

ExpeER

Distributed Infrastructure for EXPERimentation in Ecosystem Research

Grant Agreement Number: 262060

SEVENTH FRAMEWORK PROGRAMME

Capacities

Integrating activities: Networks of Research Infrastructures (RIs)

Theme: Environment and Earth Sciences

DELIVERABLE D3.3

Deliverable title: **Thesaurus for semantic interoperability of existing systems**

Abstract: The provision of common semantics for the description of datasets and their contents within a network of distributed data sources is a crucial service to any research infrastructure and network. Within the ExpeER project existing domain efforts were continued building a thesaurus for the ecosystem observation and experimentation community. The report provides a description of the underlying challenges, the main components and structure of the thesaurus, the links to other semantic sources, the governance for the development and the use of the controlled vocabulary within ExpeER and beyond.

Due date of deliverable: M48

Actual submission date: M54

Start date of the project: December 1st, 2010

Duration: 54 months

Organisation name of lead contractor: Umweltbundesamt GmbH (EAA)

Contributors: Schentz, H.; Watkins, J.; Bertrand, N.; Frenzel, M.; Kertész, M.; Grandin, U.; Stoll, St.; Pfaff, G.; Oggioni, A.; Damien M.; Zavattaro, L.; Sacco, D.; Danuso, F. & Peterseil, J.

Revision N°: V2.0

Dissemination level:

PU Public (must be available on the website)	<input checked="" type="checkbox"/>
PP Restricted to other programme participants (including the Commission Services)	<input type="checkbox"/>
RE Restricted to a group specified by the consortium (including the Commission Services)	<input type="checkbox"/>
CO Confidential, only for members of the consortium (including the Commission Services)	<input type="checkbox"/>

Revision N°	Date	Changes between different revision versions	Distribution (who have reviewed)
V1.0	17.04.2015	Basic version of the document	Schentz / Peterseil
V2.0	05.05.2015	Inclusion of final comments	Peterseil, Schentz, Watkins, Bertrand

Contact details on the authors and contributors

Name	Organisation	
<i>Schentz, Herbert</i> <i>Peterseil, Johannes</i> <i>Pfaff, Günter</i>	EAA	Umweltbundesamt GmbH, Ecosystem research and Environmental Information Management Spittelauer Lände 5, 1090 Vienna, Austria
<i>Watkins, John</i> <i>Bertrand, Nic</i>	CEH	Centre for Ecology & Hydrology, Lancaster Environment Centre Library Avenue, Bailrigg, LA1 4AP Lancaster, UK
<i>Frenzel, Mark</i>	UFZ	Helmholtz-Centre for Environmental Research (UFZ), Department Biozönoseforschung Theodor-Lieser-Str. 4, 06120 Halle, Germany
<i>Kertész, Miklós</i>	IEB	MAGYAR TUDOMANYOS AKADEMIA OKOLOGIAI KUTATOKOZPONT KleBELSberg Kuno utca 3., 8237 Tihany, Hungary
<i>Grandin, Ulf</i>	SLU	Swedish University of Agricultural Sciences (SLU), Department of Aquatic Sciences and Assessment P.O. Box 7070, SE-750 07 Uppsala, Sweden
<i>Stoll, Stefan</i>	Senckenberg	Senckenberg Research Institute and Natural History Museum Frankfurt, Abteilung Fließgewässerökologie und Naturschutzforschung (Senckenberg Gesellschaft für Naturforschung) Clamecystraße 12, 63571 Gelnhausen, Germany
<i>Oggioni, Alessandro</i>	CNR	CONSIGLIO NAZIONALE DELLE RICERCHE (CNR), Institute for Electromagnetic Sensing of the Environment (IREA) Via Bassini 15, I-20133 Milano, Italy
<i>Damien Maurice</i>	INRA	INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE (INRA), Forest Ecology and Ecophysiology (INRA – Nancy) Allée de l'Arboretum, 54280 Champenoux, France
<i>Zavattaro, Laura</i> <i>Sacco, Dario</i>	UNITO	UNIVERSITA DEGLI STUDI DI TORINO (DISAFA) Largo Braccini 2, 10095 Grugliasco, Italy
<i>Danuso, Francesco</i>	UNIUD	Università degli Studi di Udine via Palladio 8, 33100 Udine, Italy

Acknowledgment

Special acknowledgement needs to be given to all additional contributors to the discussions on building the EnvThes thesauri not listed here in the list of authors. Especially people contributing to the ExpeER and EnvEurope project as well as people in the related research infrastructures and networks like LTER Europe, AnaEE/AnaEE-France, LifeWatch and ILTER.

Table of Content

1	Introduction.....	5
2	Use cases and applications of a common semantic.....	6
2.1	Controlled terms as keywords.....	6
2.2	Controlled terms to correct, annotate and publish data	7
2.3	Machine readable references	8
2.4	Establishment of interlinkages to other controlled vocabularies	9
3	Design principles of EnvThes	9
3.1	Foundational vocabularies	9
3.2	Requirements for extensions	10
3.3	Main structure of revised EnvThes.....	11
3.4	Building an interlinked thesaurus.....	13
4	Workflow and governance	14
4.1	Governance	14
4.2	Editing rules.....	15
4.3	Workflows	16
5	Use of the thesaurus	17
5.1	Keyword source in DEIMS Research Site and Dataset Registry.....	17
5.2	Lookup of Terms and SPARQL endpoint.....	18
5.3	Semantic Annotation of data elements and datasets.....	20
5.4	Use by ILTER	22
6	Conclusions and outlook	23
7	References.....	24
8	Glossary	26
	Annex 1 EnvThes – Access and Management Interfaces	29
	User Guide for EnvThes	30
	Editors Guide for EnvThes	33
	SPARQL Endpoint.....	41
	Annex 2 – Documentation of work	47
	Workshops.....	47
	The ILTER workshop	50
	Skype conferences.....	52

Executive summary

The long term ecological research community as well as the experimental community need to reuse and integrate data and knowledge, which so far are generated and managed by many different institutes with differing IT systems, differing data structures and semantics. Inter alia, therefore, a strong effort has to be undertaken to disambiguate the meaning of the data by the use of controlled vocabularies for metadata and data management.

To cover these aspects a common semantic source for ExpeER was developed integrating already existing work for this domain. This thesaurus, an extension of the existing EnvThes developed by EnvEurope, was established to (a) serve as source for keywords in the metadata management system DEIMS and (b) as controlled vocabulary for properties, methods, units, objects of interest and others within the data management systems. In order to cover the needs of the long term experimentation and observation communities, inter-alia the following topics need to be covered:

- Ecosystem properties
- Ecological experiments and treatments
- Agricultural and forestry management practices
- Ecological observation and sampling strategies
- Ecosystem, habitat and species catalogues

The development of this thesaurus did not start from the scratch, but was based on EnvThes which, itself has been created by extension of already existing US-LTER vocabulary, QUDT ontology (NASA Sweet Ontology – Units), EUNIS Habitat list, INSPIRE spatial data themes. In a first attempt species names from the Catalogue of Life (CoL) have been included. Because of performance issues it was decided, that the CoL should be an own SKOS/RDF project rather than a part of the thesaurus, or in case, the CoL team establishes its own Linked Data access for the CoL, simply use this interface.

A simple governance and workflow has been established, developed in intensive cooperation within an editor's group, working with face to face meetings and skype conferences. The basic principles and guidelines for the development and implementation for EnvThes are:

- (1) Implementation of a *production* and *development* version of EnvThes. The production version is a released copy from the development version, which represents the work in progress.
- (2) The root elements of the thesaurus need to have a clear domain relevant structure resembling the basic elements of ISO 19156 Observation and Measurements (O&M), in order to make use of the controlled vocabulary together with data structured according to O&M easier.
- (3) The atoms of compound terms should be within the thesaurus and linked to the compound terms by *SKOS:related* concept relation.
- (4) It should not be attempted to implement additional complex relations which cannot be modelled by the basic SKOS relationships within EnvThes
- (5) Basic terms should be spelled in singular and written in lower case.

The development of a controlled vocabulary is a community effort driven by their needs. Therefore continuous developments and adaptations by the community are necessary. Basic criteria for governance of the thesaurus were developed by the thesaurus working group and need to be implemented into the landscape of existing research infrastructures.

D3.3 Thesaurus for semantic interoperability of existing systems

1 Introduction

In order to improve environmental management and to support the development and implementation of targeted environmental policies at European scale, information on the status and trends as well as on the expected processes and reactions of ecosystems towards pressures is needed. To provide a sound basis for these analysis, ecological monitoring and long-term studies (including experiments) of ecological systems need a shared scientific basis as well as harmonised methodologies addressing the different ecosystem components and processes. This can only be reached through an integrated approach addressing common objectives and research topics, providing sufficient resources and include the relevant disciplines. Research infrastructures like AnaEE¹ (Analysis and Experimentation on Ecosystems) or LTER Europe² (European Long-Term ecological Research network) are addressing these aspects by streamlining and harmonizing monitoring procedures and experimental approaches across the different ecosystem domains. This work is supported in addition by scientific projects (e.g. ExpeER³ or EnvEurope⁴) tackling specific subjects in this respect.

