.381805559999997/10.045638889999999&MergeObservationsIntoDat
aArray=trueDistribution"/>
<ef:hasObservation xlink:title="IT_SI001230_Lake Scuro_Conductivity_20130301"
xlink:href="http://bolegweb.geof.unizg.hr/deims/node/8713/iso19139"/>
```

[64] https://data.lter-europe.net/deims/documentation/site#R38

The first hasObservation element provides in xlink:href attribute a link to GetObservation request extracted from the dataset metadata. The second hasObservation element provides a link to iso19139 metadata representation because it does not provide any SOS link. The xlink:title attribute is populated by the dataset title provided by users in the DEIMS-SDR.

3.5.1.2 Environmental Monitoring Acitivity [EMA]

The **Environmental Monitoring Acitivity [EMA]** is mapped to a specific **DEIMS-SDR:DataProducts** collected within the observation and experimentation facilities (site). In extension to the original definition (see Poursanides et al. 2017) the data products can also be defined as monitoring (collection) activities or events which generate a series of datasets, e.g. Biodiversity observation or Soil water sampling.

The linking between the DEIMS-SDR:Site and the DEIMS-SDR:DataProduct is done using the element '*involvedIn*'. The INSPIRE optional multiple element *involvedIn* has been mapped to relevant data products content type fields linked to site via a view as in the following example:

```
<ef:involvedIn>
	<ef:EnvironmentalMonitoringActivity	gml:id="Data_collection_activity_8689b125-
	ee46-4d09-9e46-640f9c5c6eab_5877c2697a53f">
		<gml:description>
		landscape measure, elevation, land cover, land use, landscape cover
		</gml:description>
		<gml:identifier codeSpace="Airborne Images"/>
		<gml:name>LTER Zöbelboden Austria aerial data</gml:name>
		<ef:activityTime>
			<gml:TimePeriod
			gml:id="Data_collection_activity_TimePeriod_8689b125-ee46-4d09-9e46-
			640f9c5c6eab_5877c2697a580">
			<gml:beginPosition>1995-07-15</gml:beginPosition>
			<gml:endPosition>2016-07-15</gml:endPosition>
			</gml:TimePeriod>
		</ef:activityTime>
		<ef:activityConditions>
		Aerial photographs and LiDAR data of the LTER Zöbelboden catchment
		</ef:activityConditions>
		<ef:responsibleParty>
			<base2:RelatedParty>
			<base2:individualName>
			<gco:CharacterString>Johannes		Kobler,		Thomas
			Dirnboeck</gco:CharacterString>
			</base2:individualName>
			</base2:RelatedParty>
		</ef:responsibleParty>
		<ef:inspireId>
			<base:Identifier>
			<base:localId>8689b125-ee46-4d09-9e46-640f9c5c6eab</base:localId>
			<base:namespace>https://data.lter-europe.net/deims</base:namespace>
			</base:Identifier>
		</ef:inspireId>
		<ef:onlineResource>https://data.lter-
		europe.net/deims/node/10282</ef:onlineResource>
	</ef:EnvironmentalMonitoringActivity>
</ef:involvedIn>
```

The metadata available in the product data model are aggregated into the EnvironmentalMonitoringActivity complex type. It's gml identifier is composed using standard pattern ContentType_UUID_uniqid() as in the example above: `Data_collection_activity_TimePeriod_8689b125-ee46-4d09-9e46-640f9c5c6eab_5877c2697a580`. Data product title[65] is placed into the gml:name element; Data product type[66] in gml:identifier element attribute codeSpace; Data product abstract in the ef:activityConditions, parameters are listed in gml:description, Data product temporal extent in ef:activityTime and finally Contact information simplified to full names are aggregated in ef:responsibleParty.

3.5.1.3 Environmental Monitoring Network [EMN]

The relation to a **Environmental Monitoring Network [EMN]** is mapped to a specific **DEIMS-SDR:Network** to which a DEIMS-SDR:Site belongs to. This can either be a long term monitoring network like LTER Europe or a project like ECOPOTENTIAL. The linking between the DEIMS-SDR:Site and the DEIMS-SDR:Network is done using the INSPIRE element '*belongsTo*'. *The* INSPIRE optional multiple element *belongsTo* has been mapped to LTER National Network[67] and Networks (in addition to LTER)[68] elements as follows:

```
<ef:belongsTo xlink:href="https://data.lter-europe.net/deims/node/222">
    <ef:NetworkFacility gml:id="Network_e3d5faf7-5603-4acd-b835-eaedca8e5dec">
        <gml:name>Austria (LTER-Austria)</gml:name>
        <ef:linkingTime>
            <gml:TimePeriod          gml:id="timePeriod_e3d5faf7-5603-4acd-b835-
            eaedca8e5dec">
            <gml:beginPosition>2012-10-22</gml:beginPosition>
            <gml:endPosition indeterminatePosition="now"/>
            </gml:TimePeriod>
        </ef:linkingTime>
        <ef:belongsTo/>
            <ef:contains/>
    </ef:NetworkFacility>
</ef:belongsTo>
<ef:belongsTo>
    <ef:NetworkFacility gml:id="Network_b4f312d0-ad5d-4421-9876-3403329d2132">
        <gml:name>UNECE ICP Integrated Monitoring</gml:name>
        <ef:linkingTime>
            <gml:TimePeriod          gml:id="timePeriod_b4f312d0-ad5d-4421-9876-
            3403329d21321">
            <gml:beginPosition>2012-10-22</gml:beginPosition>
            <gml:endPosition indeterminatePosition="now"/>
            </gml:TimePeriod>
        </ef:linkingTime>
        <ef:belongsTo/>
        <ef:contains/>
    </ef:NetworkFacility>
</ef:belongsTo>
```

[65] https://data.lter-europe.net/deims/documentation/activity#DP1

[66] https://data.lter-europe.net/deims/documentation/activity#DP3

[67] https://data.lter-europe.net/deims/documentation/site#R30

[68] https://data.lter-europe.net/deims/documentation/site#R32

The first section appearance of ef:belongsTo element always renders data from linked LTER National Network, if any, with xlink:href value linking to the node of that network, gml:id containing the network node UUID, gml:name network node title and ef:linkingTime is for gml:beginPosition identified as the date of creation for network node and gml:endPosition as value now. The above mentioned applies also for other networks (e.g. project affiliations), but it has no link to any node, since it is a list of networks without any further information available in DEIMS-SDR.

3.5.1.4 Source code

The application is based on Drupal queries providing the necessary data. The software is open source and can be downloaded from https://github.com/klimeto/deims_emf_module. Within the H2020 project eLTER the application developed in ECOPOTENTIAL was reused and extended by creating automatically creating descriptive metadata for the INSPIRE EF data exchange format using ISO19139 standard.

3.5.2 Visualisation of site information

The EF Viewer allows to access the GeoServer[69] instance of DEIMS-SDR and to visualise the location of each site on a map. GeoServer is an OGC (Open Geospatial Consortium) compliant implementation of a number of open standards such as Web Feature Service (WFS), Web Map Service (WMS), and Web Coverage Service (WCS). The EF viewer utilises both the WMS and the WFS of GeoServer.

It is designed to easily access any GeoServer and can be configured through a settings file (settings.js) that stores the necessary information (URL of GeoServer instance, name of workspace, style settings such as the map centre or colours for the points to be visualised).

The viewer uses the GeoServer getcapabilities link to fetch all available layers and print them to the “Layers” menu, thus allow a user to switch between any given layer.

The same layers are also dynamically loaded to the “Export” menu where they can be exported as a Shapefile [.shp].

[69] http://geoserver.org/

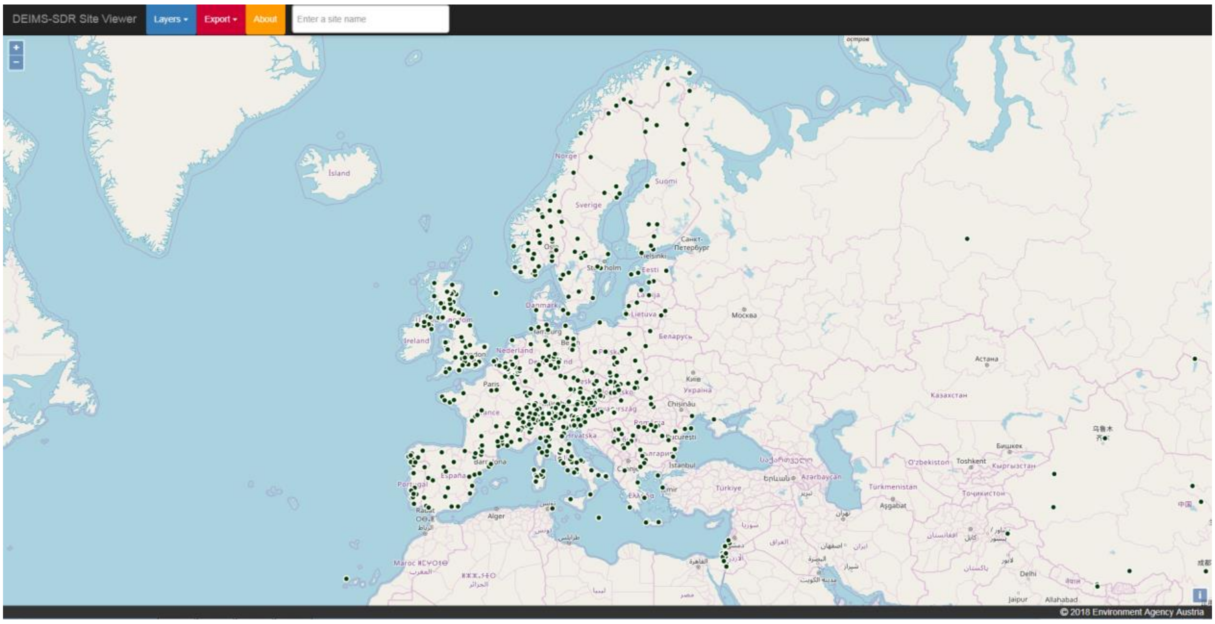


Fig. 3.6 EF Viewer Interface

The interface also includes a search box where sites can be listed by their site name. If a name is typed in and selected the map will pan to the site location and display a pop up.

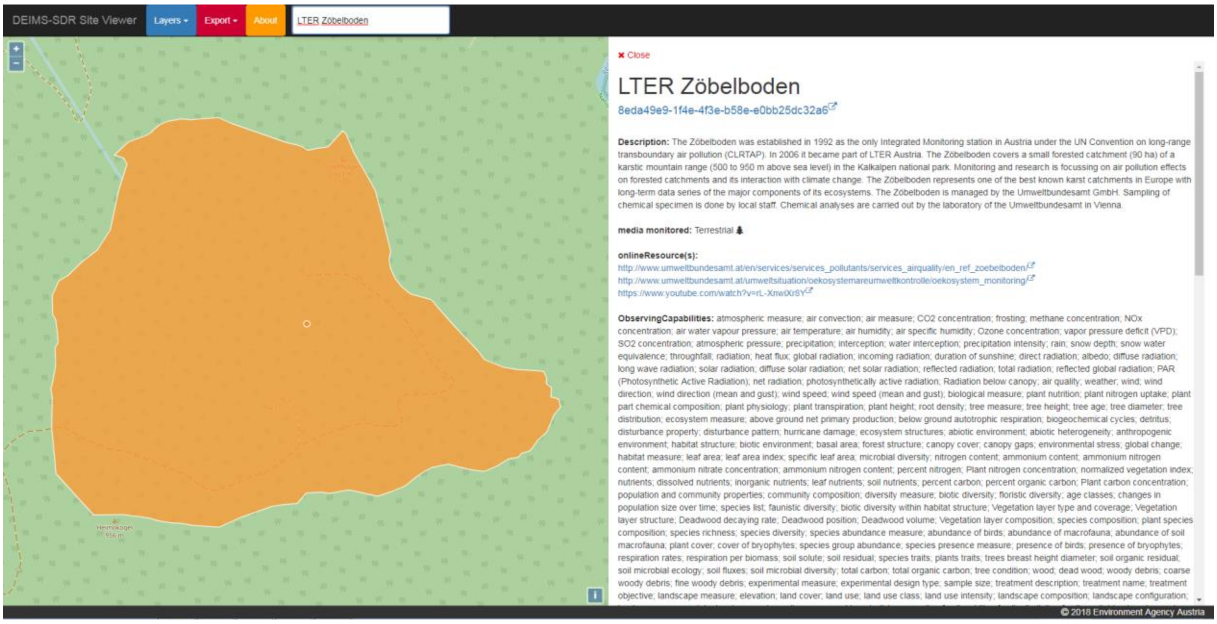


Fig. 3.7 EF Viewer Detailed information about a site

Alternatively a site can also be selected by clicking one of the dots on the map, which also causes a popup to appear that provides the URL to the site records on DEIMS-SDR (“View records on DEIMS-SDR”), the link to

Co-funded by the European Union

the metadata record as Inspire EF[70] ("Download metadata record") and a link to display more information ("Show more details ..."). When the latter link is clicked the corresponding metadata record is parsed and the information is displayed in a sidebar.

The information detail level depends on the information provided on DEIMS-SDR and is limited to the extent of Inspire EF. It includes:

- a resolvable identifier that refers to the DEIMS-SDR landing page
- an abstract
- the media monitored
    - Terrestrial
    - Marine
    - Fresh water rivers
    - Fresh water Lakes
- onlineResources listing url associated with the site
- ObservingCapabilities
- Activities
- hasObservations linking to datasets in ISO19139
- responsibleParties
- belongsTo describing project or network affiliation

With the interactive map viewer an easy tool to be integrated in any website was created allowing to link information from observation sites and the resulting data.

3.5.2.1 Source code

The source code is open source and available under via https://github.com/stopopol/ef_viewer. The reference implementation of the service is deployed under: https://data.lter-europe.net/map/.

[70] http://inspire.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_EF_v3.0.pdf

4 Summary

In order to allow automisation of workflows the standardisation of data and data flows is needed. This implies not only structural (syntactic) interoperability of the data used, but also semantic and moreover in the long term legal interoperability. Despite many community efforts a unique standard or format across different environmental domains is still missing. On European scale with the INSPIRE directive an important initiative in the harmonisation of environmental data from public administration was started in 2007. With a great effort common data schemas were developed which can be used beyond INSPIRE. These specifications are mainly tackling with spatial data but also allow the linking to observation data via the Environmental Monitoring Facility. The underlying data model for these in-situ data is based on the OGC Observation and Measurements (O&M) which is proving to be a common and unifying data model for any environmental observation. The OGC standard on Observation and Measurements[71] can be seen as attempt to unify and harmonise data collected from field campaigns. The work on O&M arises from work originally undertaken through the Open Geospatial Consortium's Sensor Web Enablement (SWE) activity. SWE is concerned with establishing interfaces and protocols that will enable a "Sensor Web" through which applications and services will be able to access sensors of all types, and observations generated by them, over the Web. The model encompasses the basic elements which defines any observation characterised by the observation property (phenomenon) the observation location (feature of interest) and the measurement itself (see Fig. 4.1).

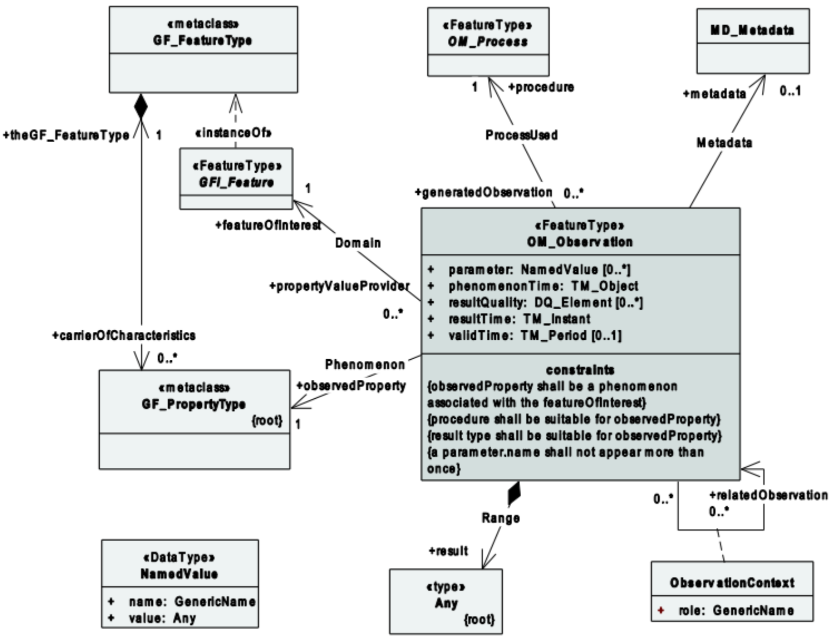


Fig. 4.1 Basic observation type according to OGC Observation and Measurements[72]

---

[71] See http://www.opengeospatial.org/standards/om

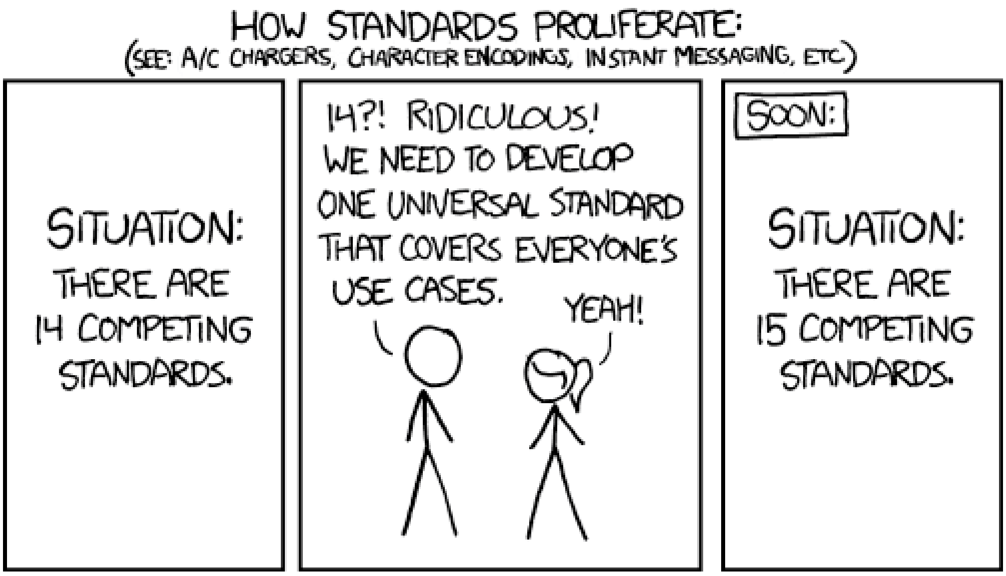
[72] See file:///C:/Users/JPeterseil/Downloads/10-004r3_Topic_20_Observations_and_Measurements.pdf

Due to the scope of observations in the ecosystem domain based on the research topic temporal, spatial and thematic resolutions might vary quite significantly. Depending on the usage of data harmonisation efforts in terms of temporal and thematic aggregation can be done. These aspects were not tackled by the current task, but will be an outcome of the modelling WPs collected in the last phase of the project. The current work focused on the evaluation of relevant controlled vocabularies and the data models for data coming from the different PAs and research projects.

Therefore the basic idea of the O&M data model was implemented on different scales in order to provide the basic steps for data harmonisation. The basic structure of the common controlled vocabulary EnvThes further developed in ECOPOTENTIAL was following this basic distinction. This in the long will enable the harmonisation of parameter names is extended to ontology build on top of the common vocabulary.

Also with the extension of DarwinCore with the EventCore and the related MeasurementOfFact (see e.g. DePooter et al. 2017) a step in a similar direction was done. The basic set of data items provided by the eLTER data reporting format can easily mapped to the common O&M model.

As outlined in the current report, a range of community standards is already provided by different communities. In order to foster data mobilisation and sharing these data formats should be applied for the data provisioning. The documentation of the data with high quality metadata including e.g. also the documentation of the data structure and reference lists used is an important step into the right direction.



© https://xkcd.com/927/

5 References

Alonso, J., Castro, P., Martins, I. , Honrado, J.P., Calafate, M., Gonçalves, J. , Peterseil, J., Monteiro, A., Santos, C., Arenas-Castro, S., Marcos, B., Poursanidis, D., Prados, J., Pratola, Ch. (2017). D5.3 Framework for user-oriented quality evaluation routines. ECOPOTENTIAL Project (H2020 GANr. 641762) Deliverable. [online http://www.ECOPOTENTIAL-project.eu/images/ECOPOTENTIAL/documents/D5.3.pdf]. 156pp.

d'Aquin, M., Noy, N. F. (2012) "Where to publish and find ontologies? A survey of ontology libraries," Web Semantics: Science, Services and Agents on the World Wide Web, vol. 11, pp. 96–111, Mar. 2012.

Davenport, T.H.; Prusak, L. (1997) Working Knowledge : How Organizations Manage What They Know, 1997.

De Pooter D, Appeltans W, Bailly N, Bristol S, Deneudt K, Eliezer M, Fujioka E, Giorgetti A, Goldstein P, Lewis M, Lipizer M, Mackay K, Marin M, Moncoiffé G, Nikolopoulou S, Provoost P, Rauch S, Roubicek A, Torres C, van de Putte A, Vandepitte L, Vanhoorne B, Vinci M, Wambiji N, Watts D, Klein Salas E, Hernandez F (2017) Toward a new data standard for combined marine biological and environmental datasets - expanding OBIS beyond species occurrences. Biodiversity Data Journal 5: e10989. https://doi.org/10.3897/BDJ.5.e10989

Esser, M. (2015): Semantic enrichment of terminology. Bachelor thesis, University of Applied Sciences, Potsdam.

Fiore, N., Magagna, B., Goldfarb, D. (2017) EcoPortal: Proposition for a Semantic Repository Dedicated to Ecology and Biodiversity. In: Alsayed Algergawy, Naouel Karam, Friederike Klan and Clement Jonquet (Eds) Proceedings of the 2nd International Workshop on Semantics for Biodiversity, co-located with 16th International Semantic Web Conference {(ISWC} 2017), Vienna, Austria, October 22nd, 2017.}, online http://ceur-ws.org/Vol-1933/poster-paper-13.pdf

Frege, G. (1967) Begriffsschrift (1879); English Translation in van Heijenoort, J.(ed.): From Frege to Gödel, pp. 1–82. Harvard University Press, Cambridge (1967)

Goldfarb, D., Le Franc, Y. (2017). Enhancing the Discoverability and Interoperability of Multi-Disciplinary Semantic Repositories. In: A. Algergawy, N. Karam, F. Klan, C. Jonquet: Proc. of the 2nd International Workshop on Semantics for Biodiversity co-located with 16th International Semantic Web Conference (ISWC 2017), Vienna, Austria, October 22nd, 2017. CEUR Workshop Proceedings 1933

Guralnick, R., Walls, R. & Jetz, W. (2017) Humboldt Core – toward a standardized capture of biological inventories for biodiversity monitoring, modeling and assessment. Ecography 40:001–012.

Harpring, P. (2010). Introduction to controlled vocabularies: terminology for art, architecture, and other cultural works. Getty Publications, 2010.

Kissling, W.D., Hardisty, A., García, E.A., Santamaria, M., De Leo, F., Pesole, G., Freyhof, J., Manset, D., Wissel, S., Konijn, J., Los, W. (2015) Towards global interoperability for supporting biodiversity research on essential biodiversity variables (EBVs), Biodiversity, 16:2-3, 99-107, DOI: 10.1080/14888386.2015.1068709.

Laport, M.-A., Mougenot, I., Garnier, E. (2012): ThesauForm-Traits: A web based collaborative tool to develop a thesaurus for plant functional diversity research. Ecological Informatics, vol.11, September 2012, Pages 34-44, DOI:https://doi.org/10.1016/j.ecoinf.2012.04.004

Mader, C., Haselhofer, B., Isaac, A. (2012): Finding quality issues in SKOS vocabularies, International Conference on Theory and Practice of Digital Libraries, p 222-233, Springer Verlag.

Michener, W.K., Brunt, J.W., Helly, J.J., Kirchner, Th.B., Stafford, S.G. (1997) Nongeospatial metadata for the Ecological Sciences. Ecological Applications 7:330-342

Ogden, C.K., Richards, I.A. (1923) The meaning of meaning. Harvest Book. Harcourt, Brace & World, Inc., New York. 363pp.

Oggioni, A., Carrara, P., Kliment, T., Peterseil, J., Schentz, H. (2012). Monitoring of Environmental Status through Long Term Series: Data Management System in the EnvEurope Project. EnviroInfo 2012, Shaker Verlag, Aachen.

Orbst, L (2006): The ontology spectrum and semantic models. Online presentation. Information Semantics Group Information Discovery & Understanding, Center for Innovative Computing & Informatics.

Petersen, M., Glöckler, F., Kiessling, W., Doring, M., Fichtmuller, D., Laphakorn, L., Baltruschat, B., Hoffmann, J. (2018) History and development of ABCDEFG: a data standard for geosciences. Fossil Record. DOI: 10.5194/fr-21-47-2018.

Pickert, G. (2011) „Einführung in Ontologien", Humboldt-Universität Berlin, Feb. 2011. http://www.dbis.informatik.huberlin.de/dbisold/lehre/WS0203/SemWeb/artikel/2/Pickert_Ontologien_final.pdf

Pickert, G. (2011): Einführung in Ontologien, Humbold-Universität Berlin, 2011.

Porter, J.H. (2010) A controlled vocabulary for LTER datasets. (2010) (link: http://databits.lternet.edu/spring-2010/controlled-vocabulary-lter-datasets)

Poursanidis, D., Peterseil, J., Wohner, C. , Chrysoulakis, N., Wetzel, F. , Alonso, J. , Castro, P. , Beierkuhnlein, C. , Bernd, A. , Zabala, A. , Masó, J. , Domingo, C. , Vetaas, O., Bargmann, T. , Bosch, S. (2017). D5.2 Metadata for pre-existing datasets. ECOPOTENTIAL Project (H2020 GANr. 641762) Deliverable. [online http://www.ECOPOTENTIAL-project.eu/images/ECOPOTENTIAL/documents/D5.2.pdf] 116pp.

Saarenmaa, H., Camacho, A.G., Ó Tuama, É., Agosti, D., Bailly, N., Calderesi, M., Drucker, D., Güntsch, A., Jacobsen, K., Kempa, M., Köljag, U., Mergen, P., Peer, I., Penev, L., Valland, N., de Wever, A. (2014) Deliverable D2.1 Architectural design, review and guidelines for using standards. Project report EU-BON (GANr. 308454). Deliverable [online http://www.eubon.eu/getatt.php?filename=EU%20BON_D2.1_Architectural%20design%20review%20and%20guidelines%20for%20using%20standards_10716.pdf]. 88 pp.

Schentz, H., Peterseil, J. & Bertrand, N. (2013) EnvThes- interlinked thesaurus for long term ecological research, monitoring, and experiments. Proceedings EnviroInfo 2013: Environmental Informatics and Renewable Energies. Shaker Verlag, Aachen.

Schentz, H., Peterseil, J., Magagna, B. & Mirtl, M. (2011) Semantics in Ecosystem Research and Monitoring. Proceedings EnviroInfo 2011: Innovations in Sharing Environmental Observation and Information. Shaker Verlag, Aachen.

Sowa J.F. (2000a) Ontology, Metadata, and Semiotics. In: Ganter B., Mineau G.W. (eds) Conceptual Structures: Logical, Linguistic, and Computational Issues. ICCS-ConceptStruct 2000. Lecture Notes in Computer Science, vol 1867. Springer, Berlin, Heidelberg

Sowa, J. F. (2000b). Knowledge Representation: Logical, philosophical, and computational foundations. Brooks/Cole, Pacific Grove, 2000.

Studer, R. , Oppermann, H., Schnurr, H.-P. (2011). "Die Bedeutung von Ontologien für das Wissensmanagement," Institut AIFB, Universität Karlsruhe, 2001.

Studer, R., Oppermann, H., Schnurr, H.-P. (2001) Die Bedeutung von Ontologien für das Wissensmanagement. Institut AIFB, Universität Karlsruhe (2001)

Turnitsa, C. and Tolk, A. (2006). Battle Management Language: A Triangle with Five Sides Proceedings of the Simulation Interoperability Standards Organization (SISO) Spring Simulation Interoperability Workshop (SIW), Huntsville, AL, April 2-7, 2006

Vanderbilt, K., and E. Gaiser. (2017). The International Long Term Ecological Research Network: a platform for collaboration. Ecosphere 8(2):e01697. 10.1002/ecs2.1697

Vanderbilt, K., John H. Porter, Sheng-Shan Lu, Nic Bertrand, David Blankman, Xuebing Guo, Honglin He, Don Henshaw, Karpjoo Jeong, Eun-Shik Kim, Chau-Chin Lin, Margaret O'Brien, Takeshi Osawa, Éamonn Ó Tuama, Wen Su, Haibo Yang (2017) A prototype system for multilingual data discovery of International Long-Term Ecological Research (ILTER) Network data. Ecological Informatics 40:93-101, http://dx.doi.org/10.1016/j.ecoinf.2016.11.011.

Vanderbilt, K., Lin, Ch.Ch., Lu, Sh-Sh., Kassim, A.R., He., H., Guo, X., San Gil, I., Blankman, D. & Porter, J. (2015) Forstering ecological data sharing: collaborations in the international Long Term Ecological Research Network. Ecosphere(10) Article 204 18pp.

Vanderbilt, K.L., Blankman, D., Guo, X., He, H., Lin. Ch-Ch., Lu, S.-S., Ogawa, A., Ó Tuama, É., Schentz, H., Su, W. (2010) A multilingual metadata catalog for the ILTER: Issues and approaches, Ecological Informatics 5:187-193.

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)

Co-funded by the European Union

6 Annex

6.1 Annex 1 – Analysis of data formats for different data products

Tab. 6.1 Most common data formats for documented data products (n=103; Source: DEIMS-SDR)[73]

LEVEL_1	LEVEL_2	n	Table	Vector	Raster
Topography	Elevation model	1			x
Topography	Topography	2		x	x
Meteorology		6	x		
Meteorology	Air humidity	1	x		
Meteorology	Air pressure	1	x		
Meteorology	Air temperature	2	x		
Meteorology	Precipitation	1	x		
Meteorology	Radiation	1	x		
Meteorology	Wind speed and direction	1	x		
Deposition	Precipitation chemistry	1	x		
Geology and Geomorphology		2		x	x
Hydrology and water budget		2	x		
Hydrology and water budget	Groundwater	1	x	x	x
Hydrology and water budget	Runoff and discharge	2	x		
Hydrology and water budget	Snow cover and duration	1	x	x	
Hydrology and water budget	Water use	1	x	x	
Land Cover and habitats	Current land cover and habitats	5		x	x
Land Cover and habitats	Historic land cover and habitats	3		x	x
Land Cover and habitats	Land Cover and habitats	1	x	x	x
Terrestrial systems characteristics		2	x		
Terrestrial systems characteristics	Biodiversity and species richness (incl. population)	4	x		
Terrestrial systems characteristics	Ecosystem biomass and structure	2	x		
Terrestrial systems characteristics	Phenology	2	x	x	x
Terrestrial systems characteristics	Plant and foliage chemistry	2	x		
Terrestrial systems characteristics	Soil	6	x	x	x
Terrestrial systems characteristics	Vegetation structure and species cover	9	x		
Aquatic systems characteristics		1	x		

[73] Updated 13.05.2018

LEVEL_1	LEVEL_2	n	Table	Vector	Raster
Aquatic systems characteristics	Running water (River and streams)	4	x	x	
Aquatic systems characteristics	Standing water (Lakes)	3	x	x	x
Aquatic systems characteristics	Wetlands	3	x	x	x
Social systems characteristics		1	x	x	x
Social systems characteristics	Economic data	1	x		
Social systems characteristics	Ecosystem services	3		x	x
Social systems characteristics	Ecosystem use and management	2	x		
Disturbance and disaster events		2	x	x	x
Disturbance and disaster events	Fire	1	x	x	x
Disturbance and disaster events	Infestation	2	x	x	x
Disturbance and disaster events	Other disturbance	1	x	x	x
Nature conservation and management		1	x	x	
Nature conservation and management	Focal species	6	x		
Nature conservation and management	Governance and planning	2		x	
Remote sensing data products		1		x	x
Remote sensing data products	Derived EO Data products	4		x	x
Remote sensing data products	Original images	3			x
TOTAL		**103**			

6.2 Annex 2 - Mapping of Site and Related Documentation to INSPIRE EMF Data Model

Author: Tomáš Kliment[74] with contributions from Alessandro Oggioni[75], Christoph Wohner[76] & Johannes Peterseil[3]

Version: 1.1 Draft 2017-01-13

Last edit: 2017-01-13

Document history

2016-10-27	Peterseil, Johannes	Compilation of the existing material and preparation of the first version of the document
2017-01-11	Kliment, Tomas	Reporting the job done to implement the EMF XML export module for the DEIMS-SDR system
2017-01-13	Peterseil, Johannes Christoph Wohner	Final comments

6.2.1 Definition of terms

DEIMS-SDR SITE A site is defined as an area where the long term observations or experiments are carried out. A site can be described with a set of data (characteristics) defined by networks (e.g. LTER Europe).

EXCHANGE FORMAT Syntactically and semantically given data specification to ensure interoperable exchange of information between information and communication systems (ICT) and to allow for a machine-to-machine communication.

INSPIRE ENVIRONMENTAL MONITORING FACILITIES Location and operation of environmental monitoring facilities includes observation and measurement of emissions, of the state of environmental media and of other ecosystem parameters (biodiversity, ecological conditions of vegetation, etc.) by or on behalf of public authorities.

[74] MK18, s.r.o., Slovakia

[75] CNR, Italy

[76] Umweltbundesamt GmbH, Austria

6.2.2 Objective

The document tries to summarise the status of the work done with regard to the mapping and exchange of site information based on the concepts used in the OGC Environmental Monitoring Facility specification. The work is done within the H2020 projects ECOPOTENTIAL (contract Tomas Kliment), including requirements from eLTER and ILTER/LTER Europe.

6.2.3 Use cases

The use cases described in the following section should be supported by the current developments. The aim of the work is to define DEIMS-SDR Site Exchange Model (SEM) and its technical implementation.

Tab. 6.2. Overview on use cases

ID	Use case name	PRIO
UC 1	Sharing of site information for distributed discovery and reuse	1
UC 2	Providing standardised exchange of site boundaries	1
UC 3	Import of external input for site information	2
UC 4	Provide full INSPIRE compliant EMF dataset record	3

6.2.3.1 UC 1 Sharing site information for distributed discovery

UC Element	Description
UCID	1
UC Name	Sharing site information for distributed discovery and reuse
UC Description	Mapping of local site and other data models MD fields (based on DEIMS-SDR SMM 1.1) on a common site exchange model (SEM) and exchange the information as eXtensible Markup Language (XML) files between different systems. The exchange model should include all information specified in the SMM and should follow, where possible, European standard exchange models (e.g. INSPIRE EF Specification). The model also should allow the import of site information to DEIMS-SDR from any external sources (e.g. excel)
Application	DEIMS-SDR EMF XML Module
Primary Actor	DEIMS
Precondition	Site information needs to be provided in DEIMS-SDR together with its spatial representation in form of a centroid / bounding box coordinates.
Trigger	On demand by the user, or by scheduler (e.g. a eLTER integration portal harvester)

Basic Flow	1. User describes site based on standard attributes (or updates existing information). 2. XML document is created on the fly dynamically triggered by an user demand based on mapping implemented in the EMF XML module. 3. The batch harvesting of EMF XML site representations is executed using the harvest list providing the basic metadata as record identifier, update information and uniform resource locator pointing to an XML file implemented as a view in DEIMS.
Alternate Flows	-

6.2.3.2 UC 2 Providing standardised exchange format for research site boundaries

UC Element	Description
UCID	2
UC Name	Provide standard exchange format for research site boundaries
UC Description	The research site boundaries and basic information, based on the information provided in DEIMS, are exposed in standardized geographic data model using GML geometry objects included in the EMF. By this datasets can be used in any external web or desktop GIS applications (e.g. Openlayers, QGIS). The multi geometry objects should include a) Centroid coordinates b) Bounding box c) Site boundaries
Application	DEIMS-SDR
Primary Actor	DEIMS-SDR
Precondition	Geographic information needs to be provided in DEIMS-SDR within the editor form for Site Metadata Model[77] by placing the centroid coordinate and drawing the bounding box or more exact site boundaries delineating the spatial extent.
Trigger	On demand by the user
Basic Flow	1. User describes site's geospatial information using geographic module of the linked research site editing form. 2. XML document created on the fly provides standard GML encoding of the coordinates in the standard geometry element defined in EMF model. 3. Site data in XML are harvested / downloaded and used in remote systems to visualize.
Alternate Flows	Harvested site information are shared using INSPIRE download service based on Web Feature Service (WFS) to provide an API for filtering and further data preparation and manipulation (e.g. StoredQueries).

6.2.3.3 UC 3 Import of external input for site information

UC Element	Description

[77] https://data.lter-europe.net/deims/documentation/site

UCID	3
UC Name	Import of external inputfor site information
UC Description	Extension of UC 1 but with direction to import information into the DEIMS-SDR system based on the site exchange model specification
Application	DEIMS-SDR
Primary Actor	Site manager
Precondition	-
Trigger	Creation or update of site information
Basic Flow	1. Site manager provides new or updated site information in an external data source (e.g. excel) 2. External conversion to SEM (site exchange model XML) 3. Import of SEM to DEIMS-SDR 3.1 update of existing site record (identified based on UUID) 3.2 create new site entry in DEIMS-SDR
Alternate Flows	-

6.2.3.4 UC 4 Provide full INSPIRE compliant EF dataset record

UC Element	Description
UCID	4
UC Name	Provide full INSPIRE compliant EF dataset record
UC Description	Provide a full INSPIRE compliant EF dataset record based on the site information provided in DEIMS-SDR. These standard spatial dataset can be accessed by any GIS client. This includes the mapping to complex feature types.
Application	DEIMS-SDR (providing site information and boundaries) Transformation ETL from site to EF Geoserver (OGC WxS)
Primary Actor	DEIMS-SDR
Precondition	Mapping of all SMM fields to EF fields, open issue is what to do with the extensions
Trigger	On demand by users (e.g. GIS client)

Basic Flow	1. Mapping of SMM fields to EF fields 2. Creation of GML (e.g. by geoserver) 3. Expose as OGC WxS
Alternate Flows	Predefined dataset (GML)

6.2.4 MATERIALS

- DEIMS-SDR Site Metadata Model description https://data.lter-europe.net/deims/content/Site-Metadata-Model-version-11
- DEIMS-SDR Data Product Metadata Model description https://data.lter-europe.net/deims/documentation/activity
- DEIMS-SDR Dataset Metadata Model description https://data.lter-europe.net/deims/documentation/dataset
- DEIMS-SDR Dataset Person Model description https://data.lter-europe.net/deims/documentation/person
- DEIMS-SDR Dataset Network Model description https://data.lter-europe.net/deims/documentation/network
- EF Mapping material including examples https://drive.google.com/drive/folders/0Bx5B7ytC5F2ZUmdtQUItU0xMYk0
- INSPIRE Environmental Monitoring Facilities Data specification http://inspire.ec.europa.eu/Themes/120/2892
- INSPIRE Consolidated UML Model http://inspire.ec.europa.eu/data-model/approved/r4618-ir/html/index.htm?goto=2:3:6:1:1:7980
- INSPIRE registry http://inspire.ec.europa.eu/registry

6.2.5 METHODOLOGY

6.2.5.1 DEIMS-SDR Site Metadata Model

The current documentation of the site metadata model for ILTER/LTER was taken for the mapping exercise. The full documentation can be found at https://data.lter-europe.net/deims/content/Site-Metadata-Model-version-11. The metadata on the sites (including the boundaries) are stored within the DEIMS Site and Dataset Registry (https://data.lter-europe.net/deims/).

6.2.5.2 Mapping of the Site documentation MD fields to EF

Author: Alessandro Oggioni (CNR)

In a common folder the mapping tables as well as examples are provided (https://drive.google.com/drive/folders/0Bx5B7ytC5F2ZUmdtQUItU0xMYk0)

Mapping table (EMF_Mapping Table_2003 - https://docs.google.com/spreadsheets/d/1cdmUC2x5FUPFs4Oglj-E3dcwH8_bfq0_Q2dXoHabwq8/edit#gid=1256544976) is using the last edition of EMF Schema (v. 4.0). It was difficult especially for understand the different relations among different type of EF classes. I added in the table 3 columns:

- column I is dedicated to the sub element of the XML element mentioned in the column C;
- column J mentions a path in DEIMS-SDR site or person forms;
- column k reports my personal comment.

Additional information and examples are provided in the folder (https://drive.google.com/drive/folders/0Bx5B7ytC5F2ZUmdtQUItU0xMYk0):

- pictures (EML_LTER) to represent the relations among different Environmental Monitoring Networks (EMN) and different Environmental Monitoring Facilities (EMF);
- XML EF examples of EMN for LTER Eu network and for LTER-Italy network;
- XML EF examples of EMF for LTER parent site and LTER research site;
- a presentation by INSPIRE group with some examples use case.

6.2.5.3 Exposing Site Documentation using XML encoding - proposed options

Author: Tomas Kliment (MK18)

Based on available technologies the following use case implementation options were proposed:

Option 1. Native Drupal XML data export

We can use the native Drupal XML data export to encode the site info in an “unstandardized” XML as the following example (LTER_EU_IT_044):

Schemata: Basic fields (where possible) is mapped to EF compliant fields, but the model is extended to an DEIMS-SDR native Site Exchange Model (SEM). This allows for a fast and pragmatic mapping of all information in the DEIMS-SDR Site Metadata model (SMM) as defined in DEIMS-SDR.

Example XML code snippet:

```
<?xml version="1.0" encoding="UTF-8"?>
<deimsSites>
        <site>
                <siteLocalId>2e2aba16-6ee9-4508-a43f-343d9dbf514f</siteLocalId>
```

```
                <namespace>data.lter-europe.net/deims</namespace>
                <siteName>Lago di Garda - Italy</siteName>
                <siteCode>LTER_EU_IT_044</siteCode>
                <siteManager>Nico SalmasoNico Salmaso</siteManager>
                <geometry>POINT (10.6205 45.5806)</geometry>
                <broader/>
                <involvedIn>Formal LTER/LTSER</involvedIn>
                <belongsTo>Italy (LTER-ITalia)</belongsTo>
        </site>
</deimsSites>
```

Evaluation:

+ Easy and fast to implement (based on Drupal Views and XML preview)

- No "standardized" data model and service API available

Implementation options

a) Drupal views and XML export (like in DEIMS1)
b) Drupal Module XML Export (like done for the EML/ISO export)

Option 2. GeoServer WFS API

We can use the GeoServer WFS API connected directly to MySQL database, which is underneath the DEIMS-SDR creating a data store by using the MySQL plugin for GeoServer.

Schemata: Basic fields (where possible) is mapped to EF compliant fields, but the model is extended to a DEIMS-SDR native Site Exchange Model (SEM). This allows for a fast and pragmatic mapping of all information in the DEIMS-SDR Site Metadata model (SMM) as defined in DEIMS-SDR.

Example XML code snippet:

```
<?xml version="1.0" encoding="UTF-8"?>
<wfs:FeatureCollection xsi:schemaLocation="http://www.opengis.net/wfs/2.0
http://schemas.opengis.net/wfs/2.0/wfs.xsd http://www.opengis.net/gml/3.2
http://schemas.opengis.net/gml/3.2.1/gml.xsd http://data.lter-europe.net/
http://klimeto.com:80/geoserver/deims/wfs?service=WFS&version=2.0.0&request=Descr
ibeFeatureType&typeName=deims%3AResearchSites" timeStamp="2016-10-
28T06:12:55.948Z" numberReturned="1" numberMatched="1"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:deims="http://data.lter-europe.net/"
xmlns:gml="http://www.opengis.net/gml/3.2"
xmlns:wfs="http://www.opengis.net/wfs/2.0"
xmlns:xs="http://www.w3.org/2001/XMLSchema">
  <wfs:member>
    <deims:ResearchSites gml:id="ResearchSites.2e2aba16-6ee9-4508-a43f-
            343d9dbf514f"><deims:SITE_NAME>Lago di Garda</deims:SITE_NAME>
```

```
        <deims:SHORT_NAME>IT08-005-A</deims:SHORT_NAME>
        <deims:SITE_CODE>LTER_EU_IT_044</deims:SITE_CODE>
        <deims:SITE_SIZE>36800.0</deims:SITE_SIZE>
        <deims:SITE_DESCRIPTION>Lake Garda has the largest volume (49 billions of
             cubic meters) and the most extensive area (368 km2) of Italian lakes.
             Along with lakes Orta, Maggiore, Lugano, Como, Iseo and Idro, it forms
             part of the group of deep lakes located south of the Alps in one of
             the most densely populated and highly productive area of Italy. With
             a total volume of over 124 billions of cubic meters, these lakes
             constitute one of the largest freshwater supplies in Europe. Their
             waters are used for ……………………………………………………………………..
        </deims:SITE_DESCRIPTION>
        <deims:PURPOSE>Long Term Ecological Research. Basic and applied research.
             Impact of climatic and anthropic stressors.</deims:PURPOSE>
        <deims:geom>
        <gml:Point srsDimension="2" srsName="urn:ogc:def:crs:EPSG::4326">
        <gml:pos>45.5806 10.6205</gml:pos
        ></gml:Point></deims:geom>
      </deims:ResearchSites>
    </wfs:member>
  </wfs:FeatureCollection>
```


Evaluation:

+ Provides service API (WFS) => queryable and usable by remote clients as e.g. QGIS, ArcGIS, OpenLayers, etc.
+ Easy and efficient implementation.
- Would require extensions for complex features as e.g. contact info, parameters, EnvThes keywords etc.

Option 3. Extending the Option 2 with GeoServer AppSchema plugin

We can use Geoserver AppSchema plugin to implement mapping and provide complex features as INSPIRE EF datasets exposed by WFS API.

Schemata: Mapping of all site model contents to the INSPIRE EF Application Schema.

Evaluation:

+ Provides service API (WFS 2.0)
+ Real ETL workflow of DEIMS-SDR site data to "INSPIRE" database, which

can be done as SQL views, which offers on the fly changes instantly sharable via WFS service.

- including complex feature types, which have very low support in current

client applications from the portrayal and use point of view.

Implementation options:

- ETL transformation of DEIMS-SDR fields to INSPIRE database
- Predefined dataset (GML) as static view
- Sharing via service as 'dynamic' view

Option 4. Developing DEIMS-SDR module to provide EMF XML view

We can reuse the module created for the dataset ISO GMD XML export DEIMS-SDR module. The mapping rules will be defined by the PHP code in the templates created for each content types involved in the site description, e.g. Person, Dataset, Data Product, Research Site, Organization, etc.

Schemata: All appropriate DEIMS-SDR fields are mapped to EF elements, the rendering rules are defined either using formatters and set in content form display definition, and / or directly in the template file.

Evaluation:

+ Drupal API friendly solution reusable by multiple DEIMS-SDR sites if needed.

+ Easy to deploy and configure by site admin using web administration interface.

+ Extensible and customizable by any Drupal positive developer.

+ Produces valid INSPIRE EMF XML documents.

- Does not provide standardized service API (e.g. WFS).

6.2.6 Implementation

Option 4 was used to implement the requirements to provide site and related data model information in a standardized and exchangeable data model.

The development was done in the MK18 environment[78] on the DEIMS-SDR version backed up in November 2016. First release has been published in Github repository[79] and deployed by UBA first on the development for testing purpose and after on the production DEIMS-SDR for the real usage.

Basic description of the module with the instructions how to deploy it on a DEIMS-SDR instance is provided in the Github repository, together with 2 examples of site EMF XML encodings provided by the production instance.

The core of the mapping is implemented in the site template file[80], which defines the backbone of the EMF XML representation rendered by Drupal API for a site. Individual mandatory and optional elements defined in the EMF data model[81] and matched to corresponding DEIMS-SDR site / person / dataset / research site / data product / organization fields have been mapped as follows:

- The root element *EnvironmentalMonitoringFacility* contains used namespaces definitions, ef data model XML schema location definition used for validation and editing purposes and gml:id attribute which should serve as unique identifier of the dataset, which is generated as “Facility_SiteUUID” in examples:

gml:id="Facility_8eda49e9-1f4e-4f3e-b58e-e0bb25dc32a6"

INSPIRE ID

- INSPIRE mandatory element *inspireId* is implemented using the site UUID[82] in the element <base:localId> and DEIMS-SDR site URL in the element <base:namespace> as in the following example:

```
<ef:inspireId>
  <base:Identifier>
    <base:localId>8eda49e9-1f4e-4f3e-b58e-e0bb25dc32a6</base:localId>
    <base:namespace>https://data.lter-europe.net/deims</base:namespace>
  </base:Identifier>
</ef:inspireId>
```

NAME

78 http://bolegweb.geof.unizg.hr/deims/

79 https://github.com/MK18SRO/deims_emf_module

80 https://github.com/MK18SRO/deims_emf_module/blob/master/templates/emf--node--site.tpl.php

81 http://inspire.ec.europa.eu/data-model/approved/r4618-ir/html/index.htm?goto=2:3:6:1:1:7980

82 https://data.lter-europe.net/deims/documentation/site#R2

- INSPIRE optional element <ef:name> has been mapped to the site name[83] corresponding field as in example:

```
<ef:name>Zöbelboden LTER IM master site (ICP_IM_AT01)</ef:name>
```

ADDITIONAL INFORMATION

- INSPIRE optional element *additionalDescription* has been mapped to the site corresponding node site description field[84] as follows:

```
<ef:additionalDescription>
  The Zöbelboden was established in 1992 as the only Integrated Monitoring station in
  Austria under the UN Convention on long-range transboundary air pollution (CLRTAP). In
  2006 it became part of LTER Austria. The Zöbelboden covers a small forested catchment
  (90 ha) of a karstic mountain range (500 to 950 m above sea level) in the Kalkalpen
  national park. Monitoring and research is focussing on air pollution effects on forested
  catchments and its interaction with climate change. The Zöbelboden represents one of
  the best known karst catchments in Europe with long-term data series of the major
  components of its ecosystems. The Zöbelboden is managed by the Umweltbundesamt GmbH.
  Sampling of chemical specimen is done by local staff. Chemical analyses are carried out
  by the laboratory of the Umweltbundesamt in Vienna. All data and metadata from
  monitoring and research projects are stored in a semantically structured database.
</ef:additionalDescription>
```

MEDIA MONITORED

- INSPIRE multiple and mandatory element *mediaMonitored* has been mapped to the site corresponding node GEO-BON biome field[85] due it's high semantic similarity to the INSPIRE MediaValue codeList[86] as follows:

```
<ef:mediaMonitored xlink:title="Terrestrial"/>
```

---

83 https://data.lter-europe.net/deims/documentation/site#R1

84 https://data.lter-europe.net/deims/documentation/site#R5

85 https://data.lter-europe.net/deims/documentation/site#R22

86 http://inspire.ec.europa.eu/codeList/MediaValue

Comment: This is not fully INSPIRE compliant, we will need to either map the Geo-Bon biomes to the INSPIRE MediaValue and provide also xlink:href pointing to the INSPIRE registry or use ILTER Biome values and provide their definition in the EnvThes liked again by the xlink:href.

LEGAL BACKGROUND

- INSPIRE optional multiple element *legalBackground* has been mapped to the linked dataset metadata element Legal obligation reporting due to its absence in the site model. Each site, which has linked dataset metadata that have legal obligation data provided, render those as in the following example:

```
<ef:legalBackground>
  <base2:LegislationCitation gml:id="Dataset_d9e94776-e7a8-11e2-a655-
  005056ab003f_5877344811c32">
    <base2:name>
      Directive 2000/60/EC of the European Parliament and of the Council of 23 October
      2000 establishing a framework for Community action in the field of water policy
    </base2:name>
    <base2:date>
      <gmd:CI_Date>
        <gmd:date>
          <gco:Date>2000-10-23</gco:Date>
        </gmd:date>
        <gmd:dateType>
          <gmd:CI_DateTypeCode
          codeList="http://www.isotc211.org/2005/resources/Codelist/gmxCodelists.xml"
          codeListValue="publication"
          codeSpace="ISOTC211/19115">publication</gmd:CI_DateTypeCode>
        </gmd:dateType>
      </gmd:CI_Date>
    </base2:date>
    <base2:link nilReason="missing"/>
    <base2:level nilReason="missing"/>
  </base2:LegislationCitation>
</ef:legalBackground>
```

For all mandatory elements nested in the legalBackground object where DEIMS-SDR does not natively provide any information we used the nilReason attribute with value missing in order to ensure the data validity.

RESPONSIBLE PARTY

- INSPIRE multiple optional element *responsibleParty* has been mapped to all fields providing site contact details: Site Manager[87], Site Owner[88], Founding Organisation[89] and Metadata Provider[90]. The mapping was simplified for the base2:contact address element information where instead of using nested elements from linked INSPIRE schemes for administrative units and geographic names, we concatenate the contact address information about the person in a single base2 schema element <base2:contactInstructions> as following example:

```
<base2:contact>
  <base2:Contact>
    <base2:contactInstructions>
      <gco:CharacterString>Spittelauer Lände 5 Vienna 1090 AT</gco:CharacterString>
    </base2:contactInstructions>
    <base2:electronicMailAddress>thomas.dirnboeck@umweltbundesamt.at</base2:electronicM
    ailAddress>
  </base2:Contact>
</base2:contact>
```

Additionally, all the contact information metadata available in the person model[91] as Phone, email, web address, position and role are encoded as in following example:

```
<ef:responsibleParty>
  <base2:RelatedParty>
    <base2:individualName>
      <gco:CharacterString>Tomas Kliment</gco:CharacterString>
    </base2:individualName>
    <base2:organisationName>
      <gco:CharacterString>MK 18 s.r.o.</gco:CharacterString>
    </base2:organisationName>
    <base2:positionName>
      <gco:CharacterString>Ing., PhD.</gco:CharacterString>
    </base2:positionName>
    <base2:contact>
      <base2:Contact>
        <base2:contactInstructions>
          <gco:CharacterString>Juzná 698/4 Roznava 04801 SK</gco:CharacterString>
        </base2:contactInstructions>
        <base2:electronicMailAddress>tomas.kliment@gmail.com</base2:electronicMailAddres
        s>
        <base2:telephoneVoice>00421907232019</base2:telephoneVoice>
        <base2:website>http://about.me/klimeto</base2:website>
```

[87] https://data.lter-europe.net/deims/documentation/site#R9

[88] https://data.lter-europe.net/deims/documentation/site#R10

[89] https://data.lter-europe.net/deims/documentation/site#R11

[90] https://data.lter-europe.net/deims/documentation/site#R13

[91] https://data.lter-europe.net/deims/documentation/person#P4

```
      </base2:Contact>
    </base2:contact>
    <base2:role xlink:role="Researcher"/>
  </base2:RelatedParty>
</ef:responsibleParty>
<ef:responsibleParty>
```

Since the site editing form allows users to identify Site Contact Details as an individual, who may be linked to an organization, or only an organization, for which DEIMS-SDR contains only 2 data fields with information about the name and URL address, we provide the following valid encoding for the case when only an organization was provided:

```
<ef:responsibleParty>
  <base2:RelatedParty>
    <base2:organisationName>
      <gco:CharacterString>MK 18 s.r.o.</gco:CharacterString>
    </base2:organisationName>
    <base2:contact>
      <base2:Contact>
        <base2:website>http://www.mk18sro.sk/</base2:website>
      </base2:Contact>
    </base2:contact>
  </base2:RelatedParty>
</ef:responsibleParty>
```

GEOMETRY

- INSPIRE optional element *geometry* has been mapped into <gml:MultiGeometry> element with options to provide encoding of 2 types of geometric primitives: <gml:Polygon> and <gml:Point> as multiple geometry members. Depending what geographic data users provide the module will render EMF representation. The following example renders either possible ways (Bounds and Centroid):

```
<ef:geometry>
  <gml:MultiGeometry gml:id="LTER_EU_AT_003">
    <gml:geometryMember>
      <gml:Polygon gml:id="LTER_EU_AT_003_BOUNDS"
      srsName="http://www.opengis.net/def/crs/EPSG/0/4326">
        <gml:exterior>
          <gml:LinearRing>
            <gml:posList>
              47.837485912663 14.452910303853 47.84700622596 14.452910303853
              47.84700622596 14.43536201892 47.837485912663 14.43536201892
              47.837485912663 14.452910303853
            </gml:posList>
          </gml:LinearRing>
        </gml:exterior>
      </gml:Polygon>
```

```
    </gml:geometryMember>
    <gml:geometryMember>
      <gml:Point gml:id="LTER_EU_AT_003_POINT"
      srsName="http://www.opengis.net/def/crs/EPSG/0/4326">
        <gml:pos srsDimension="2">47.842246069311 14.444136161386</gml:pos>
      </gml:Point>
    </gml:geometryMember>
  </gml:MultiGeometry>
</ef:geometry>
```

The GML Identifier attribute for the entire geometry object is mapped to the Site Code[92] field and then individual geometric primitives have the id composed of the the Site Code and primitive type, e.g. LTER_EU_AT_003_POINT

ONLINE RESOURCE

- INSPIRE optional element *onlineResource* has been mapped to the element Online Information[93]. In addition, each site has at least one onlineResource element, which points at the HTML representation of the site metadata within DEIMS-SDR site. Example XML encoding is rendered as follows:

```
<ef:onlineResource>
  https://data.lter-europe.net/deims/site/8eda49e9-1f4e-4f3e-b58e-e0bb25dc32a6
</ef:onlineResource>
<ef:onlineResource>
  http://www.umweltbundesamt.at/en/services/services_pollutants/services_airquality/en_
  ref_zoebelboden/
</ef:onlineResource>
```

PURPOSE

- INSPIRE optional element *purpose* is defined as an empty codeList and thus we could not provide the data available from element Purpose of the site[94]. In order to identify those sites having provided the purpose information, the module renders at least the following XML snippet:

92 https://data.lter-europe.net/deims/documentation/site#R3

93 https://data.lter-europe.net/deims/documentation/site#R12

94 https://data.lter-europe.net/deims/documentation/site#R5-2

```
<ef:purpose
xlink:href="http://inspire.ec.europa.eu/codelist/PurposeOfCollectionValue"/>
```

OBSERVING CAPABILITY

- INSPIRE optional element *observingCapability* is complex object where only the property observedProperty has been mapped to the Site parameter element[95], which is linked to the EnvThes. The remaining mandatory sub elements are rendered empty with the nilReason attribute with value missing as following example:

```
<ef:observingCapability>
  <ef:ObservingCapability gml:id="ObservingCapability_LTER_EU_AT_003_58775cf204d78">
    <ef:observingTime xsi:nil="true" nilReason="missing"/>
    <ef:processType xsi:nil="true" nilReason="missing"/>
    <ef:resultNature xsi:nil="true" nilReason="missing"/>
    <ef:procedure nilReason="missing"/>
    <ef:featureOfInterest nilReason="missing"/>
    <ef:observedProperty     xlink:href="http://vocabs.ceh.ac.uk/evn/tbl/sparql?default-
    graph-uri=urn:x-evn-
    pub:EnvThes&format=text/json&query=SELECT%20%3Fresult%0AWHERE%20%7B%0A%09GRAPH%20%3
    Curn%3Ax-evn-
    pub%3AEnvThes%3E%20%7B%0A%09%09%3Fresult%20a%20%3Chttp%3A%2F%2Fwww.w3.org%2F2004%2F
    02%2Fskos%2Fcore%23Concept%3E%20.%0A%20%20%20%20%20%20%20%20%20%20%20%20%20%20%2
    0FILTER%20EXISTS%20%7B%0A%20%20%20%20%20%20%20%20%20%20%20%3Fresult%20%3FanyPropert
    y%20%3FanyValue%20.%0A%20%20%20%20%20%20%20%20%20%20%20FILTER%20(isLiteral(%3FanyVa
    lue)%20%26%26%20regex(LCASE(str(%3FanyValue))%2C%20%22(%3F%3D.*ecosystem
    measure*)%22))%20.%0A%20%20%20%20%20%20%20%7D%20.%0A%09%7D%0ABIND%20(%3Chttp%3A%2F%
    2Fwww.w3.org%2F2004%2F02%2Fskos%2Fcore%23prefLabel%3E(%3Fresult)%20AS%20%3Flabel)%2
    0.%0A%7D%20ORDER%20BY%20(LCASE(%3Flabel))"/>
  </ef:ObservingCapability>
</ef:observingCapability>
```

The value of the gml identifier for observingCapability object is created as a concatenation of the string ObservingCapability, Site Code and PHP uniqid function, which generates a unique ID based on the microtime (current time in microseconds).

The value of the site parameter provided in the site metadata is used to create a Sparql query, which is encoded in the xlink:href attribute and resolves in json representation of the query result returned by the EnvThes Sparql endpoint as following example:

```
{
	"head" : {
		"vars" : ["result"]
	},
```

95 https://data.lter-europe.net/deims/documentation/site#R38

```
	"results" : {
		"bindings" : [{
				"result" : {
					"type" : "uri",
					"Value" : "http://vocabs.lter-
europe.net/EnvThes/USLterCV_716"
				}
			}
		]
	}
}
```

Comments: The content of individual elements in the observingCapability object should be provided / linked (xlink:href) from the SOS API with O&M data if available.

BROADER

- INSPIRE optional element *broader* has been mapped to the element Parent site name[96] using the xlink:href attribute referring to EMF XML representation of the parent site as in the following example:

```
<ef:broader xlink:href="https://data.lter-europe.net/deims/node/8612/emf">
  <ef:Hierarchy gml:id="Facility_d0a8da18-0881-4ebe-bccf-bc4cb4e25701">
    <ef:linkingTime>
      <gml:TimePeriod gml:id="timePeriod01">
        <gml:beginPosition>1960-01-01</gml:beginPosition>
        <gml:endPosition indeterminatePosition="now"/>
      </gml:TimePeriod>
    </ef:linkingTime>
    <ef:broader/>
    <ef:narrower/>
  </ef:Hierarchy>
</ef:broader>
```

Additionally, the child mandatory element *Hierarchy* contains one mandatory element *linkingTime* which has been mapped to the static code due to missing information about the hierarchy establishment time stamp.

HAS OBSERVATION

[96] https://data.lter-europe.net/deims/documentation/site#R7

INSPIRE element *hasObservations* (multiple) has been mapped to dataset metadata linked to the site and provides xlink to either ISO19139 XML representation generated by iso19139 DEIMS-SDR module or if provided in the dataset metadata it provides link to SOS service.

Example of both ways is provided in the following XML code fragment:

```
<ef:hasObservation xlink:title="IT_SI001230_Lake Scuro_Chla_20130228"
xlink:href="http://sk.ise.cnr.it/observations/sos/kvp?service=SOS&version=2.0.0&request
=GetObservation&offering=offering:http://sp7.irea.cnr.it/sensors/sk.ise.cnr.it/procedur
e/noManufacturerDeclared/noModelDeclared/noSerialNumberDeclared/20150428125100001/obser
vations&observedProperty=http://vocab.nerc.ac.uk/collection/P02/current/CPWC/&procedure
=http://sp7.irea.cnr.it/sensors/sk.ise.cnr.it/procedure/noManufacturerDeclared/noModelD
eclared/noSerialNumberDeclared/20150428125100001&featureOfInterest=http://sp7.irea.cnr.
it/sensors/sk.ise.cnr.it/foi/SSF/SP/EPSG:4326/44.381805559999997/10.045638889999999&Mer
geObservationsIntoDataArray=trueDistribution"/>
<ef:hasObservation xlink:title="IT_SI001230_Lake Scuro_Conductivity_20130301"
xlink:href="http://bolegweb.geof.unizg.hr/deims/node/8713/iso19139"/>
```

The first hasObservation element provides in xlink:href attribute a link to GetObservation request extracted from the dataset metadata. The second hasObservation element provides a link to iso19139 metadata representation because it does not provide any SOS link. The xlink:title attribute is populated by the dataset title provided by users in the DEIMS.

INVOLVED IN

INSPIRE optional multiple element *involvedIn* has been mapped to relevant data products content type fields linked to site via a view as in the following example:

```
<ef:involvedIn>
 <ef:EnvironmentalMonitoringActivity  gml:id="Data_collection_activity_8689b125-ee46-
4d09-9e46-640f9c5c6eab_5877c2697a53f">
    <gml:description>
      landscape measure, elevation, land cover, land use, landscape cover
    </gml:description>
    <gml:identifier codeSpace="Airborne Images"/>
    <gml:name>LTER Zöbelboden Austria aerial data</gml:name>
    <ef:activityTime>
      <gml:TimePeriod gml:id="Data_collection_activity_TimePeriod_8689b125-ee46-4d09-
      9e46-640f9c5c6eab_5877c2697a580">
        <gml:beginPosition>1995-07-15</gml:beginPosition>
        <gml:endPosition>2016-07-15</gml:endPosition>
      </gml:TimePeriod>
    </ef:activityTime>
    <ef:activityConditions>
      Aerial photographs and LiDAR data of the LTER Zöbelboden catchment
    </ef:activityConditions>
```