A broad variety of data is generated by different research sites and communities adding up to a common data pool on ecosystem status and processes. Despite harmonisation efforts by the different research infrastructures and networks, still a lot of heterogeneity exists across the borders of different scientific domains or the different research infrastructures. In addition to the variety of disciplines involved, data management still show a high level of diversity ranging from single data centres to dislocated data storages at single institutes or research groups. Also different data formats, descriptions or reference lists are used by the different communities.

In order to allow for common analysis of data resulting from the 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 establishment 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).

The common semantic is established on the basis of existing controlled vocabularies extending them to the domain requirements identified by the different long term observation and experimentation projects.

The current report gives an overview on the work performed within the ExpeER project (FP7, 2010-2015) dealing with the implementation of common data management principles. The report is organised according to the following structure:

Chapter 2 gives a short overview on the use cases for the extended controlled vocabulary in the context of the project and beyond.

Chapter 3 describes the basic design principles and structures which were applied to construct the hierarchy of terms and allow for interlinking them to other semantic sources.

Chapter 4 describes the governance structure and workflows on the establishment and work of the controlled vocabulary.

¹ See <http://www.anaee.com/>

² See <http://www.lter-europe.net/>

³ See <http://www.expeeronline.eu/>

⁴ See <http://www.enveurope.eu/>

Chapter 5 describes the use and implementation of EnvThes within the current project as well the intended applications.

Chapter 6 provides a summary and outlook to further work planned in terms of the further development of controlled vocabularies and ontologies.

2 Use cases and applications of a common semantic

To integrate and interpret data from different sources a common language describing the content – a common semantic – is needed. This can either be implicitly by the interpretation of the data analyst or explicit by the use of a common controlled vocabulary or shared reference lists. Figure 1 shows the schematic links between the data and a common semantic. These links can either be made by simple links or by the use of annotations.

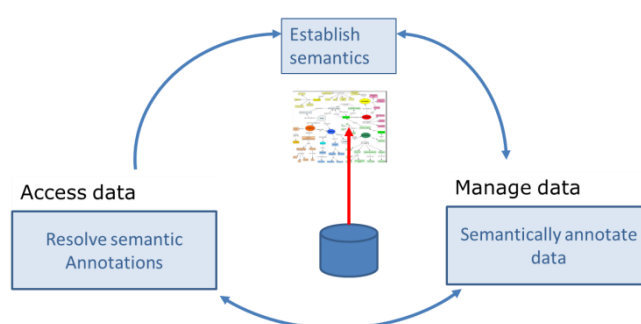


Figure 1 Conceptual schema on the use of controlled vocabulary for semantic annotation

For the use of common terms defined in a controlled vocabulary different levels can be defined, which are described in detail in this section:

- (1) Using keywords within metadata to describe the content of the data, datasets or data elements.
- (2) Using controlled vocabulary and shared reference lists within data records (e.g. species reference list) or
- (3) Using controlled vocabularies to annotate 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 which are described in Chapter 3 for EnvThes.

2.1 Controlled terms as keywords

A keyword is generally defined as a word 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 information⁵. 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

⁵ See <http://dictionary.reference.com/browse/keyword> (accessed 2015-04-17)

things, diverging spellings of the same name 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.

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. Chapter 5 gives some insights on that.

Currently EnvThes, the thesaurus of ExpeER, serves as source for the keywords of metadata managed by the metadata platform DEIMS⁶, thus allowing a cross dataset search.

ABSTRACT
Abstract: <p>Dataset provides pH measured in Lake Santo Parmense in 2012. Water samples were collected by means of a Ruttner bottle at different depths over the whole water column at the point of maximum depth of the lake. pH was measured by means of a pH meter.</p>
KEYWORDS
EnvThes Keywords: lakes LTER Site ph limnology
ACCESS&USE CONSTRAINTS
Principal and granted permission:

Figure 2 Keyword section of described dataset in DEIMS (e.g. <http://data.lter-europe.net/deims/dataset/c9d42534-7f95-11e2-b67b-005056ab003f>)

2.2 Controlled terms to correct, annotate and publish data

Exactly defined controlled vocabularies for the exchange of scientific information, as e.g. the table of chemical 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 lists 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 online available, 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.

⁶ See <http://data.lter-europe.net/deims/>

2.3 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/RDF⁷ 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 (Protocol and RDF Query Language) 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.

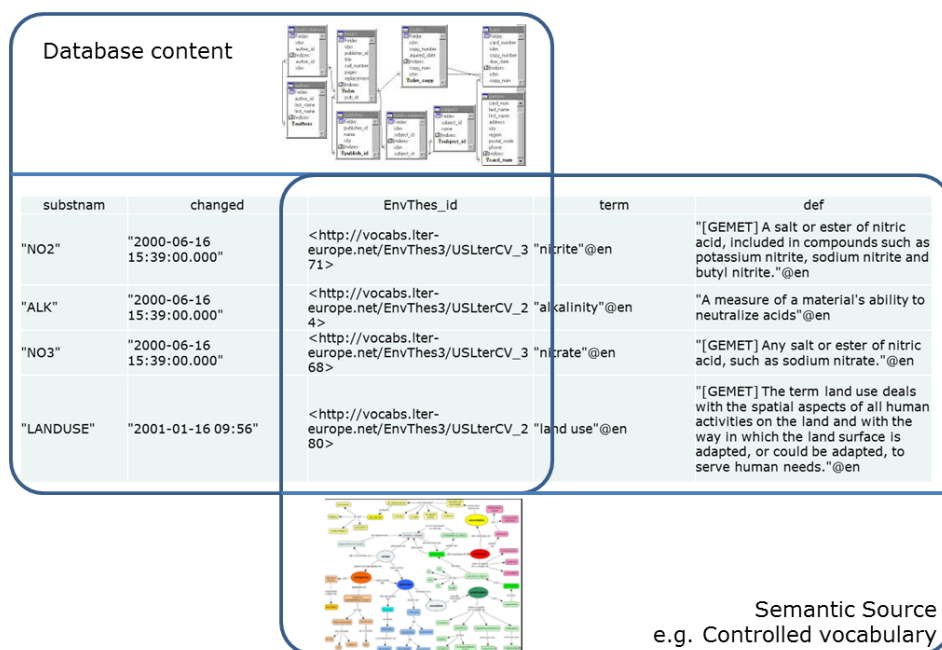


Figure 3 Linking the database content with definitions coming from a controlled vocabulary

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.

⁷ RDF – Resource description Framework (W3C Standard)

2.4 Establishment of interlinkages 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. These links, if exposed as Linked Data SPARQL endpoint can be accessed and queried through the data integration process.

For the ecosystem observation and experimental community links (partial integrations) to concepts of the following vocabularies are of interest:

- GEMET (<https://www.eionet.europa.eu/gemet/>) - the official controlled vocabulary of the EEA
- EUROVOC (<http://eurovoc.europa.eu/>) - the official controlled vocabulary of the European commission.
- AGROVOC (<http://aims.fao.org/vest-registry/vocabularies/agrovoc-multilingual-agricultural-thesaurus>) - the most comprehensive vocabulary, with translations in multiple languages for a lot of disciplines around the topic “food”. It is the vocabulary of the FAO.
- QUDT ontology (<http://www.qudt.org/>) - a comprehensive vocabulary for units and dimensions first established for the NASA.
- Several code lists of the EEA and the INSPIRE registry

It has to be pointed out, that those links must be able to be followed and must be machine readable.

3 Design principles of EnvThes

One of the main aims of the data management and integration activities of ExpeER was to bridge between existing silos of data resulting from different kind of research, observation and experimental activities. In order to achieve this goal, a common semantic basis is one of the first steps to be taken. It is important that the generated or used common semantic is not an isolated silo of concepts itself but is interlinked with existing sources of concepts and definitions used in the domain.

Therefore it was decided to build on existing vocabularies of the ecological and experimental community and extend them, where needed. Existing vocabularies were screened on the one side for their fitness for use and on the other hand a gap analysis was made identifying gaps and the need for further semantic work in the domain for long term ecosystem observation and experimentation.

3.1 Foundational vocabularies

A foundation for the ExpeER thesaurus was found in the thesaurus developed by the LTER community within the EnvEurope project, as this thesaurus already had many elements which were shared with other controlled vocabularies, covered a large conceptual part of the LTER domain and even some conceptual models of the experimental domain.

The EnvEurope thesaurus contained already imported parts from other controlled vocabularies which were taken over by ExpeER and extended by concepts, which so far have been missing but were needed by the ExpeER community.

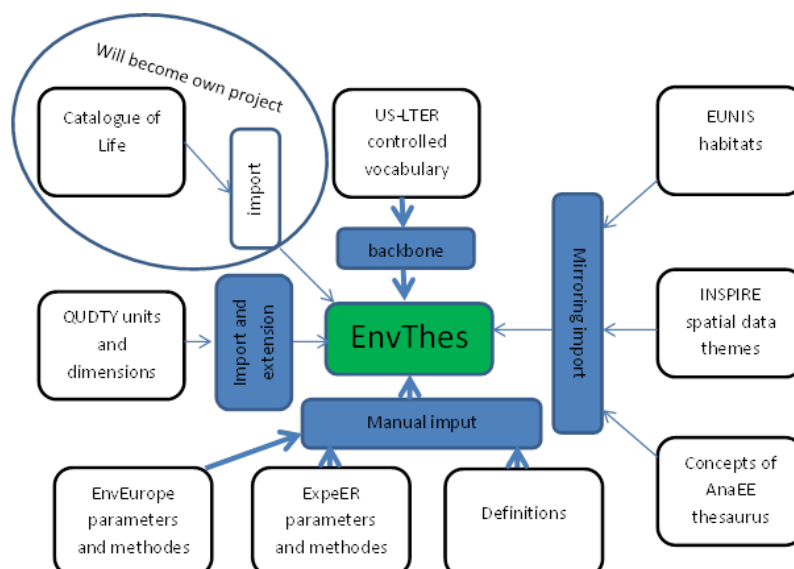


Figure 4 Foundational vocabularies of EnvThes

Figure 4 shows an overview on the different vocabularies interlinked within the work on EnvThes. These specifically are:

- **US-LTER vocabulary** (<http://vocab.lternet.edu/vocab/vocab/sobre.php>), which served as backbone vocabulary of EnvThes. As EnvThes has been extended, mainly within the ExpeER project, the US LTER vocabulary just provides about 50 % of the terms in the common controlled vocabulary.
- **EUNIS Habitats** (<http://eunis.eea.europa.eu/habitats-code-browser.jsp>) have been mirrored from the EEA List providing entries down to level 3 of EUNIS.
- **INSPIRE spatial data themes** (http://www.eionet.europa.eu/gemet/inspire_themes?langcode=en); although being on a rather coarse level, they are important as key words for all datasets, that have to be provided according to the INSPIRE Metadata regulation.
- **Units and dimensions**, as defined in QUDT ontology (<http://www.qudt.org/>) are comprehensive and served as a basis for the UNITS of EnvThes. Some ecology - specific units had to be added.

To date, the tools and services used within the architecture do not allow direct references to multiple vocabularies and therefore they have to be imported into one repository (as a local cache). However, it is important to establish links to the original vocabulary, where ever possible, for two reasons: (a) to proof the provenance of the concept and (b) to prepare the establishment of links within architectures that allow for references to multiple sources.

3.2 Requirements for extensions

EnvThes, in the status, as it was developed within the EnvEurope project, covered the basic concepts, needed by the LTER-Europe community. The granularity within the concepts of parameters (e.g. observed properties), however was coarse and concepts, needed for the description of ecosystem experiments, as e.g. needed by AnaEE-France⁸, were widely missing. These missing extensions were the main contribution of the ExpeER project by a joined effort of the long term experimental and observation community.

⁸ See <http://www.anaee-s.fr/>

3.2.1 Concepts for ecosystem observations

ECOPAR, an application, developed within EnvEurope and contributed by the ExpeER community, is based on the Ecological integrity framework facilitating the assignment of parameters to ecosystem patterns and processes, inter alia listing the observations that need to be done and the parameters that have to be determined. As EnvThes aims to cover all the concepts of the ecological community represented in the ExpeER project, it has become necessary to enrich the existing concepts by those, which are represented in ECOPAR. Combing through the terms within ECOPAR the EnvThes editing team found out, that it was necessary to check for synonyms and homonyms before adding terms which were missing in the thesaurus.

3.2.2 Concepts for ecosystem experiments

Right from the beginning of ExpeER steps for enriching EnvThes by concepts specific for experiments have been carried out, mainly by selecting concepts of the AnaEE-France thesaurus⁹, which should be entered into EnvThes (e.g. agriculture – specific terms like “fertilizer”, “landuse practices”). This covered mainly the experimental treatments of the sampling plots, which were not in the main focus of the long term observation community. In addition information on agricultural and forestry practices (e.g. fertilisation, coppicing, etc.) was needed to be added to EnvThes in order to describe the experimental system and design.

Gap analysis and work planning was done within common workshops in order to achieve a comparable granularity of the concepts.

As described above for terms coming from ECOPAR, terms within AnaEE-France thesaurus had to undergo three checks: (1) a string matching check searching for AnaEE-France thesaurus concepts, which are already present in EnvThes, (2) a check for existing synonyms to avoid duplicates, and (3) search for fitting broader terms correctly

3.3 Main structure of revised EnvThes

During the work on EnvThes the editing group found out, that in addition to the controlled vocabulary within the thesaurus, an ontology for meaningful relations between the concepts would be helpful. Thus the concepts of EnvThes with clearly defined links could become instances within the ontology. ISO 19156, observation and measurements (O&M) is a widely used core ontology for ecological observations, and therefore was envisaged as a possible basis for an ecological ontology which has yet to be created. Adapting the structure of the root concepts to ISO 19156 has been carried out alongside the EnvThes development (Leadbetter & Vodden 2015). This has informed the basic hierarchy and use of *SKOS:related* links in EnvThes. Table 1 gives an overview on the resulting root structure. As EnvThes is work in progress, the current version of the main structure, the root elements (concepts) of the thesaurus, represent a snap shot of the state of March 2015.

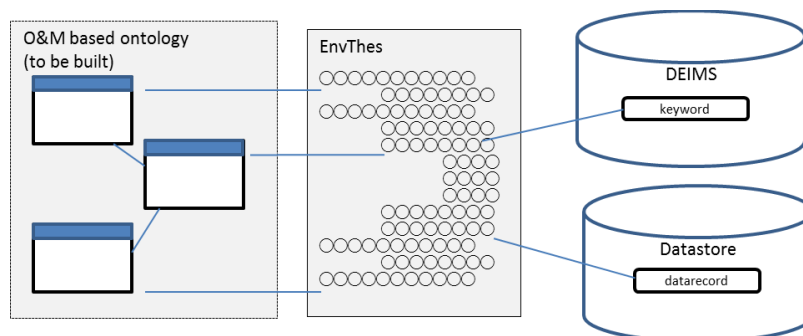


Figure 5 Links between Ontology, EnvThes, metadata keywords and data records

⁹ See http://rbdd.cnrs.fr/IMG/pdf/presentation_cpichot.pdf?98/b5855d90c511117d8637c52c1bddbd088e684eaf

Table 1 Overview on root concepts of EnvThes

root concept	definition	comment
matrix	The medium, in which the actual <i>Object Of Interest</i> is mixed, embedded, suspended, or entailed in some way - such that a process is required to extract or separate it before an observation can be made. (en)	can sometimes overlap with object of interest
measure	An Observable Property is a compound concept formed by determining an Observed Property for an Object Of Interest - for example the pH (<i>Observed Property</i>) of a sample of stream water (the Object Of Interest). Observable Properties may optionally have associated Units Of Measure, a Matrix, and various Constraints and Statistical Measures. (en)	the most important thing of measure (and other compound terms) is the fact, that with their entrance to the ExpeER controlled vocabulary, the way, how they are composed is fixed. (even if you cannot always establish a composition rule)
object of interest	The substance, taxon or other physical/chemical phenomenon of the Feature Of Interest that is being observed. E.g. waves, rainfall, Calluna vulgaris, Aluminium. Concept Source: CSIRO Observable Properties ontology http://environment.data.gov.au/def/op (en)	
property	The property or characteristic of the <i>Object Of Interest</i> that is observed during the act of observation. E.G. If the temperature of the water is measured, then the Observed Property is Temperature. Concept Source: INSPIRE Observations & Measurements (Complex Properties Model) http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/D2.9_O&M_Guidelines_v2.0.pdf (en)	There are properties which need not end up in a measure
statistical function	A statistical measure qualifies an Observable Property and represents some aggregation or derivation E.g. Mean, Mode, Median, Weekly mean, Monthly mean. Ontologically, Statistical Measures may be derived from other Statistical Measures. They may have complex attribution as well as a descriptive label. see http://uncertml.org Concept Source: INSPIRE Observations & Measurements (Complex Properties Model) http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/D2.9_O&M_Guidelines_v2.0.pdf (en)	
system classification	ecosystem, layers, spheres, etc	
discipline	A discipline is knowledge or wisdom associated with one academic field of study or profession. A discipline incorporates types of knowledge, expertise, skills, people, projects, communities, problems, challenges, studies, inquiry, approaches, and research areas that are strongly associated with academic areas of study or areas of professional practice (en)	this is of special importance for the metadata system DEIMS
event	Something that happens and modifies the system. It represents a discontinuity in time.	
method	A way of proceeding or doing something, especially a systematic or regular one. (en)	
organizational concept	container of concepts needed for the metadata management system DEIMS	important for all sorts of organisational classifications
units and dimension	based on units and dimensions of the QUDT ontology (http://www.qudt.org/)	
tool	machine, device	
unstructured concepts	container for concepts, where the position in the hierarchy of the thesaurus is not yet clear or decided	
deprecated terms	container of concepts, that might be referenced in the past, but should not be used anymore	

3.4 Building an interlinked thesaurus

As already stated in Chapter 2.4, it is important to avoid the generation of semantic silos when communities are just on the way to overcome their data silos. Avoiding the EnvThes silo is done by the establishment of links to other controlled vocabularies.