```
    <ef:responsibleParty>
      <base2:RelatedParty>
        <base2:individualName>
          <gco:CharacterString>Johannes Kobler, Thomas Dirnboeck</gco:CharacterString>
        </base2:individualName>
      </base2:RelatedParty>
    </ef:responsibleParty>
    <ef:inspireId>
      <base:Identifier>
        <base:localId>8689b125-ee46-4d09-9e46-640f9c5c6eab</base:localId>
        <base:namespace>https://data.lter-europe.net/deims</base:namespace>
      </base:Identifier>
    </ef:inspireId>
    <ef:onlineResource>https://data.lter-
    europe.net/deims/node/10282</ef:onlineResource>
  </ef:EnvironmentalMonitoringActivity>
</ef:involvedIn>
```

The metadata available in the product data model are aggregated into the EnvironmentalMonitoringActivity complex type. It's gml identifier is composed using standard pattern ContentType_UUID_uniqid() as in the example above: `Data_collection_activity_TimePeriod_8689b125-ee46-4d09-9e46-640f9c5c6eab_5877c2697a580`. Data product title[97] is placed into the gml:name element; Data product type[98] in gml:identifier element attribute codeSpace; Data product abstract in the ef:activityConditions, parameters are listed in gml:description, Data product temporal extent in ef:activityTime and finally Contact information simplified to full names are aggregated in ef:responsibleParty.

REPRESENTATIVE POINT

INSPIRE optional element *representativePoint* is mapped to Site Geographic Center Coordinates[99] field as follows:

```
<ef:representativePoint>
  <gml:Point gml:id="LTER_EU_AT_003_CENTROID"
  srsName="http://www.opengis.net/def/crs/EPSG/0/4326">
    <gml:pos srsDimension="2">47.842246069311 14.444136161386</gml:pos>
  </gml:Point>
</ef:representativePoint>
```

[97] https://data.lter-europe.net/deims/documentation/activity#DP1

[98] https://data.lter-europe.net/deims/documentation/activity#DP3

[99] https://data.lter-europe.net/deims/documentation/site#R16

MEASUREMENT REGIME

INSPIRE mandatory element *measurementRegime* has been mapped as a static value for all sites from the MeasurementRegimeValue codeList class as follows:

```
<ef:measurementRegime
xlink:href="http://inspire.ec.europa.eu/codeList/MeasurementRegimeValue/continuousDataC
ollection"/>
```

MOBILE

INSPIRE mandatory element *mobile* has been mapped as a static value to be false for all sites as follows:

```
<ef:mobile>false</ef:mobile>
```

OPERATIONAL ACTIVITY PERIOD

INSPIRE mandatory multiple element *OperationalActivityPeriod* has been mapped to the site status[100] elements as follows:

```
<ef:operationalActivityPeriod xsi:nil="false">
  <ef:OperationalActivityPeriod    gml:id="OperationalActivityPeriod_8eda49e9-1f4e-4f3e-
  b58e-e0bb25dc32a6">
    <ef:activityTime>
      <gml:TimePeriod gml:id="timePeriod_8eda49e9-1f4e-4f3e-b58e-e0bb25dc32a6">
        <gml:beginPosition>1992-01-01</gml:beginPosition>
        <gml:endPosition indeterminatePosition="now"/>
      </gml:TimePeriod>
    </ef:activityTime>
  </ef:OperationalActivityPeriod>
</ef:operationalActivityPeriod>
```

[100] https://data.lter-europe.net/deims/documentation/site#R35

The gml:TimePeriod has been used serialize the time information where the begin position is mapped to Year site established and if Year site closed is empty, which means the site is still operating, endPosition is filled by attribute indeterminatePosition by value now, otherwise endPosition is filled by date.

BELONGS TO

INSPIRE optional multiple element *belongsTo* has been mapped to LTER National Network[101] and Networks (in addition to LTER)[102] elements as follows:

```
<ef:belongsTo xlink:href="https://data.lter-europe.net/deims/node/222">
  <ef:NetworkFacility gml:id="Network_e3d5faf7-5603-4acd-b835-eaedca8e5dec">
    <gml:name>Austria (LTER-Austria)</gml:name>
    <ef:linkingTime>
      <gml:TimePeriod gml:id="timePeriod_e3d5faf7-5603-4acd-b835-eaedca8e5dec">
        <gml:beginPosition>2012-10-22</gml:beginPosition>
        <gml:endPosition indeterminatePosition="now"/>
      </gml:TimePeriod>
    </ef:linkingTime>
    <ef:belongsTo/>
    <ef:contains/>
  </ef:NetworkFacility>
</ef:belongsTo>
<ef:belongsTo>
  <ef:NetworkFacility gml:id="Network_b4f312d0-ad5d-4421-9876-3403329d2132">
    <gml:name>UNECE ICP Integrated Monitoring</gml:name>
    <ef:linkingTime>
      <gml:TimePeriod gml:id="timePeriod_b4f312d0-ad5d-4421-9876-3403329d21321">
        <gml:beginPosition>2012-10-22</gml:beginPosition>
        <gml:endPosition indeterminatePosition="now"/>
      </gml:TimePeriod>
    </ef:linkingTime>
    <ef:belongsTo/>
    <ef:contains/>
  </ef:NetworkFacility>
</ef:belongsTo>
```

The first section appearance of ef:belongsTo element always renders data from linked LTER National Network, if any, with xlink:href value linking to the node of that network, gml:id containing the network node UUID, gml:name network node title and ef:linkingTime is for gml:beginPosition identified as the date of creation for network node and gml:endPosition as value now.

101 https://data.lter-europe.net/deims/documentation/site#R30

102 https://data.lter-europe.net/deims/documentation/site#R32

The above mentioned applies also for other networks, but it has no link to any node, since it is a list of networks without any further information available in DEIMS-SDR.

Possible further development:


```
<?xml version="1.0" encoding="UTF-8"?>
<ef:EnvironmentalMonitoringNetwork xmlns:gml="http://www.opengis.net/gml/3.2"
    xmlns:xlink="http://www.w3.org/1999/xlink" xmlns:gsr="http://www.isotc211.org/2005/gsr"
    xmlns:gco="http://www.isotc211.org/2005/gco" xmlns:gts="http://www.isotc211.org/2005/gts"
    xmlns:gss="http://www.isotc211.org/2005/gss" xmlns:gmd="http://www.isotc211.org/2005/gmd"
    xmlns:base="http://inspire.ec.europa.eu/schemas/base/3.3"
    xmlns:om="http://www.opengis.net/om/2.0" xmlns:gn="http://inspire.ec.europa.eu/schemas/gn/4.0"
    xmlns:net="http://inspire.ec.europa.eu/schemas/net/4.0"
    xmlns:sc="http://www.interactive-instruments.de/ShapeChange/AppInfo"
    xmlns:ad="http://inspire.ec.europa.eu/schemas/ad/4.0"
    xmlns:base2="http://inspire.ec.europa.eu/schemas/base2/2.0"
    xmlns:tn="http://inspire.ec.europa.eu/schemas/tn/4.0"
    xmlns:au="http://inspire.ec.europa.eu/schemas/au/4.0"
    xmlns:cp="http://inspire.ec.europa.eu/schemas/cp/4.0"
    xmlns:bu-base="http://inspire.ec.europa.eu/schemas/bu-base/4.0"
    xmlns:ef="http://inspire.ec.europa.eu/schemas/ef/4.0"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:schemaLocation="http://inspire.ec.europa.eu/schemas/ef/4.0
http://inspire.ec.europa.eu/schemas/ef/4.0/EnvironmentalMonitoringFacilities.xsd"
    gml:id="LTER-UUID">
    <!-- UUID - Remember the gmlID can't start with number -->
    <!-- Example of research network LTER-Europe -->

    <ef:inspireId>
        <base:Identifier>
            <base:localId>LTER_EU</base:localId>
            <base:namespace>ILTER</base:namespace>
            <!-- The namespace value will be owned by the data provider of the spatial object and will be registered in the INSPIRE
External Object Identifier Namespaces Register. -->
        </base:Identifier>
    </ef:inspireId>

    <ef:name>LTER-Eu</ef:name>

    <ef:additionalDescription>
        network description
    </ef:additionalDescription>

    <!-- mediaMonitored is from a codeList with some values: air, biota, landscape, sediment, soil/ground, waste, water -->
    <!-- could be terms from EnvThes -->
    <ef:mediaMonitored xlink:href="http://inspire.ec.europa.eu/codeList/MediaValue/water"/>
    <ef:mediaMonitored xlink:href="http://inspire.ec.europa.eu/codeList/MediaValue/waste"/>
    <ef:mediaMonitored xlink:href="http://inspire.ec.europa.eu/codeList/MediaValue/soil/ground"/>
    <ef:mediaMonitored xlink:href="http://inspire.ec.europa.eu/codeList/MediaValue/sediment"/>
    <ef:mediaMonitored xlink:href="http://inspire.ec.europa.eu/codeList/MediaValue/landscape"/>
    <ef:mediaMonitored xlink:href="http://inspire.ec.europa.eu/codeList/MediaValue/biota"/>
    <ef:mediaMonitored xlink:href="http://inspire.ec.europa.eu/codeList/MediaValue/air"/>
    <!-- mediaMonitored is from a codeList (http://inspire.ec.europa.eu/codeList/MediaValue) with some values:
        air, biota, landscape, sediment, soil/ground, waste, water -->

    <ef:legalBackground/>
    <!-- ??? -->

    <!-- Research network responsible -->
    <ef:responsibleParty>
        <base2:RelatedParty>
            <base2:individualName>
                <gco:CharacterString>Mikael Mirtle</gco:CharacterString>
            </base2:individualName>

            <base2:organisationName>
                <gco:CharacterString>UBA</gco:CharacterString>
            </base2:organisationName>

            <base2:positionName>
                <gco:CharacterString>Reasercher</gco:CharacterString>
            </base2:positionName>

            <base2:contact>
                <base2:Contact>
                    <base2:address>
                        <!-- all the elements below are mandatory. We need to undestand how to match with DEIMS fileds -->
```

```
                        <ad:AddressRepresentation>
                            <ad:adminUnit>
                                <gn:GeographicalName>
                                    <gn:language/>
                                    <gn:nativeness/>
                                    <gn:nameStatus/>
                                    <gn:sourceOfName/>
                                    <gn:pronunciation>
                                        <gn:PronunciationOfName/>
                                    </gn:pronunciation>
                                    <gn:spelling>
                                        <gn:SpellingOfName>
                                            <gn:text/>
                                            <gn:script/>
                                        </gn:SpellingOfName>
                                    </gn:spelling>
                                </gn:GeographicalName>
                            </ad:adminUnit>
                        </ad:AddressRepresentation>
                    </base2:address>
                    <base2:electronicMailAddress>x.xxxxx@xxx.au</base2:electronicMailAddress>
                    <base2:telephoneFacsimile>+xx xxxxxxxxxx</base2:telephoneFacsimile>
                    <base2:telephoneVoice>+xx xxxxxxxxxx</base2:telephoneVoice>
                    <base2:website>http://www.xxx.xxx/xxxxxxxxxxx</base2:website>
                </base2:Contact>
            </base2:contact>

            <base2:role xlink:role="Site Manager"
                xlink:href="http://xxx.xxx.xx/roleCodeList/siteManager"/>
            <!-- for xlink:href needs to create a codlist for different role terms:
                network chair, researcher, site manager, data manger, faculty associate, post doctoral associate, other professional,
graduate student, udergraduate student
            -->
        </base2:RelatedParty>
    </ef:responsibleParty>

    <ef:geometry xlink:href="http://www.geonodeLTER-Italy.xx/WFS/LTER_EU_IT" />

    <ef:purpose/>

    <!--<ef:observingCapability>
        <ef:ObservingCapability gml:id="aaaac">
            <ef:observingTime/>
            <ef:processType/>
            <ef:resultNature/>
            <ef:procedure/>
            <ef:observedProperty/>
        </ef:ObservingCapability>
    </ef:observingCapability>-->
    <!-- relation with procedure, FOI and ObservedProp -->

    <ef:broader xlink:href="http://xxx.xxxx.xxx/LTER_EMN_International.xml">
        <ef:Hierarchy gml:id="LTER-International">
            <ef:linkingTime>
                <gml:TimePeriod gml:id="timePeriod04">
                    <gml:beginPosition>1960-01-01</gml:beginPosition>
                    <gml:endPosition indeterminatePosition="now"/>
                    <!-- or
                        <gml:endPosition>2016-01-01</gml:endPosition>
                    -->
                </gml:TimePeriod>
            </ef:linkingTime>
            <ef:broader/>
            <ef:narrower/>
        </ef:Hierarchy>
    </ef:broader>

    <ef:narrower xlink:href="http://xxx.xxxx.xxx/LTER_EMN_Italy.xml">
        <ef:Hierarchy gml:id="LTER-Italy">
            <ef:linkingTime/>
            <ef:broader/>
            <ef:narrower/>
        </ef:Hierarchy>
    </ef:narrower>

    <!--<ef:supersedes/>-->
    <!-- In a genealogy, the AbstractMonitoringObject(s) that has(have) been deactivated/replaced by another one. -->
    <!-- To use only in this case -->

    <!--<ef:supersededBy/>-->
    <!-- In a genealogy, the newly active AbstractMonitoringObject(s) that replaces(replace) the superseded one. -->
    <!-- To use only in this case -->

    <!--<ef:reportedTo>
        <ef:ReportToLegalAct>
```

```
            <ef:legalAct/>
            <ef:reportDate/>
            <ef:reportedEnvelope/>
            <ef:observationRequired/>
            <ef:observingCapabilityRequired/>
            <ef:description/>
        </ef:ReportToLegalAct>
    </ef:reportedTo>-->

    <!--<ef:hasObservation />-->
    <!-- OM observation -->

    <!--<ef:involvedIn/>-->

    <ef:organisationLevel xlink:href="http://inspire.ec.europa.eu/codeList/LegislationLevelValue/european"/>
    <!-- organisationLevel is from a codeList (http://inspire.ec.europa.eu/codeList/LegislationLevelValue) with some values:
        european, international, national, sub-national -->

    <ef:contains xlink:href="http://xxx.xxxx.xxx/LTER_EMF_ParentSite008.xml"/>
    <ef:contains xlink:href="http://xxx.xxxx.xxx/LTER_EMF_ResearchSite_045.xml"/>
    <ef:contains xlink:href="http://xxx.xxxx.xxx/LTER_EMF_ResearchSite_043.xml"/>
    <ef:contains xlink:href="http://xxx.xxxx.xxx/LTER_EMF_ResearchSite_087.xml"/>
    <ef:contains xlink:href="http://xxx.xxxx.xxx/LTER_EMF_ResearchSite_044.xml"/>
    <ef:contains xlink:href="http://xxx.xxxx.xxx/LTER_EMF_ResearchSite_042.xml"/>
    <ef:contains xlink:href="http://xxx.xxxx.xxx/LTER_EMF_ResearchSite_102.xml"/>
    <!-- all europe research and parent sites -->

</ef:EnvironmentalMonitoringNetwork>
```

6.3 Annex 3 – Example INSPIRE EF record for Nationalpark Kalkalpen

Link: https://data.lter-europe.net/deims/node/8642/emf


```
<?xml version="1.0" encoding="utf-8"?>
<ef:EnvironmentalMonitoringFacility xmlns:gml="http://www.opengis.net/gml/3.2"
xmlns:xlink="http://www.w3.org/1999/xlink" xmlns:gsr="http://www.isotc211.org/2005/gsr"
xmlns:gco="http://www.isotc211.org/2005/gco" xmlns:gts="http://www.isotc211.org/2005/gts"
xmlns:gss="http://www.isotc211.org/2005/gss" xmlns:gmd="http://www.isotc211.org/2005/gmd"
xmlns:base="http://inspire.ec.europa.eu/schemas/base/3.3" xmlns:om="http://www.opengis.net/om/2.0"
xmlns:gn="http://inspire.ec.europa.eu/schemas/gn/4.0" xmlns:net="http://inspire.ec.europa.eu/schemas/net/4.0"
xmlns:sc="http://www.interactive-instruments.de/ShapeChange/AppInfo"
xmlns:ad="http://inspire.ec.europa.eu/schemas/ad/4.0"
xmlns:base2="http://inspire.ec.europa.eu/schemas/base2/2.0"
xmlns:tn="http://inspire.ec.europa.eu/schemas/tn/4.0" xmlns:au="http://inspire.ec.europa.eu/schemas/au/4.0"
xmlns:cp="http://inspire.ec.europa.eu/schemas/cp/4.0" xmlns:bu-base="http://inspire.ec.europa.eu/schemas/bu-
base/4.0" xmlns:ef="http://inspire.ec.europa.eu/schemas/ef/4.0" xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance" xsi:schemaLocation="http://inspire.ec.europa.eu/schemas/ef/4.0
http://inspire.ec.europa.eu/schemas/ef/4.0/EnvironmentalMonitoringFacilities.xsd" gml:id="Facility_49515dda-
1198-4013-8f43-c33e107af081">
  <ef:inspireId>
    <base:Identifier>
      <base:localId>49515dda-1198-4013-8f43-c33e107af081</base:localId>
      <base:namespace>https://data.lter-europe.net/deims</base:namespace>
    </base:Identifier>
  </ef:inspireId>
  <ef:name>Kalkalpen National Park</ef:name>
  <ef:additionalDescription>
Kalkalpen National Park is made up of two mountain ranges * The Reichraminger Hintergebirge is one of Austria's
largest distinct forest areas - a sea of forest, which has not yet been dissected by public transportation
routes and human habitation. Here, you will also find one of the longest intact stream systems of the Eastern
Alps. Old shelters and overgrown trails remind us today of how wood was used and harvested in earlier times. *
The Sengsengebirge is a northern outpost of the Limestone Alps. The ca. 20 km long main ridge reaches its
highest point at the Hoher Nock (1,963 m). The name Sengsengebirge can be traced back to the use of the forests
as a source of energy for the numerous scythe smithies once located here. Facts and Figures Established: July
25, 1997 Area: 20,850 ha Zoning: 89% nature zone, 11% conservation area Property: 88% federally owned (Austrian
Forestry Service), 11% privately owned, and 1% municipal property Internationally recognized: as a national
park (IUCN category II) since 1998, Ramsar protected area (wetland of global importance), and Natura 2000 area
(European nature reserve) since 2004 Elevation: 385 to 1,963 m (Hohe Nock) Main types of rock: Wetterstein
limestone, primary dolomite Landscape Classification of Kalkalpen National Park (Updated June 2011) 81% forest
8% mountain pine 6% alpine pastures and meadows 5% rock and scree
</ef:additionalDescription>
  <ef:legalBackground>
    <base2:LegislationCitation gml:id="Dataset_915b6ea1-d4e3-4b29-8e0f-33ca90db5718_5afac0fbd7371">
      <base2:name>None</base2:name>
      <base2:date>
        <gmd:CI_Date>
          <gmd:date>
            <gco:Date>2000-01-01</gco:Date>
          </gmd:date>
          <gmd:dateType>
            <gmd:CI_DateTypeCode codeList="http://www.isotc211.org/2005/resources/Codelist/gmxCodelists.xml"
codeListValue="publication" codeSpace="ISOTC211/19115">publication</gmd:CI_DateTypeCode>
          </gmd:dateType>
        </gmd:CI_Date>
      </base2:date>
      <base2:link nilReason="missing" />
      <base2:level nilReason="missing" />
    </base2:LegislationCitation>
  </ef:legalBackground>
  <ef:legalBackground>
    <base2:LegislationCitation gml:id="Dataset_fbaff6f6-a378-4df9-8bed-416433149456_5afac0fbd73c8">
      <base2:name>None</base2:name>
      <base2:date>
        <gmd:CI_Date>
          <gmd:date>
            <gco:Date>2000-01-01</gco:Date>
```

```
          </gmd:date>
          <gmd:dateType>
            <gmd:CI_DateTypeCode   codeList="http://www.isotc211.org/2005/resources/Codelist/gmxCodelists.xml"
codeListValue="publication" codeSpace="ISOTC211/19115">publication</gmd:CI_DateTypeCode>
          </gmd:dateType>
        </gmd:CI_Date>
      </base2:date>
      <base2:link nilReason="missing" />
      <base2:level nilReason="missing" />
    </base2:LegislationCitation>
  </ef:legalBackground>
  <ef:legalBackground>
    <base2:LegislationCitation gml:id="Dataset_d25d745d-f887-48fd-9df0-80f9b961f3f5_5afac0fbd7419">
      <base2:name>None</base2:name>
      <base2:date>
        <gmd:CI_Date>
          <gmd:date>
            <gco:Date>2000-01-01</gco:Date>
          </gmd:date>
          <gmd:dateType>
            <gmd:CI_DateTypeCode   codeList="http://www.isotc211.org/2005/resources/Codelist/gmxCodelists.xml"
codeListValue="publication" codeSpace="ISOTC211/19115">publication</gmd:CI_DateTypeCode>
          </gmd:dateType>
        </gmd:CI_Date>
      </base2:date>
      <base2:link nilReason="missing" />
      <base2:level nilReason="missing" />
    </base2:LegislationCitation>
  </ef:legalBackground>
  <ef:legalBackground>
    <base2:LegislationCitation gml:id="Dataset_c79a7d6a-8643-4108-aa6c-5198bd68d9af_5afac0fbd7468">
      <base2:name>None</base2:name>
      <base2:date>
        <gmd:CI_Date>
          <gmd:date>
            <gco:Date>2000-01-01</gco:Date>
          </gmd:date>
          <gmd:dateType>
            <gmd:CI_DateTypeCode   codeList="http://www.isotc211.org/2005/resources/Codelist/gmxCodelists.xml"
codeListValue="publication" codeSpace="ISOTC211/19115">publication</gmd:CI_DateTypeCode>
          </gmd:dateType>
        </gmd:CI_Date>
      </base2:date>
      <base2:link nilReason="missing" />
      <base2:level nilReason="missing" />
    </base2:LegislationCitation>
  </ef:legalBackground>
  <ef:legalBackground>
    <base2:LegislationCitation gml:id="Dataset_ee2912df-6c4a-467f-97df-5427f4f5986b_5afac0fbd74b6">
      <base2:name>EU-Biodiversitätsstrategie target 2/action 5</base2:name>
      <base2:date>
        <gmd:CI_Date>
          <gmd:date>
            <gco:Date>2000-01-01</gco:Date>
          </gmd:date>
          <gmd:dateType>
            <gmd:CI_DateTypeCode   codeList="http://www.isotc211.org/2005/resources/Codelist/gmxCodelists.xml"
codeListValue="publication" codeSpace="ISOTC211/19115">publication</gmd:CI_DateTypeCode>
          </gmd:dateType>
        </gmd:CI_Date>
      </base2:date>
      <base2:link nilReason="missing" />
      <base2:level nilReason="missing" />
    </base2:LegislationCitation>
  </ef:legalBackground>
  <ef:legalBackground>
    <base2:LegislationCitation gml:id="Dataset_723fc9a3-d4ee-4409-ad4f-2e8d21555b72_5afac0fbd7504">
      <base2:name>None</base2:name>
      <base2:date>
        <gmd:CI_Date>
          <gmd:date>
```

```
            <gco:Date>2000-01-01</gco:Date>
          </gmd:date>
          <gmd:dateType>
            <gmd:CI_DateTypeCode   codeList="http://www.isotc211.org/2005/resources/Codelist/gmxCodelists.xml"
codeListValue="publication" codeSpace="ISOTC211/19115">publication</gmd:CI_DateTypeCode>
          </gmd:dateType>
        </gmd:CI_Date>
      </base2:date>
      <base2:link nilReason="missing" />
      <base2:level nilReason="missing" />
    </base2:LegislationCitation>
  </ef:legalBackground>
  <ef:legalBackground>
    <base2:LegislationCitation gml:id="Dataset_f358ffa6-eaa5-42ed-952f-9d41b1d15fdf_5afac0fbd7554">
      <base2:name>None</base2:name>
      <base2:date>
        <gmd:CI_Date>
          <gmd:date>
            <gco:Date>2000-01-01</gco:Date>
          </gmd:date>
          <gmd:dateType>
            <gmd:CI_DateTypeCode   codeList="http://www.isotc211.org/2005/resources/Codelist/gmxCodelists.xml"
codeListValue="publication" codeSpace="ISOTC211/19115">publication</gmd:CI_DateTypeCode>
          </gmd:dateType>
        </gmd:CI_Date>
      </base2:date>
      <base2:link nilReason="missing" />
      <base2:level nilReason="missing" />
    </base2:LegislationCitation>
  </ef:legalBackground>
  <ef:legalBackground>
    <base2:LegislationCitation gml:id="Dataset_33ea62e4-a7b1-435d-98e7-49d2e31cb3da_5afac0fbd75a3">
      <base2:name>None</base2:name>
      <base2:date>
        <gmd:CI_Date>
          <gmd:date>
            <gco:Date>2000-01-01</gco:Date>
          </gmd:date>
          <gmd:dateType>
            <gmd:CI_DateTypeCode   codeList="http://www.isotc211.org/2005/resources/Codelist/gmxCodelists.xml"
codeListValue="publication" codeSpace="ISOTC211/19115">publication</gmd:CI_DateTypeCode>
          </gmd:dateType>
        </gmd:CI_Date>
      </base2:date>
      <base2:link nilReason="missing" />
      <base2:level nilReason="missing" />
    </base2:LegislationCitation>
  </ef:legalBackground>
  <ef:legalBackground>
    <base2:LegislationCitation gml:id="Dataset_f58b8401-1082-49ad-851c-196a87284c01_5afac0fbd75f1">
      <base2:name />
      <base2:date>
        <gmd:CI_Date>
          <gmd:date>
            <gco:Date>2000-01-01</gco:Date>
          </gmd:date>
          <gmd:dateType>
            <gmd:CI_DateTypeCode   codeList="http://www.isotc211.org/2005/resources/Codelist/gmxCodelists.xml"
codeListValue="publication" codeSpace="ISOTC211/19115">publication</gmd:CI_DateTypeCode>
          </gmd:dateType>
        </gmd:CI_Date>
      </base2:date>
      <base2:link nilReason="missing" />
      <base2:level nilReason="missing" />
    </base2:LegislationCitation>
  </ef:legalBackground>
  <ef:legalBackground>
    <base2:LegislationCitation gml:id="Dataset_e0f479dd-a907-44c8-8c0a-ebcdd247fa2e_5afac0fbd763f">
      <base2:name />
      <base2:date>
        <gmd:CI_Date>
```

```
          <gmd:date>
            <gco:Date>2000-01-01</gco:Date>
          </gmd:date>
          <gmd:dateType>
            <gmd:CI_DateTypeCode   codeList="http://www.isotc211.org/2005/resources/Codelist/gmxCodelists.xml"
codeListValue="publication" codeSpace="ISOTC211/19115">publication</gmd:CI_DateTypeCode>
          </gmd:dateType>
        </gmd:CI_Date>
      </base2:date>
      <base2:link nilReason="missing" />
      <base2:level nilReason="missing" />
    </base2:LegislationCitation>
  </ef:legalBackground>
  <ef:legalBackground>
    <base2:LegislationCitation gml:id="Dataset_3acfa10b-daa9-47ce-893f-96e0ead9de11_5afac0fbd768d">
      <base2:name />
      <base2:date>
        <gmd:CI_Date>
          <gmd:date>
            <gco:Date>2000-01-01</gco:Date>
          </gmd:date>
          <gmd:dateType>
            <gmd:CI_DateTypeCode   codeList="http://www.isotc211.org/2005/resources/Codelist/gmxCodelists.xml"
codeListValue="publication" codeSpace="ISOTC211/19115">publication</gmd:CI_DateTypeCode>
          </gmd:dateType>
        </gmd:CI_Date>
      </base2:date>
      <base2:link nilReason="missing" />
      <base2:level nilReason="missing" />
    </base2:LegislationCitation>
  </ef:legalBackground>
  <ef:responsibleParty>
    <base2:RelatedParty>
      <base2:individualName>
        <gco:CharacterString>Franziska Pöpperl</gco:CharacterString>
      </base2:individualName>
      <base2:organisationName>
        <gco:CharacterString>LTER Europe</gco:CharacterString>
      </base2:organisationName>
      <base2:contact>
        <base2:Contact>
          <base2:electronicMailAddress>Franziska.Poepperl@kalkalpen.at</base2:electronicMailAddress>
        </base2:Contact>
      </base2:contact>
      <base2:role xlink:role="Data Manager" />
    </base2:RelatedParty>
  </ef:responsibleParty>
  <ef:responsibleParty>
    <base2:RelatedParty>
      <base2:individualName>
        <gco:CharacterString>Christian Fuxjaeger</gco:CharacterString>
      </base2:individualName>
      <base2:organisationName>
        <gco:CharacterString>LTER Europe</gco:CharacterString>
      </base2:organisationName>
      <base2:contact>
        <base2:Contact>
          <base2:electronicMailAddress>christian.fuxjaeger@kalkalpen.at</base2:electronicMailAddress>
        </base2:Contact>
      </base2:contact>
    </base2:RelatedParty>
  </ef:responsibleParty>
  <ef:responsibleParty>
    <base2:RelatedParty>
      <base2:organisationName>
        <gco:CharacterString>Nationalpark Kalkalpen</gco:CharacterString>
      </base2:organisationName>
      <base2:contact>
        <base2:Contact>
          <base2:website>http://www.kalkalpen.at/system/web/default.aspx</base2:website>
        </base2:Contact>
```

```
        </base2:contact>
      </base2:RelatedParty>
    </ef:responsibleParty>
    <ef:responsibleParty>
      <base2:RelatedParty>
        <base2:individualName>
          <gco:CharacterString>Franziska Pöpperl</gco:CharacterString>
        </base2:individualName>
        <base2:organisationName>
          <gco:CharacterString>LTER Europe</gco:CharacterString>
        </base2:organisationName>
        <base2:contact>
          <base2:Contact>
            <base2:electronicMailAddress>Franziska.Poepperl@kalkalpen.at</base2:electronicMailAddress>
          </base2:Contact>
        </base2:contact>
        <base2:role xlink:role="Data Manager" />
      </base2:RelatedParty>
    </ef:responsibleParty>
    <ef:responsibleParty>
      <base2:RelatedParty>
        <base2:individualName>
          <gco:CharacterString>Christian Fuxjaeger</gco:CharacterString>
        </base2:individualName>
        <base2:organisationName>
          <gco:CharacterString>LTER Europe</gco:CharacterString>
        </base2:organisationName>
        <base2:contact>
          <base2:Contact>
            <base2:electronicMailAddress>christian.fuxjaeger@kalkalpen.at</base2:electronicMailAddress>
          </base2:Contact>
        </base2:contact>
      </base2:RelatedParty>
    </ef:responsibleParty>
    <ef:responsibleParty>
      <base2:RelatedParty>
        <base2:individualName>
          <gco:CharacterString>Forschung Kalkalpen</gco:CharacterString>
        </base2:individualName>
        <base2:organisationName>
          <gco:CharacterString>LTER Europe</gco:CharacterString>
        </base2:organisationName>
        <base2:contact>
          <base2:Contact>
            <base2:electronicMailAddress>forschung@kalkalpen.at</base2:electronicMailAddress>
          </base2:Contact>
        </base2:contact>
      </base2:RelatedParty>
    </ef:responsibleParty>
    <ef:geometry>
      <gml:MultiGeometry gml:id="LTER_EU_AT_008">
        <gml:geometryMember>
          <gml:Polygon gml:id="LTER_EU_AT_008_BOUNDS" srsName="http://www.opengis.net/def/crs/EPSG/0/4326">
            <gml:exterior>
              <gml:LinearRing>
                <gml:posList>
47.795866525545 14.487113747121 47.79546095925 14.486909871843 47.795316765818 14.486842606486 47.794716562245
14.486577039909 47.794333124532 14.48640971347 47.794109093818 14.486317995633 47.79314143092
... DELETED TO SAVE SPACE
14.482888324227 47.795999913313 14.483027831415 47.795688098256 14.48434299101 47.795638080474 14.484576914821
47.795619276679 14.484929032443 47.795641126016 14.485264683424 47.795679353564 14.485581998303 47.795815396905
14.486105771698 47.795867794348 14.486355297457 47.795882110369 14.486724168738 47.795866525545 14.487113747121
</gml:posList>
              </gml:LinearRing>
            </gml:exterior>
          </gml:Polygon>
        </gml:geometryMember>
        <gml:geometryMember>
          <gml:Point gml:id="LTER_EU_AT_008_POINT" srsName="http://www.opengis.net/def/crs/EPSG/0/4326">
            <gml:pos srsDimension="2">47.772360000000 14.392820000000</gml:pos>
          </gml:Point>
```