Those links are implemented by the use of SKOS/RDF xx-matches¹⁰:

- SKOS:exactMatch
- SKOS:closeMatch
- SKOS:relatedMatch
- SKOS:broaderMatch
- SKOS:narrowerMatch

Most of the relations are established as exact matches. Concepts that have been imported from sources which provide resolvable unique identifiers always have an *SKOS:exactMatch* link back to those sources. This was done e.g. for EUNIS Habitats, INSPIRE spatial data themes, QUDTY units and dimensions.

As GEMET is the official Thesaurus of the European Environment Agency (EEA), links to that vocabulary help to know, whether terms are compliant with EEA. The same compliance aspect is true for links to EuroVoc, the official thesaurus of the European Commission. AGROVOC is the most comprehensive controlled vocabulary in the field of agriculture and is developed and hosted by the FAO, therefore used globally. A link to Wikipedia and later also to dbpedia helps for getting extensive explanations and pictures and could be used in the future for getting integrated into the Linked Data cloud.

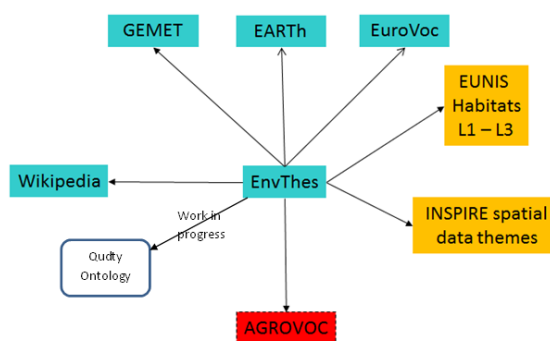


Figure 6 Links from EnvThes to other controlled vocabularies

It is important to point out, that not every concept can have links to all sources.

Links are established in a two-step process: (a) look for string matches followed by the second step (b) review of found matching strings by an expert.

¹⁰ See <http://www.w3.org/2004/02/skos/>

4 Workflow and governance

A simple guidance and workflow has been established, developed in intensive cooperation within an editor's group, working with face to face meetings and skype conferences. The basic guidelines for the development and implementation for EnvThes are:

- (1) Implementation of a *production* and *development* version of EnvThes. The production version is a released copy from the development version, which represents the work in progress.
- (2) The root elements of the thesaurus need to have a clear domain relevant structure resembling the basic elements of ISO 19156 Observation and Measurements (O&M), in order to make use of the controlled vocabulary together with data structured according to O&M easier.
- (3) The atoms of compound terms should be within the thesaurus and linked to the compound terms by SKOS:related concept relation.
- (4) It should not be attempted to implement additional complex relations which cannot be modelled by the basic SKOS relationships within EnvThes.
- (5) Basic terms should be spelled in singular and written in small letters.

The work on EnvThes is still ongoing, as it will be the semantic basis for the further work within the long term experimentation and observation communities.

4.1 Governance

EnvThes is a community effort which can only be done by the input and the contribution of different members of the community. To ensure the use and the long term maintenance in the frame of long term ecosystem research and monitoring a clear but simple guidance has been established.

EnvThes is a **free and open thesaurus** for the domain of long term ecosystem research and monitoring including all related domains (e.g. biodiversity, agriculture, forestry, etc.). It can be used by everyone but acknowledgement is needed.

EnvThes is a collaborative effort of supporting projects to the research infrastructures of AnaEE/AnaEE-France and LTER Europe. The project EnvEurope and ExpeER contributed to the establishment and development of EnvThes. As common working bodies are currently not defined, existing working groups, e.g. LTER Europe Expert Panel on Information Management, are suggested to steer the work until organisational structures, e.g. joint working groups, are established.

Therefore, the following organisational structure is suggested and has been established:

- **"EnvThes"** as a working group within the Expert Panel on Information Management of LTER Europe joint with external experts. The formation of a joint expert working group is discussed within WP4 of ExpeER.
- **"Editor Team EnvThes (ET)"** has the role to coordinate the work on EnvThes and make final decisions on the content. The "Editor Team EnvThes" consists of 5-9 (actually 6) persons including the facilitator. The members of the "Editor Team EnvThes" are fixed and can be changed on an annual basis within the EP IM meeting at the LTER Europe conference.
- **"Collaboration Team EnvThes"** who is contributing to the content of EnvThes. Every scientist and expert can join the Collaboration Team EnvThes by in-kind contribution.
- **"Technical Coordinator EnvThes (TC)"** which is named as person. The technical hosting of EnvThes is located at CEH (UK) with Nic Bertrand and Mike Wilson as the contact points.

EnvThes has to be integrative and therefore a clear link and permanent relation and communication with relevant networks need to be ensured (e.g. US LTER).

The following tasks will have to be carried out by the organisational units:

- Editor Team EnvThes
 - Develop a maintenance concept for the EnvThes working group in LTER Europe / ILTER based on the experience of the ExpeER project
 - Editorial work on the EnvThes
 - Coordinate the further development of EnvThes and organise user inputs
 - Coordinate requests from different networks or projects and sustain the core structure and focus of EnvThes.
 - Coordinate and streamline the work with ongoing activities of the semantic group in US LTER
 - Coordinate and streamline the work with ongoing activities of the semantic group in ILTER
- Technical Coordinator EnvThes
 - Management of the content (e.g. import)
 - Support work on the EnvThes
- Hosting
 - EnvThes is currently hosted by CEH, thus supporting the long term experimental and observation community

The work on EnvThes will be done as in kind contribution. The work can further be financed by different projects (e.g. eLTER, LTER Europe, AnaEE, AnaEE-France, etc.) as well as formally linked to ILTER.

4.2 Editing rules

Within ExpeER a lot of work to refine EnvThes was done, including restructuring of the concepts according to the following rules:

- (1) No attempt shall be made, to create ontology-like structures (meaningful relations, e.g. *'observed by'*) within the thesaurus, by (mis)using SKOS-relations *"SKOS:broader"*, *"SKOS:narrower"*, and *"SKOS:related"* for encoding (implicitly) more meaningful relations.
- (2) To make annotation of data easier the main structure should be aligned with observation and measurement, the basic structure of INSPIRE.
- (3) Compound terms should, as much as possible, be separated into their atoms, which have to be within EnvThes and should have the link "related term" towards the compound term.
- (4) Concepts have to have a preferred label in singular. The only exceptions are concepts, where there is no singular.
- (5) EnvThes has English as a primary language, which means, that every concept has to have a preferred label in English, but may have preferred labels in other languages. If there is a definition, there must be a definition in English.
- (6) No concept of the productive version may be deleted. If a concept should not be used any more, it has to be moved to the deprecated concepts.
- (7) A concept may have a scope note that indicates where the concept can be used. So far the scope note indicates at least, whether the concept is derived from the US-LTER controlled vocabulary. In the near future there will also be a scope note for the usage in DEIMS, which will be necessary for DEIMS, running with DRUPAL 7.

4.3 Workflows

In order to allow for an extensible and adaptable procedure when updating EnvThes also beyond the runtime of the current projects, the partners in WP3 decided on a simple workflow. Figure 7 shows a schematic picture of the workflow ranging from the identification of gaps to the formulation of a proposal and the acceptance of the term.

For this the development version was created as a copy of the production version, which remains as stable version used e.g. in the DEIMS Research Site and Dataset Registry.