```
      </gml:geometryMember>
    </gml:MultiGeometry>
  </ef:geometry>
  <ef:onlineResource>
https://data.lter-europe.net/deims/site/49515dda-1198-4013-8f43-c33e107af081
</ef:onlineResource>
  <ef:onlineResource>http://www.kalkalpen.at/system/web/default.aspx</ef:onlineResource>
  <ef:onlineResource>
http://www.kalkalpen.at/system/web/
</ef:onlineResource>
  <ef:purpose xlink:href="http://inspire.ec.europa.eu/codelist/PurposeOfCollectionValue" />
  <ef:observingCapability>
    <ef:ObservingCapability gml:id="ObservingCapability_LTER_EU_AT_008_5afac0fbde951">
      <ef:observingTime xsi:nil="true" nilReason="missing" />
      <ef:processType xsi:nil="true" nilReason="missing" />
      <ef:resultNature xsi:nil="true" nilReason="missing" />
      <ef:procedure nilReason="missing" />
      <ef:featureOfInterest nilReason="missing" />
      <ef:observedProperty _text="ecosystem measure" />
    </ef:ObservingCapability>
  </ef:observingCapability>
  <ef:observingCapability>
    <ef:ObservingCapability gml:id="ObservingCapability_LTER_EU_AT_008_5afac0fbde9b7">
      <ef:observingTime xsi:nil="true" nilReason="missing" />
      <ef:processType xsi:nil="true" nilReason="missing" />
      <ef:resultNature xsi:nil="true" nilReason="missing" />
      <ef:procedure nilReason="missing" />
      <ef:featureOfInterest nilReason="missing" />
      <ef:observedProperty _text="biological measure" />
    </ef:ObservingCapability>
  </ef:observingCapability>
  <ef:observingCapability>
    <ef:ObservingCapability gml:id="ObservingCapability_LTER_EU_AT_008_5afac0fbdea06">
      <ef:observingTime xsi:nil="true" nilReason="missing" />
      <ef:processType xsi:nil="true" nilReason="missing" />
      <ef:resultNature xsi:nil="true" nilReason="missing" />
      <ef:procedure nilReason="missing" />
      <ef:featureOfInterest nilReason="missing" />
      <ef:observedProperty _text="atmospheric measure" />
    </ef:ObservingCapability>
  </ef:observingCapability>
  <ef:observingCapability>
    <ef:ObservingCapability gml:id="ObservingCapability_LTER_EU_AT_008_5afac0fbdea53">
      <ef:observingTime xsi:nil="true" nilReason="missing" />
      <ef:processType xsi:nil="true" nilReason="missing" />
      <ef:resultNature xsi:nil="true" nilReason="missing" />
      <ef:procedure nilReason="missing" />
      <ef:featureOfInterest nilReason="missing" />
      <ef:observedProperty _text="experimental measure" />
    </ef:ObservingCapability>
  </ef:observingCapability>
  <ef:observingCapability>
    <ef:ObservingCapability gml:id="ObservingCapability_LTER_EU_AT_008_5afac0fbdeaa0">
      <ef:observingTime xsi:nil="true" nilReason="missing" />
      <ef:processType xsi:nil="true" nilReason="missing" />
      <ef:resultNature xsi:nil="true" nilReason="missing" />
      <ef:procedure nilReason="missing" />
      <ef:featureOfInterest nilReason="missing" />
      <ef:observedProperty _text="landscape measure" />
    </ef:ObservingCapability>
  </ef:observingCapability>
  <ef:observingCapability>
    <ef:ObservingCapability gml:id="ObservingCapability_LTER_EU_AT_008_5afac0fbdeaec">
      <ef:observingTime xsi:nil="true" nilReason="missing" />
      <ef:processType xsi:nil="true" nilReason="missing" />
      <ef:resultNature xsi:nil="true" nilReason="missing" />
      <ef:procedure nilReason="missing" />
      <ef:featureOfInterest nilReason="missing" />
      <ef:observedProperty _text="soil measure" />
    </ef:ObservingCapability>
  </ef:observingCapability>
```

```
  <ef:observingCapability>
    <ef:ObservingCapability gml:id="ObservingCapability_LTER_EU_AT_008_5afac0fbdeb38">
      <ef:observingTime xsi:nil="true" nilReason="missing" />
      <ef:processType xsi:nil="true" nilReason="missing" />
      <ef:resultNature xsi:nil="true" nilReason="missing" />
      <ef:procedure nilReason="missing" />
      <ef:featureOfInterest nilReason="missing" />
      <ef:observedProperty _text="water measure" />
    </ef:ObservingCapability>
  </ef:observingCapability>
  <ef:observingCapability>
    <ef:ObservingCapability gml:id="ObservingCapability_LTER_EU_AT_008_5afac0fbdeb85">
      <ef:observingTime xsi:nil="true" nilReason="missing" />
      <ef:processType xsi:nil="true" nilReason="missing" />
      <ef:resultNature xsi:nil="true" nilReason="missing" />
      <ef:procedure nilReason="missing" />
      <ef:featureOfInterest nilReason="missing" />
      <ef:observedProperty xlink:title="" />
    </ef:ObservingCapability>
  </ef:observingCapability>
  <ef:broader xlink:href="https://data.lter-europe.net/deims/node/8612/emf">
    <ef:Hierarchy gml:id="Facility_d0a8da18-0881-4ebe-bccf-bc4cb4e25701">
      <ef:linkingTime>
        <gml:TimePeriod gml:id="timePeriod01">
          <gml:beginPosition>1960-01-01</gml:beginPosition>
          <gml:endPosition indeterminatePosition="now" />
        </gml:TimePeriod>
      </ef:linkingTime>
    </ef:Hierarchy>
  </ef:broader>
  <ef:hasObservation xlink:title="Kalkalpen National Park - Combined soil type-depth map"
xlink:href="https://data.lter-europe.net/deims/node/10325/iso19139" />
  <ef:hasObservation xlink:title="Kalkalpen National Park - Corine Land Cover 2006"
xlink:href="https://data.lter-europe.net/deims/node/10306/iso19139" />
  <ef:hasObservation xlink:title="Kalkalpen National Park - Corine Land Cover 2012"
xlink:href="https://data.lter-europe.net/deims/node/10319/iso19139" />
  <ef:hasObservation xlink:title="Kalkalpen National Park - Corine Land Cover Changes 2006-2012"
xlink:href="https://data.lter-europe.net/deims/node/10320/iso19139" />
  <ef:hasObservation xlink:title="Kalkalpen National Park - EUNIS Habitat map" xlink:href="https://data.lter-
europe.net/deims/node/10321/iso19139" />
  <ef:hasObservation xlink:title="Kalkalpen National Park - Forest structure (Arial photo interpretation)"
xlink:href="https://data.lter-europe.net/deims/node/10328/iso19139" />
  <ef:hasObservation xlink:title="Kalkalpen National Park - Habitat and biotop map"
xlink:href="https://data.lter-europe.net/deims/node/10322/iso19139" />
  <ef:hasObservation xlink:title="Kalkalpen National Park - Road network" xlink:href="https://data.lter-
europe.net/deims/node/10334/iso19139" />
  <ef:hasObservation xlink:title="Kalkalpen National Park - Running surface waters"
xlink:href="https://data.lter-europe.net/deims/node/10331/iso19139" />
  <ef:hasObservation xlink:title="Kalkalpen National Park - Standing surface waters"
xlink:href="https://data.lter-europe.net/deims/node/10332/iso19139" />
  <ef:hasObservation xlink:title="Kalkalpen National Park - Wetlands" xlink:href="https://data.lter-
europe.net/deims/node/10333/iso19139" />
  <ef:involvedIn>
    <ef:EnvironmentalMonitoringActivity gml:id="Data_collection_activity5a18453e-8016-4d02-8bd3-
e6a7802b6698_5afac0fbdebe6">
      <gml:description />
      <gml:identifier codeSpace="Original images" />
      <gml:name>
Kalkalpen National Park (Austria) - Aerophoto 2009/2010/2013
</gml:name>
      <ef:activityTime>
        <gml:TimePeriod gml:id="Data_collection_activity_TimePeriod5a18453e-8016-4d02-8bd3-
e6a7802b6698_5afac0fbdec25">
          <gml:beginPosition>2016-01-01</gml:beginPosition>
          <gml:endPosition />
        </gml:TimePeriod>
      </ef:activityTime>
      <ef:activityConditions />
      <ef:responsibleParty>
        <base2:RelatedParty>
          <base2:individualName>
```

```
            <gco:CharacterString>Franziska Pöpperl</gco:CharacterString>
          </base2:individualName>
        </base2:RelatedParty>
      </ef:responsibleParty>
      <ef:inspireId>
        <base:Identifier>
          <base:localId>5a18453e-8016-4d02-8bd3-e6a7802b6698</base:localId>
          <base:namespace>https://data.lter-europe.net/deims</base:namespace>
        </base:Identifier>
      </ef:inspireId>
      <ef:onlineResource>https://data.lter-europe.net/deims/node/10338</ef:onlineResource>
    </ef:EnvironmentalMonitoringActivity>
  </ef:involvedIn>
  <ef:involvedIn>
    <ef:EnvironmentalMonitoringActivity              gml:id="Data_collection_activityc632b61a-9437-4d77-b5b7-
d1c344b40031_5afac0fbdec6b">
      <gml:description>biological measure</gml:description>
      <gml:identifier codeSpace="Infestation" />
      <gml:name>Kalkalpen National Park (Austria) - Bark Beetle</gml:name>
      <ef:activityTime>
        <gml:TimePeriod                 gml:id="Data_collection_activity_TimePeriodc632b61a-9437-4d77-b5b7-
d1c344b40031_5afac0fbdeca9">
          <gml:beginPosition>2016-01-01</gml:beginPosition>
          <gml:endPosition />
        </gml:TimePeriod>
      </ef:activityTime>
      <ef:activityConditions>
Gis layers with the management points over the national park from 2010 - 2015
</ef:activityConditions>
      <ef:responsibleParty>
        <base2:RelatedParty>
          <base2:individualName>
            <gco:CharacterString>Franziska Pöpperl</gco:CharacterString>
          </base2:individualName>
        </base2:RelatedParty>
      </ef:responsibleParty>
      <ef:inspireId>
        <base:Identifier>
          <base:localId>c632b61a-9437-4d77-b5b7-d1c344b40031</base:localId>
          <base:namespace>https://data.lter-europe.net/deims</base:namespace>
        </base:Identifier>
      </ef:inspireId>
      <ef:onlineResource>https://data.lter-europe.net/deims/node/10340</ef:onlineResource>
    </ef:EnvironmentalMonitoringActivity>
  </ef:involvedIn>
  <ef:involvedIn>
    <ef:EnvironmentalMonitoringActivity              gml:id="Data_collection_activity45cd080e-2661-4a71-9648-
fd8196501a9b_5afac0fbdecec">
      <gml:description>biological measure</gml:description>
      <gml:identifier codeSpace="Vegetation structure and species cover" />
      <gml:name>
Kalkalpen National Park (Austria) - Biotope mapping
</gml:name>
      <ef:activityTime>
        <gml:TimePeriod                 gml:id="Data_collection_activity_TimePeriod45cd080e-2661-4a71-9648-
fd8196501a9b_5afac0fbded2a">
          <gml:beginPosition>2016-01-01</gml:beginPosition>
          <gml:endPosition />
        </gml:TimePeriod>
      </ef:activityTime>
      <ef:activityConditions />
      <ef:responsibleParty>
        <base2:RelatedParty>
          <base2:individualName>
            <gco:CharacterString>Franziska Pöpperl</gco:CharacterString>
          </base2:individualName>
        </base2:RelatedParty>
      </ef:responsibleParty>
      <ef:inspireId>
        <base:Identifier>
          <base:localId>45cd080e-2661-4a71-9648-fd8196501a9b</base:localId>
```

```
          <base:namespace>https://data.lter-europe.net/deims</base:namespace>
        </base:Identifier>
      </ef:inspireId>
      <ef:onlineResource>https://data.lter-europe.net/deims/node/10341</ef:onlineResource>
    </ef:EnvironmentalMonitoringActivity>
  </ef:involvedIn>
  <ef:involvedIn>
    <ef:EnvironmentalMonitoringActivity              gml:id="Data_collection_activity217f59ce-9120-4881-b724-
bb46035fc380_5afac0fbded6d">
      <gml:description />
      <gml:identifier codeSpace="Disturbance and desaster events" />
      <gml:name>Kalkalpen National Park (Austria) - Dynamic</gml:name>
      <ef:activityTime>
        <gml:TimePeriod                   gml:id="Data_collection_activity_TimePeriod217f59ce-9120-4881-b724-
bb46035fc380_5afac0fbdedab">
          <gml:beginPosition>2016-01-01</gml:beginPosition>
          <gml:endPosition />
        </gml:TimePeriod>
      </ef:activityTime>
      <ef:activityConditions>Dynamic processes like fire and avalanches</ef:activityConditions>
      <ef:responsibleParty>
        <base2:RelatedParty>
          <base2:individualName>
            <gco:CharacterString>Franziska Pöpperl</gco:CharacterString>
          </base2:individualName>
        </base2:RelatedParty>
      </ef:responsibleParty>
      <ef:inspireId>
        <base:Identifier>
          <base:localId>217f59ce-9120-4881-b724-bb46035fc380</base:localId>
          <base:namespace>https://data.lter-europe.net/deims</base:namespace>
        </base:Identifier>
      </ef:inspireId>
      <ef:onlineResource>https://data.lter-europe.net/deims/node/10339</ef:onlineResource>
    </ef:EnvironmentalMonitoringActivity>
  </ef:involvedIn>
  <ef:involvedIn>
    <ef:EnvironmentalMonitoringActivity              gml:id="Data_collection_activitybc6cc12f-1882-4cad-a0c9-
b0979b768144_5afac0fbdedee">
      <gml:description />
      <gml:identifier codeSpace="Vegetation structure and species cover" />
      <gml:name>Kalkalpen National Park (Austria) - Forest</gml:name>
      <ef:activityTime>
        <gml:TimePeriod                   gml:id="Data_collection_activity_TimePeriodbc6cc12f-1882-4cad-a0c9-
b0979b768144_5afac0fbdee2c">
          <gml:beginPosition>2016-01-01</gml:beginPosition>
          <gml:endPosition />
        </gml:TimePeriod>
      </ef:activityTime>
      <ef:activityConditions />
      <ef:responsibleParty>
        <base2:RelatedParty>
          <base2:individualName>
            <gco:CharacterString>Franziska Pöpperl</gco:CharacterString>
          </base2:individualName>
        </base2:RelatedParty>
      </ef:responsibleParty>
      <ef:inspireId>
        <base:Identifier>
          <base:localId>bc6cc12f-1882-4cad-a0c9-b0979b768144</base:localId>
          <base:namespace>https://data.lter-europe.net/deims</base:namespace>
        </base:Identifier>
      </ef:inspireId>
      <ef:onlineResource>https://data.lter-europe.net/deims/node/10301</ef:onlineResource>
    </ef:EnvironmentalMonitoringActivity>
  </ef:involvedIn>
  <ef:involvedIn>
    <ef:EnvironmentalMonitoringActivity              gml:id="Data_collection_activityb1d83441-2c49-4d79-9f35-
a11aa3f8ecce_5afac0fbdee6f">
      <gml:description>landscape measure, topography</gml:description>
      <gml:identifier codeSpace="Geology and Geomorphology" />
```

```
      <gml:name>
Kalkalpen National Park (Austria) - Geology and geomorphology
</gml:name>
      <ef:activityTime>
        <gml:TimePeriod                   gml:id="Data_collection_activity_TimePeriodb1d83441-2c49-4d79-9f35-
a11aa3f8ecce_5afac0fbdeead">
          <gml:beginPosition>2016-09-26</gml:beginPosition>
          <gml:endPosition />
        </gml:TimePeriod>
      </ef:activityTime>
      <ef:activityConditions>GIS maps of geology and geomorphology</ef:activityConditions>
      <ef:responsibleParty>
        <base2:RelatedParty>
          <base2:individualName>
            <gco:CharacterString>Franziska Pöpperl</gco:CharacterString>
          </base2:individualName>
        </base2:RelatedParty>
      </ef:responsibleParty>
      <ef:inspireId>
        <base:Identifier>
          <base:localId>b1d83441-2c49-4d79-9f35-a11aa3f8ecce</base:localId>
          <base:namespace>https://data.lter-europe.net/deims</base:namespace>
        </base:Identifier>
      </ef:inspireId>
      <ef:onlineResource>https://data.lter-europe.net/deims/node/10303</ef:onlineResource>
    </ef:EnvironmentalMonitoringActivity>
  </ef:involvedIn>
  <ef:involvedIn>
    <ef:EnvironmentalMonitoringActivity              gml:id="Data_collection_activityf54a0688-5e1f-4b1a-9bd8-
cb2f9133e976_5afac0fbdeef0">
      <gml:description />
      <gml:identifier codeSpace="Social systems characteristics" />
      <gml:name>
Kalkalpen National Park (Austria) - Human infrastructure
</gml:name>
      <ef:activityTime>
        <gml:TimePeriod                   gml:id="Data_collection_activity_TimePeriodf54a0688-5e1f-4b1a-9bd8-
cb2f9133e976_5afac0fbdef2e">
          <gml:beginPosition>2016-09-26</gml:beginPosition>
          <gml:endPosition />
        </gml:TimePeriod>
      </ef:activityTime>
      <ef:activityConditions>GIS-layers of roads, buildings and parking places</ef:activityConditions>
      <ef:responsibleParty>
        <base2:RelatedParty>
          <base2:individualName>
            <gco:CharacterString>Franziska Pöpperl</gco:CharacterString>
          </base2:individualName>
        </base2:RelatedParty>
      </ef:responsibleParty>
      <ef:inspireId>
        <base:Identifier>
          <base:localId>f54a0688-5e1f-4b1a-9bd8-cb2f9133e976</base:localId>
          <base:namespace>https://data.lter-europe.net/deims</base:namespace>
        </base:Identifier>
      </ef:inspireId>
      <ef:onlineResource>https://data.lter-europe.net/deims/node/10300</ef:onlineResource>
    </ef:EnvironmentalMonitoringActivity>
  </ef:involvedIn>
  <ef:involvedIn>
    <ef:EnvironmentalMonitoringActivity              gml:id="Data_collection_activity932158a9-42a9-47f8-ada4-
543a5f3a52d9_5afac0fbdef71">
      <gml:description>landscape measure, land cover, landscape cover</gml:description>
      <gml:identifier codeSpace="Current land cover and habitats" />
      <gml:name>Kalkalpen National Park (Austria) - Land cover</gml:name>
      <ef:activityTime>
        <gml:TimePeriod                   gml:id="Data_collection_activity_TimePeriod932158a9-42a9-47f8-ada4-
543a5f3a52d9_5afac0fbdefaf">
          <gml:beginPosition>1990-01-01</gml:beginPosition>
          <gml:endPosition />
        </gml:TimePeriod>
```

```
      </ef:activityTime>
      <ef:activityConditions />
      <ef:responsibleParty>
        <base2:RelatedParty>
          <base2:individualName>
            <gco:CharacterString />
          </base2:individualName>
        </base2:RelatedParty>
      </ef:responsibleParty>
      <ef:inspireId>
        <base:Identifier>
          <base:localId>932158a9-42a9-47f8-ada4-543a5f3a52d9</base:localId>
          <base:namespace>https://data.lter-europe.net/deims</base:namespace>
        </base:Identifier>
      </ef:inspireId>
      <ef:onlineResource>https://data.lter-europe.net/deims/node/10297</ef:onlineResource>
    </ef:EnvironmentalMonitoringActivity>
  </ef:involvedIn>
  <ef:involvedIn>
    <ef:EnvironmentalMonitoringActivity              gml:id="Data_collection_activity9f44f6c7-33b6-461e-a0c8-
7792e1965394_5afac0fbdf01d">
      <gml:description>
soil measure, soil properties, thickness of soil horizon
</gml:description>
      <gml:identifier codeSpace="Soil map" />
      <gml:name>Kalkalpen National Park (Austria) - Soil</gml:name>
      <ef:activityTime>
        <gml:TimePeriod                     gml:id="Data_collection_activity_TimePeriod9f44f6c7-33b6-461e-a0c8-
7792e1965394_5afac0fbdf05b">
          <gml:beginPosition>2004-01-01</gml:beginPosition>
          <gml:endPosition />
        </gml:TimePeriod>
      </ef:activityTime>
      <ef:activityConditions>Modelled combined soil depth-soil type map</ef:activityConditions>
      <ef:responsibleParty>
        <base2:RelatedParty>
          <base2:individualName>
            <gco:CharacterString>Franziska Pöpperl</gco:CharacterString>
          </base2:individualName>
        </base2:RelatedParty>
      </ef:responsibleParty>
      <ef:inspireId>
        <base:Identifier>
          <base:localId>9f44f6c7-33b6-461e-a0c8-7792e1965394</base:localId>
          <base:namespace>https://data.lter-europe.net/deims</base:namespace>
        </base:Identifier>
      </ef:inspireId>
      <ef:onlineResource>https://data.lter-europe.net/deims/node/10298</ef:onlineResource>
    </ef:EnvironmentalMonitoringActivity>
  </ef:involvedIn>
  <ef:involvedIn>
    <ef:EnvironmentalMonitoringActivity              gml:id="Data_collection_activitye12f8253-608f-47e6-9125-
cf47831b9358_5afac0fbdf09f">
      <gml:description>landscape measure, elevation, topography</gml:description>
      <gml:identifier codeSpace="Topography" />
      <gml:name>Kalkalpen National Park (Austria) - Topographie</gml:name>
      <ef:activityTime>
        <gml:TimePeriod                     gml:id="Data_collection_activity_TimePeriode12f8253-608f-47e6-9125-
cf47831b9358_5afac0fbdf0dd">
          <gml:beginPosition>2011-01-01</gml:beginPosition>
          <gml:endPosition />
        </gml:TimePeriod>
      </ef:activityTime>
      <ef:activityConditions>
Airborne Laserscanning data (Elevation, surface model)
</ef:activityConditions>
      <ef:responsibleParty>
        <base2:RelatedParty>
          <base2:individualName>
            <gco:CharacterString>Franziska Pöpperl</gco:CharacterString>
          </base2:individualName>
```

```
        &lt;/base2:RelatedParty&gt;
      &lt;/ef:responsibleParty&gt;
      &lt;ef:inspireId&gt;
        &lt;base:Identifier&gt;
          &lt;base:localId&gt;e12f8253-608f-47e6-9125-cf47831b9358&lt;/base:localId&gt;
          &lt;base:namespace&gt;https://data.lter-europe.net/deims&lt;/base:namespace&gt;
        &lt;/base:Identifier&gt;
      &lt;/ef:inspireId&gt;
      &lt;ef:onlineResource&gt;https://data.lter-europe.net/deims/node/10302&lt;/ef:onlineResource&gt;
    &lt;/ef:EnvironmentalMonitoringActivity&gt;
  &lt;/ef:involvedIn&gt;
  &lt;ef:involvedIn&gt;
    <ef:EnvironmentalMonitoringActivity              gml:id="Data_collection_activityca3a98c7-82ac-43a5-b72d-
b45ce2817bec_5afac0fbdf120">
      &lt;gml:description /&gt;
      &lt;gml:identifier codeSpace="Running water (River and streams)" /&gt;
      &lt;gml:name&gt;Kalkalpen National Park (Austria) - Water&lt;/gml:name&gt;
      &lt;ef:activityTime&gt;
        <gml:TimePeriod                    gml:id="Data_collection_activity_TimePeriodca3a98c7-82ac-43a5-b72d-
b45ce2817bec_5afac0fbdf15e">
          &lt;gml:beginPosition&gt;2016-09-26&lt;/gml:beginPosition&gt;
          &lt;gml:endPosition /&gt;
        &lt;/gml:TimePeriod&gt;
      &lt;/ef:activityTime&gt;
      &lt;ef:activityConditions&gt;Spatial layer of running water&lt;/ef:activityConditions&gt;
      &lt;ef:responsibleParty&gt;
        &lt;base2:RelatedParty&gt;
          &lt;base2:individualName&gt;
            &lt;gco:CharacterString /&gt;
          &lt;/base2:individualName&gt;
        &lt;/base2:RelatedParty&gt;
      &lt;/ef:responsibleParty&gt;
      &lt;ef:inspireId&gt;
        &lt;base:Identifier&gt;
          &lt;base:localId&gt;ca3a98c7-82ac-43a5-b72d-b45ce2817bec&lt;/base:localId&gt;
          &lt;base:namespace&gt;https://data.lter-europe.net/deims&lt;/base:namespace&gt;
        &lt;/base:Identifier&gt;
      &lt;/ef:inspireId&gt;
      &lt;ef:onlineResource&gt;https://data.lter-europe.net/deims/node/10299&lt;/ef:onlineResource&gt;
    &lt;/ef:EnvironmentalMonitoringActivity&gt;
  &lt;/ef:involvedIn&gt;
  &lt;ef:involvedIn&gt;
    <ef:EnvironmentalMonitoringActivity              gml:id="Data_collection_activityd72bb1f7-cd95-4dc2-b3a1-
dd7f43ea4c8b_5afac0fbdf1a3">
      &lt;gml:description /&gt;
      &lt;gml:identifier codeSpace="Vegetation structure and species cover" /&gt;
      &lt;gml:name&gt;
Kalkalpen National Park (Austria) - Wild animal management
&lt;/gml:name&gt;
      &lt;ef:activityTime&gt;
        <gml:TimePeriod                    gml:id="Data_collection_activity_TimePeriodd72bb1f7-cd95-4dc2-b3a1-
dd7f43ea4c8b_5afac0fbdf1e1">
          &lt;gml:beginPosition&gt;2016-01-01&lt;/gml:beginPosition&gt;
          &lt;gml:endPosition /&gt;
        &lt;/gml:TimePeriod&gt;
      &lt;/ef:activityTime&gt;
      &lt;ef:activityConditions /&gt;
      &lt;ef:responsibleParty&gt;
        &lt;base2:RelatedParty&gt;
          &lt;base2:individualName&gt;
            &lt;gco:CharacterString&gt;Franziska Pöpperl&lt;/gco:CharacterString&gt;
          &lt;/base2:individualName&gt;
        &lt;/base2:RelatedParty&gt;
      &lt;/ef:responsibleParty&gt;
      &lt;ef:inspireId&gt;
        &lt;base:Identifier&gt;
          &lt;base:localId&gt;d72bb1f7-cd95-4dc2-b3a1-dd7f43ea4c8b&lt;/base:localId&gt;
          &lt;base:namespace&gt;https://data.lter-europe.net/deims&lt;/base:namespace&gt;
        &lt;/base:Identifier&gt;
      &lt;/ef:inspireId&gt;
      &lt;ef:onlineResource&gt;https://data.lter-europe.net/deims/node/10342&lt;/ef:onlineResource&gt;
```

```
    </ef:EnvironmentalMonitoringActivity>
  </ef:involvedIn>
  <ef:representativePoint>
    <gml:Point gml:id="LTER_EU_AT_008_CENTROID" srsName="http://www.opengis.net/def/crs/EPSG/0/4326">
      <gml:pos srsDimension="2">47.772360000000 14.392820000000</gml:pos>
    </gml:Point>
  </ef:representativePoint>
  <ef:measurementRegime
xlink:href="http://inspire.ec.europa.eu/codeList/MeasurementRegimeValue/continuousDataCollection" />
  <ef:mobile>false</ef:mobile>
  <ef:operationalActivityPeriod xsi:nil="false">
    <ef:OperationalActivityPeriod gml:id="operationalActivityPeriod_49515dda-1198-4013-8f43-c33e107af081">
      <ef:activityTime>
        <gml:TimePeriod gml:id="timePeriod_49515dda-1198-4013-8f43-c33e107af081">
          <gml:beginPosition>1993-01-01</gml:beginPosition>
          <gml:endPosition indeterminatePosition="now" />
        </gml:TimePeriod>
      </ef:activityTime>
    </ef:OperationalActivityPeriod>
  </ef:operationalActivityPeriod>
  <ef:belongsTo xlink:href="https://data.lter-europe.net/deims/node/222">
    <ef:NetworkFacility gml:id="Network_e3d5faf7-5603-4acd-b835-eaedca8e5dec">
      <gml:name>Austria (LTER-Austria)</gml:name>
      <ef:linkingTime>
        <gml:TimePeriod gml:id="timePeriod_e3d5faf7-5603-4acd-b835-eaedca8e5dec">
          <gml:beginPosition>2012-10-22</gml:beginPosition>
          <gml:endPosition indeterminatePosition="now" />
        </gml:TimePeriod>
      </ef:linkingTime>
    </ef:NetworkFacility>
  </ef:belongsTo>
  <ef:belongsTo>
    <ef:NetworkFacility gml:id="Network_6b0abbb7-b970-45cd-b52a-9b1e4098f8b9">
      <gml:name>Ramsar</gml:name>
      <ef:linkingTime>
        <gml:TimePeriod gml:id="timePeriod_6b0abbb7-b970-45cd-b52a-9b1e4098f8b91">
          <gml:beginPosition>2012-10-22</gml:beginPosition>
          <gml:endPosition indeterminatePosition="now" />
        </gml:TimePeriod>
      </ef:linkingTime>
    </ef:NetworkFacility>
  </ef:belongsTo>
  <ef:belongsTo>
    <ef:NetworkFacility gml:id="Network_73a46951-a1bf-43fc-997e-69fa1c0e4da5">
      <gml:name>Natura 2000</gml:name>
      <ef:linkingTime>
        <gml:TimePeriod gml:id="timePeriod_73a46951-a1bf-43fc-997e-69fa1c0e4da51">
          <gml:beginPosition>2012-10-22</gml:beginPosition>
          <gml:endPosition indeterminatePosition="now" />
        </gml:TimePeriod>
      </ef:linkingTime>
    </ef:NetworkFacility>
  </ef:belongsTo>
  <ef:belongsTo>
    <ef:NetworkFacility gml:id="Network_f8deab13-795e-4087-814d-9088e05926f4">
      <gml:name>Nationalpark</gml:name>
      <ef:linkingTime>
        <gml:TimePeriod gml:id="timePeriod_f8deab13-795e-4087-814d-9088e05926f41">
          <gml:beginPosition>2012-10-22</gml:beginPosition>
          <gml:endPosition indeterminatePosition="now" />
        </gml:TimePeriod>
      </ef:linkingTime>
    </ef:NetworkFacility>
  </ef:belongsTo>
  <ef:belongsTo>
    <ef:NetworkFacility gml:id="Network_2b2136b0-0330-44aa-bbc3-15acaae3d534">
      <gml:name>IUCN National park</gml:name>
      <ef:linkingTime>
        <gml:TimePeriod gml:id="timePeriod_2b2136b0-0330-44aa-bbc3-15acaae3d5341">
          <gml:beginPosition>2012-10-22</gml:beginPosition>
          <gml:endPosition indeterminatePosition="now" />
```

```
        </gml:TimePeriod>
      </ef:linkingTime>
    </ef:NetworkFacility>
  </ef:belongsTo>
  <ef:belongsTo>
    <ef:NetworkFacility gml:id="Network_1a1330fe-3ef1-4698-8f03-69fd65141a3a">
      <gml:name>ENFORS</gml:name>
      <ef:linkingTime>
        <gml:TimePeriod gml:id="timePeriod_1a1330fe-3ef1-4698-8f03-69fd65141a3a1">
          <gml:beginPosition>2012-10-22</gml:beginPosition>
          <gml:endPosition indeterminatePosition="now" />
        </gml:TimePeriod>
      </ef:linkingTime>
    </ef:NetworkFacility>
  </ef:belongsTo>
  <ef:belongsTo>
    <ef:NetworkFacility gml:id="Network_a1fe4e57-1a8a-46b7-9478-450bfcca166b">
      <gml:name>GTOS</gml:name>
      <ef:linkingTime>
        <gml:TimePeriod gml:id="timePeriod_a1fe4e57-1a8a-46b7-9478-450bfcca166b1">
          <gml:beginPosition>2012-10-22</gml:beginPosition>
          <gml:endPosition indeterminatePosition="now" />
        </gml:TimePeriod>
      </ef:linkingTime>
    </ef:NetworkFacility>
  </ef:belongsTo>
  <ef:belongsTo>
    <ef:NetworkFacility gml:id="Network_d6d25a37-9b75-4c0d-b5ed-723d7e17190e">
      <gml:name>ECOPOTENTIAL (H2020)</gml:name>
      <ef:linkingTime>
        <gml:TimePeriod gml:id="timePeriod_d6d25a37-9b75-4c0d-b5ed-723d7e17190e1">
          <gml:beginPosition>2012-10-22</gml:beginPosition>
          <gml:endPosition indeterminatePosition="now" />
        </gml:TimePeriod>
      </ef:linkingTime>
    </ef:NetworkFacility>
  </ef:belongsTo>
  <ef:belongsTo>
    <ef:NetworkFacility gml:id="Network_8305d5a6-ec9e-4721-b1f8-1c48d309fe1b">
      <gml:name>LTER</gml:name>
      <ef:linkingTime>
        <gml:TimePeriod gml:id="timePeriod_8305d5a6-ec9e-4721-b1f8-1c48d309fe1b1">
          <gml:beginPosition>2012-10-22</gml:beginPosition>
          <gml:endPosition indeterminatePosition="now" />
        </gml:TimePeriod>
      </ef:linkingTime>
    </ef:NetworkFacility>
  </ef:belongsTo>
</ef:EnvironmentalMonitoringFacility>
```