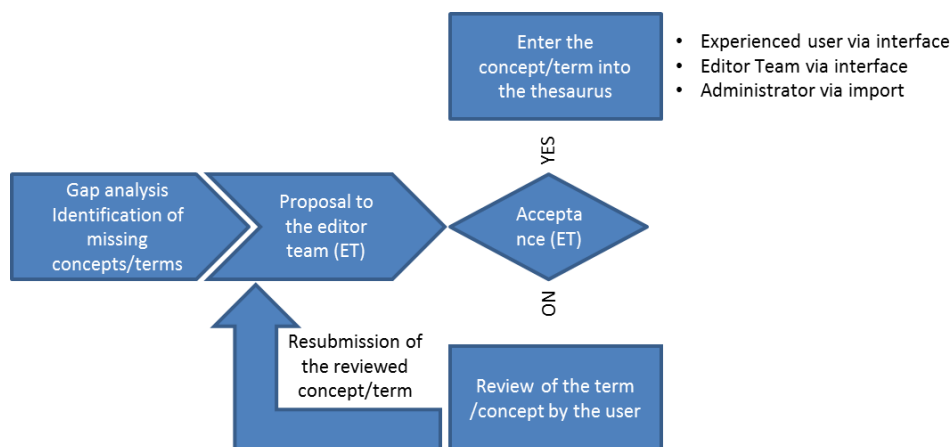


Figure 7 Simple workflow from detection of missing content (concept, definition, translation, link) to its entry

Whoever decides to contribute to the work on EnvThes and is accepted as contributor may look for concepts, definitions or translations that are missing to his/her opinion and propose changes to the Editor Team, who decides that the changes are done or rejected. Once the decision is taken, the changes take place, either by manual input by someone of the Editor Team, by the contributor if capable handling of poolParty, or by the import by the technical coordinator.

Within the editing process of EnvThes it proved to be most efficient to first distribute the draft work on portions of the thesaurus within the editing team, each person trying to look for a correct position within the hierarchy, trying to find or establish an appropriate definition, looking for synonyms and translations. In a second step the group discusses the first draft findings and either accepts them, changes details or establishes rules for the draft editor to revise the work.

In general it can be said, that the distributed individual work on concepts dominated over the common work and the common work (in face to face sessions or skype meetings) rather served as a review process.

As, for technical reasons, two instances of EnvThes exist, regular updates from the development version to the production version have to be carried out. This is done by replacement of the current production version by a new version of the thesaurus (see Figure 8). As concepts are not deleted, but referred to as "*deprecated terms*", links to existing concepts will not break. This workflow will be changed in future depending also on the technical possibilities.

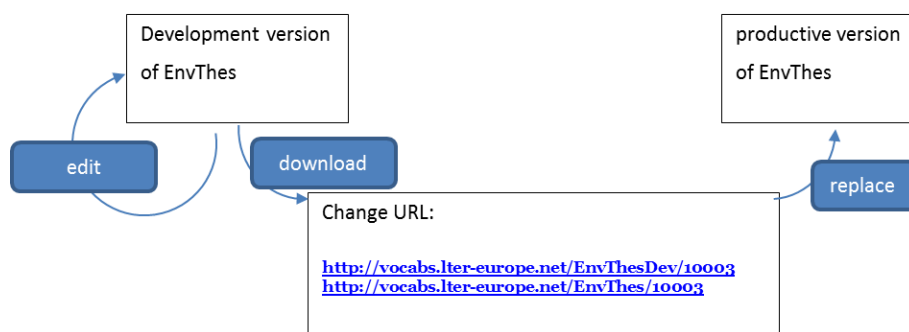


Figure 8 Relation between development version and production version of EnvThes

On the decision of the editor team, due to the resources of the technical coordinator, the development version is copied to the productive version changing the URL of the concepts, as shown in Figure 8.

5 Use of the thesaurus

The central metadata repository DEIMS manages metadata on Research sites, Datasets, Persons, Publications, and Networks. Metadata on datasets can be exported in EML and ISO format for the ILTER community and in INSPIRE format and are exposed via a CSW catalogue service.

This central metadata repository uses concepts of the thesaurus EnvThes for keywords. The concepts of EnvThes are also exposed via a Linked Data interface, offering also a SPARQL query interface.

5.1 Keyword source in DEIMS Research Site and Dataset Registry

The metadata platform DEIMS Research Site and Dataset Registry was developed to edit, manage and discover metadata on research sites, datasets and persons related to long term ecosystem observation and experimentation. Within the ExpeER project the DEIMS instance developed by the EnvEurope project was used and extended to the needs of ExpeER.

One of the main features is to provide information and keywords within the different modules of DEIMS, e.g. keywords for datasets or research topics for research sites. These different semantic terms are defined and described in the EnvThes thesaurus.

Currently there is just an offline link between the semantic source EnvThes and the metadata editor. This is due to the technical capabilities of the content management system (CMS) Drupal 6, which allows limited capabilities to allow direct linking of semantic sources. Therefore in the current version of DEIMS, which is based on Drupal 6, the thesaurus was to be imported to allow for a stable metadata editor interface. The thesaurus is updated when major updates of EnvThes are released. This is discussed and decided within the Expert Panel on Information Management and the Joint Working Groups.

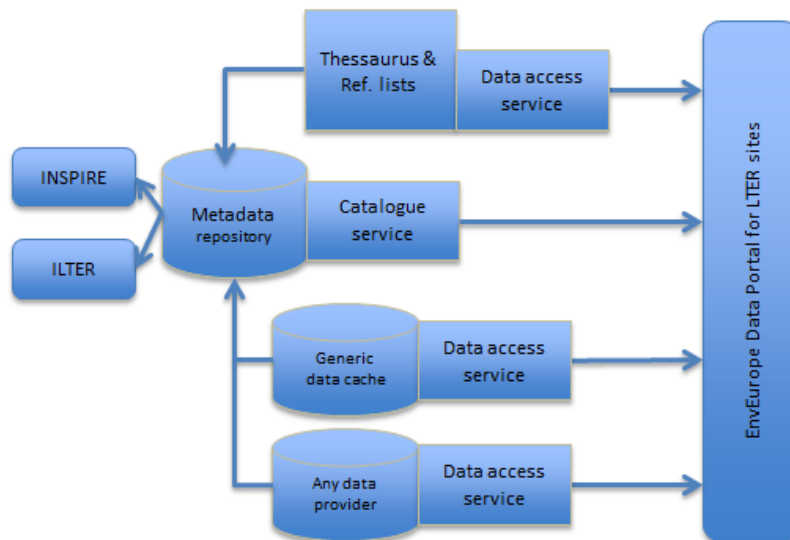


Figure 9 Basic architecture on the link between the DEIMS Research Site and Dataset Registry and Controlled Vocabularies based on an example from the EnvEurope logical system architecture

Inter-alia, for easier updates future versions of DEIMS will be upgraded to DRUPAL 7 allowing use of live links to Linked Data sources. It is planned to establish markers which serve as selection keys for special groups of concepts, thus reducing the available terms/concepts for certain selection fields of DEIMS.

5.2 Lookup of Terms and SPARQL endpoint

There are two possibilities to search and discover concepts on EnvThes. This is (a) the navigation in the hierarchical tree of the concepts and (b) to search using the autocomplete function. Both mechanisms provide a user friendly way to discover the content.

To discover a given concept by navigation in the hierarchical tree of concepts information on hierarchical relations are needed. Given the case, someone wants to know, which ecosystem types are covered by the controlled vocabulary, looking whether something like boreal forest or similar is provided. In such a case clicking down the hierarchy, e.g. *'ecosystems'* → *'terrestrial ecosystems'* → *'forest ecosystems'* → *'forest'* leads to *'boreal forest'*. Figure 10 gives an example of the web interface for this case.

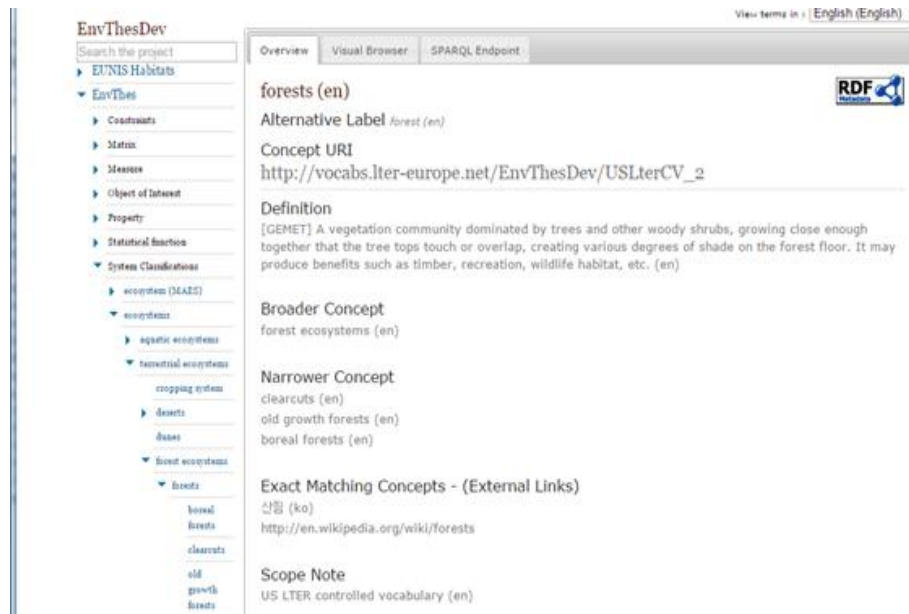


Figure 10 Example showing the results for a search with the help of the hierarchy

Alternatively a user can use the search functions of the web interface to discover the concepts and terms. In this case the search is supported by an autocomplete function. Given the case, a user wants to find the concept “forest” and all the concepts related to the concept “forest”. In this case the search word “forest” is entered to the search field and the query results in the details on the concepts showing also the related terms.

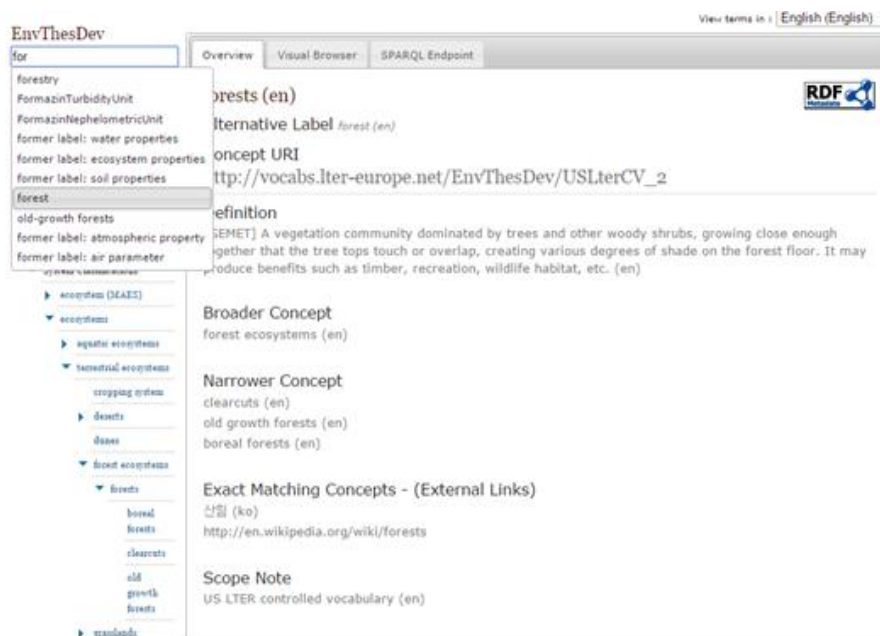


Figure 11 Example for alphabetical search using the autocomplete function of poolParty

Using the autocomplete function, after typing for (to look for forest) already shows all terms that contain the search word ‘for’. By clicking on the concept ‘forest’ in the list of results the user is forwarded to the detailed description of the selected concept.

For the establishment of user defined queries poolParty offers a SPARQL (W3C compliant query language) endpoint. That endpoint, which of course can just be used by advanced non IT users or by IT specialists, offers the possibility to make special searches and extractions of the thesaurus content. A more detailed description can be found in the Annex SPARQL Endpoint within this report.

5.3 Semantic Annotation of data elements and datasets

5.3.1 Developments initiated by the FP7 EUDAT project

Within the FP7 project EUDAT¹¹ generic infrastructure services for research infrastructures are developed in order to allow safe and efficient data management. Inter-alia the subject of missing semantics is addressed. A prototype for supporting end users in the workflow of semantic annotation was developed. The issue was raised, as very often data sets will be stored in repositories that come from various sources (including hardly standardized human observations) and may show a large heterogeneity with respect to the used vocabularies. They may even include typo errors, lack references to recognized domain knowledge sources (ontologies, thesauri, category lists), etc. When data are generated by crowd-sourcing, e.g. via unstructured information collected by smartphone devices, semantic analysis and tagging will become a crucial aspect to be solved.

Semantic annotation was identified as community service needed, since researchers and research communities will use the simple file store facilities for example to store unstructured data sets and to make them accessible. Without any supporting tools to increase the data quality and semantic description data reuse and sharing will be hampered by missing information.

Within the diagram below, the service would provide a generic tool that is anchored in the loop from users as data generators across communities that offer recognized knowledge sources, common services that help improving the data sets and the users as data consumers.

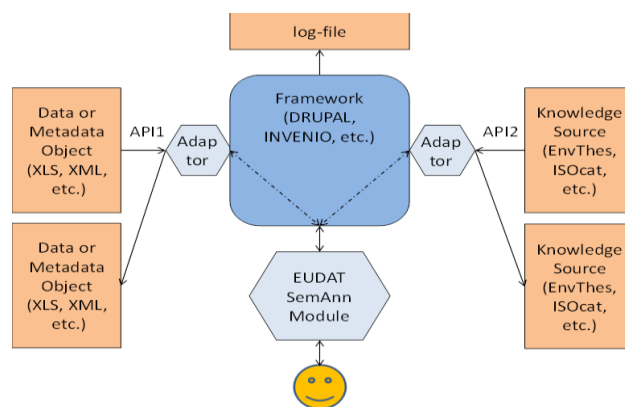


Figure 12 Elements for semantic annotation of metadata

EUDAT is working on a cross-disciplinary solution for (a) importing/storing data sets, (b) establishing references to recognized knowledge sources (“reference lists”) and (c) increasing the quality of the kind of observational data as indicated above - both of which are the results of the same look-up process. For a first pilot, a use case motivated from a concrete data flow provided by LTER/LifeWatch shall trigger the conceptualization and specification of services.

To simplify the approach, the following assumptions in the development process were made: (a) data sets do not have any graph-like structure in the first instance, but exist of a flat list of for example

¹¹ See <http://www.eudat.eu/>

key-value pairs and (b) users provide knowledge source data, structured in different form of flat lists, thesauri or even complex graphs that contain the terms to be accepted in the data sets.

Figure 13 shows the workflow implemented in a prototype with “user-friendly” tools to: (1) Import data in a generic format, (2) allow for indication of variables, which refer to existing knowledge sources (“reference lists”), (3) use acceptable terms and attributes from these specified knowledge source (by export or interactive links). (4) cross-check terms in specified data sets against specified knowledge sources to identify those not in compliance, (5) report to the data depositor about the analysis, and (6) support manual and (semi) automatic annotations to the entries such as adding references to the knowledge source.

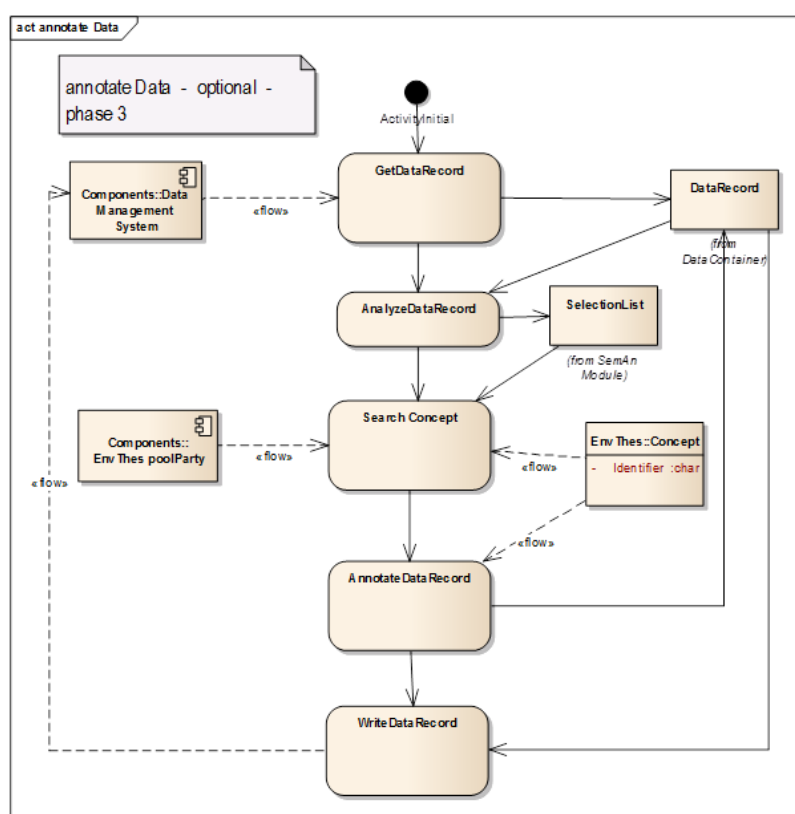


Figure 13 UML diagram showing the workflow and objects of semantic annotation

Within EUDAT the work was started on the subject providing a valuable tool set in the future. Linking the semantics to data and datasets will need support by tools.

5.3.2 Test of D2RQ Linked Data service

Within the framework of the EnvEurope project a test on linking semantics and data was carried out. This work was conceptually continued within the ExpeER project checking if this application would also fit to this community. Based on the logical system architecture which is the basis for the LTER data flows, a test with Linked Data services has been carried out.

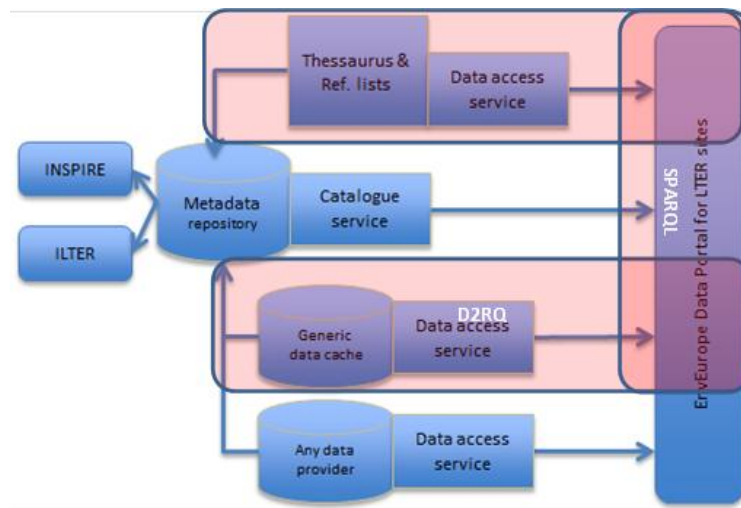


Figure 14 components used for testing the linked Data service D2RQ

Within that test, two versions of linkages between parameters in EnvThes and parameters in the data service have been carried out:

- linking by name (string match)
- linking by identifier, where the identifier has been inserted manually into the table “Parameter”

In both cases queries joining the data of the data services with definitions and descriptions in EnvThes and other vocabularies revealed interesting results. Whereas string matches allow links to any vocabulary, they can be ambiguous, especially when homonyms, terms with two diverting meanings, exist. Linking by identifiers on the other hand needs insertion processes into the storage of the data or into “association storages”.

The result of the test can be summarized as:

- having multilingual controlled vocabularies with good definitions allows use of the translations and definitions in data queries to semantically enrich the results.
- the link by identifiers is preferable, as it is the only way to achieve a precise, unambiguous connection between data and semantics.
- the establishment of the links has to be supported by tools in order to allow non IT specialists to do the work.
- the establishment of joining queries has to be supported by user friendly query tools.

5.4 Use by ILTER

ILTER (International Long Term Ecological Research) the world wide community of Long Term Ecological Researchers has the same harmonization needs as the LTER Europe community and therefore wants to use EnvThes as controlled vocabulary for metadata and data. EnvThes was adopted as ILTER thesaurus as a result of the proceedings of a workshop in Shanghai on dealing with issues with discovering metadata in multiple languages.

EnvThes is being developed as a multilingual effort and already uses terms and concepts of US LTER and other thesauri relevant to Long Term Ecosystem Research. It was therefore reasonable to adopt EnvThes as the backbone thesaurus for an ILTER Environmental Thesaurus.

6 Conclusions and outlook

The daily work of experimental and monitoring ecologists with their data and metadata has proven, that a common controlled vocabulary is of great benefit for interoperability and reuse of metadata and data.

As ecology is a science with a broad range of special sub-disciplines, it cannot be awaited, that structural and/or semantic interoperability can be established in one single step, but rather is a continuous work. Although comprehensive cross domain semantic interoperability nearly is impossible, common concepts can be found.

The formal language SKOS/RDF allows for meeting those issues, providing sufficient relations between concepts within a vocabulary and the *xxxxMATCH* relations to other vocabularies that can be followed. The decision to use this SKOS/RDF for EnvThes has proven right, and the establishment of links to other vocabularies like GEMET, EUROVOC, AGROVOC and EARTH helps to overcome semantic silos and will, of course be widened.

The reorganization of the root concepts, which has been carried out in order to be somehow compliant with ISO19156 (observation & measurement) is promising, but the usability has to be attested when EnvThes concepts are mapped to an ISO19156 compliant ontology and such an ontology is used for data publishing and data sharing.

Links cannot be established to other vocabularies and ontologies only, but also to metadata, data and publications, and all those links can be bidirectional thus opening a vast amount of pathways to data, making data discovery and data reuse easier. First steps to put those possibilities into practice have successfully been tested within the project EnvEurope, ExpeER and EUDAT. But there is still a lot to do in order to take advantage of the growing semantic interoperability of which the controlled vocabulary EnvThes is just a first step.

In future development of the EnvThes infrastructure, we intend to merge the development and production versions into one single system. This requires the implementation of the editing workflow within the host software so that new terms can be proposed, reviewed, amended and published within the host software. This will make the development of the vocabulary more accessible and auditable to its users.

7 References

- Adamescu, M., Peterseil, J., Dactu, S., Cazacu, C., Vadineanu, A. (2010). Elements for the design of a General Ecological Database. In: Maurer, I. and Tochtermann, K. (eds.) Information and Communication Technologies for Biodiversity and Agriculture. Shaker Verlag, Aachen. pp. 49-66.
- Berners-Lee, T. (2006), Design Issues: Linked Data. Online <http://www.w3.org/DesignIssues/LinkedData.html>
- Bizer, C. (et al) (2011) Linked Data: Evolving the Web into a Global Data Space, (1st edition). Synthesis Lectures on the Semantic Web: Theory and Technology, 1:1, 1-136. Morgan & Claypool. (Full text available for free at <http://linkeddatabook.com/editions/1.0/>).
- Bizer, C., Talis, T.H., Berners-Lee, T. (2009), Linked Data - The Story So Far. International Journal on Semantic Web and Information Systems (IJSWIS), 5(3), 1-22. DOI: 10.4018/jswis.2009081901
- European Environment Agency (1999) GEMET, GEneral Multilingual Envrionmental Thesaurus <http://www.eionet.europa.eu/gemet/>
- European Environment Agency (2009) European Nature Information System (EUNIS). <http://eunis.eea.europa.eu/about.jsp> and <http://eunis.eea.europa.eu/habitats.jsp>
- European Union (2013) EuroVoc Multilingual Thesaurus of the European Union. <http://eurovoc.europa.eu/>
- Food and Agriculture Organization of the United Nations (2013) AGROVOC. <http://aims.fao.org/standards/agrovoc/about>
- Gruber, T.R. (1993). Toward principles for the design of ontologies used for knowledge sharing. Originally in N. Guarino and R. Poli, (Eds.), International Workshop on Formal Ontology, Padova, Italy. Revised August 1993. International Journal of Human-Computer Studies 43(5-6):907-928.
- Isaac, A. (et al.) (2009) Simple Knowledge Organization System Primer. W3C Working Group Note 18 August 2009. <http://www.w3.org/TR/skos-primer/>
- Klyne G., Carroll J.J. (2004) Resource Description Framework (RDF): Concepts and Abstract Syntax. <http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/>
- Leadbetter, A.M. & Vodden. P.N. (2015). Semantic linking of complex properties, monitoring processes and facilities in web-based representations of the environment. International Journal of Digital Earth. (DOI: 10.1080/17538947.2015.1033483)
- LTER Controlled Vocabulary Working Group (2011) Terms of Reference. <http://im.lternet.edu/VocabTOR>
- Miles, A., Bechhofer S. (2009) SKOS Simple Knowledge Organization System Reference. <http://www.w3.org/TR/2009/REC-skos-reference-20090818/>
- Mirtl, M., Krauze, K. (2007). Developing a new Strategy for Environmental Research and Monitoring: The European Long-Term Ecological Research Network's (LTER Europe) role and perspective. In: Chmielewski, T.J. (Ed.) Nature Conservation Management: From Idea to practical Results. Lublin – Lodz – Helsinki – Aarhus. pp. 36-52.
- Mirtl, M., Orenstein, D., Wildenberg, M., Peterseil, J., Frenzel, M. (2013). Development of LTSE Platforms in LTER Europe: Challenges and Experiences in implementing Place-Based Long Term-Socio-ecological Research in Selected Regions. In: Singh, S.J., Haberl, H., Chertow, M., Mirtl, M. and Schmid, M. (Eds.). Long Term Socio-Ecological Research. Studies in Society-Nature Interactions Across Spatial and Temporal Scales. Springer , Dodrecht Heidelberg New York London. pp. 409-442.
- National Aeronautics and Space Administration (2011) Semantic Web for Earth and Environmental Terminology (SWEET) <http://sweet.jpl.nasa.gov/ontology/> and <http://www.qudt.org/qudt/owl/1.0.0/quantity/>
- National Centers for Biomedical Computing (2005) BioPortal. <http://bioportal.bioontology.org/>
- 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.
- poolParty <http://www.poolparty.biz/>
- Roskov Y., Abucay L., Orrell T., Nicolson D., Kunze T., Culham A., Bailly N., Kirk P., Bourgoin T., DeWalt R.E., Decock W., De Wever A., eds. (2015). Species 2000 & ITIS Catalogue of Life, 18th March 2015. Digital resource at www.catalogueoflife.org/col. Species 2000: Naturalis, Leiden, the Netherlands.
- Schentz, H., Schleidt, K., Lane, M. (2005). Requirements for an IT Framework for ALTER-Net – Results from the Requirment Analysis conducted in Work Package I6. ALTER-Net Deliverable WPI6_2005_01.

Abbildungsverzeichnis

<i>Figure 1 Conceptual schema on the use of controlled vocabulary for semantic annotation</i>	<i>6</i>
<i>Figure 2 Keyword section of described dataset in DEIMS (e.g. http://data.lter-europe.net/deims/dataset/c9d42534-7f95-11e2-b67b-005056ab003f)</i>	<i>7</i>
<i>Figure 3 Linking the database content with definitions coming from a controlled vocabulary</i>	<i>8</i>
<i>Figure 4 Foundational vocabularies of EnvThes</i>	<i>10</i>
<i>Figure 5 Links between Ontology, EnvThes, metadata keywords and data records</i>	<i>11</i>
<i>Figure 6 Links from EnvThes to other controlled vocabularies</i>	<i>13</i>
<i>Figure 7 Simple workflow from detection of missing content (concept, definition, translation, link) to its entry</i>	<i>16</i>
<i>Figure 8 Relation between development version and production version of EnvThes</i>	<i>17</i>
<i>Figure 10 Basic architecture on the link between the DEIMS Research Site and Dataset Registry and Controlled Vocabularies based on an example from the EnvEurope logical system architecture</i>	<i>18</i>
<i>Figure 11 Example showing the results for a search with the help of the hierarchy</i>	<i>19</i>
<i>Figure 12 Example for alphabetical search using the autocomplete function of poolParty</i>	<i>19</i>
<i>Figure 13 Elements for semantic annotation of metadata</i>	<i>20</i>
<i>Figure 14 UML diagram showing the workflow and objects of semantic annotation</i>	<i>21</i>
<i>Figure 15 components used for testing the linked Data service D2RQ</i>	<i>22</i>

8 Glossary

AGROVOC	Multilingual agricultural thesaurus (http://aims.fao.org/vest-registry/vocabularies/agrovoc-multilingual-agricultural-thesaurus)
AnaEE	Analysis and Experimentation on Ecosystems is a research infrastructure for experimental manipulation of managed and unmanaged terrestrial and aquatic ecosystems (http://www.anaee.com/)
AnaEE-France	The french contribution of the european project ANAEE (http://www.anaee-s.fr/)
CoL	Catalogue of Life is planned to become a comprehensive catalogue of all known species of organisms on <u>Earth</u> (http://www.catalogueoflife.org/)
CSIRO	Commonwealth Scientific and Industrial Research Organisation is the federal government agency for scientific research in <u>Australia</u> (http://www.csiro.au/en)
CSW	Web catalogue service is the OGC metadata service (http://www.opengeospatial.org/standards/cat)
D2RQ	The D2RQ Platform is a system for accessing relational databases as virtual, read-only RDF graphs. It offers RDF-based access to the content of relational databases without having to replicate it into an RDF store (http://d2rq.org/)
DEIMS	The DEIMS Research Site and Dataset Registry (abbreviation for Drupal ecological Information Management System) is the metadata editor and discovery portal for Research sites, Datasets, Persons and Publications building the long term observation and experimentation network in Europe. DEIMS is used by LTER Europe as the central catalogue for its site network (http://data.lter-europe.net/deims/)
DRUPAL	Content management System used for DEIMS. (https://www.drupal.org/)
EARTH	Environment Related Controlled Vocabulary (http://thesaurus.iaa.cnr.it/index.php/vocabularies/earth)
ECOPAR	ECOPAR is an interactive web tool based on a database of indicators, parameters and corresponding methods for ecosystem monitoring and research: (http://www.ufz.de/lter-d/index.php?en=32141&contentonly=1)
EML	Ecological Metadata Language (https://knbn.ecoinformatics.org/#external/emlparser/docs/eml-2.1.1/index.html)
EnvThes	Thesaurus for long term ecological research, monitoring, experiments (http://vocabs.lter-europe.net/EnvThesDev.html)
EU-BON	The main objective of EU BON is to build a substantial part of the Group on Earth Observation's Biodiversity Observation Network (GEO BON) (http://www.eubon.eu/show/project_2731/)
EUDAT	EUDAT (FP7 2011-2014) was a project under FP7 aiming at the provision of generic services to broad scale requirements of European scale research infrastructures. It addressed the specific challenges of data management, and to ensure a coherent approach to research data access and preservation and exploit new opportunities using its vision of a Collaborative Data Infrastructure (EUDAT CDI). (http://www.eudat.eu/)
EUDAT2020	EUDAT2020 (Horizon 2020 2014-2016) is a project under H2020 funding schema extending the already developed EUDAT service suite and fostering community uptake and sustainable governance (http://www.eudat.eu/)

EUNIS habitat list	The European nature information system, EUNIS, brings together European data from several databases and organisations into three interlinked modules on protection sites, species and habitat types. (http://eunis.eea.europa.eu/habitats-code-browser.jsp)
EUROVOC	EuroVoc is a multilingual, multidisciplinary thesaurus covering the activities of the EU, the European Parliament in particular. (http://eurovoc.europa.eu/)
GEMET	(www.eionet.europa.eu/gemet/en/themes/)
ILTER	International Long Term Ecological Research, is a 'network of networks', a global network of research sites located in a wide array of ecosystems that can help understand environmental change across the globe. ILTER's focus is on long-term, site-based research and monitoring. (http://www.ilternet.edu/)
INSPIRE	The INSPIRE directive aims to create a European Union (EU) spatial data infrastructure. This will enable the sharing of environmental spatial information among public sector organisations and better facilitate public access to spatial information across Europe. (http://inspire.ec.europa.eu/)
INSPIRE REGISTRY	The INSPIRE registry provides a central access point to a number of centrally managed INSPIRE registers. The content of these registers are based on the INSPIRE Directive, Implementing Rules and Technical Guidelines http://inspire.ec.europa.eu/registry/
ISO 19115	ISO 19115-1:2014 defines the schema required for describing geographic information and services by means of metadata. It provides information about the identification, the extent, the quality, the spatial and temporal aspects, the content, the spatial reference, the portrayal, distribution, and other properties of digital geographic data and services. (http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=53798)
ISocat	The ISO TC37 Data Category Registry (DCR) was created in 2008 as one of the first ISO standards delivered in the form of a database (ISocat). The Max Planck Institute for Psycholinguistics (MPI) has provided development, hosting, and support services and acted as the Registration Authority (RA). (http://www.isocat.org/)
lifeWatch	The e-Science and Technology European Research Infrastructure Consortium for Biodiversity and Ecosystem Research --LifeWatch ERIC-- will be established in accordance with Council Regulation (EC) No 723/2009 of 25 June 2009 on the Community legal framework for a European Research Infrastructure Consortium (ERIC). (http://www.lifewatch.eu/)
Linked 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 <u>data</u> , <u>information</u> , and <u>knowledge</u> on the Semantic Web using <u>URIs</u> and <u>RDF</u> ." (http://linkeddata.org/)
LTER	Long-Term Ecosystem Research is an essential component of world-wide efforts to better understand ecosystems.
LTER EUROPE	The European Long Term Ecological Research Network (LTER Europe) is the European regional group within the global ILTER network. It connects National networks and is the European contribution to the global International Long Term Ecological Research (ILTER) with over half of the ILTER members belonging to LTER-Europe (http://www.lter-europe.net/).
O&M	Observation & Measurement. This standard specifies an XML implementation for the OGC and ISO Observations and Measurements (O&M) conceptual model (OGC Observations and Measurements v2.0 also published as ISO/DIS 19156), including a schema for Sampling Features (http://schemas.opengis.net/om/2.0/ , http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=325

	74)
poolParty	PoolParty Thesaurus Server (PPT) is an advanced software platform to manage enterprise metadata and linked data based on semantic knowledge models (taxonomies, thesauri, ontologies and knowledge graphs). PPT's metadata management is based on W3C's Semantic Web standards RDF, SKOS and OWL and is combined with text mining and linked data mapping technologies (http://www.poolparty.biz/)
QUDT ontology	Ontology specifying quantities, units, dimensions, originally developed for the NASA Exploration Initiatives Ontology Models (NExIOM) project, a Constellation Program initiative at the AMES Research Center (ARC). (http://www.qudt.org/)
RDF	Resource Description Framework is a standard model for data interchange on the Web, extending the linking structure of the Web to use URIs to name the relationship between things as well as the two ends of the link (this is usually referred to as a "triple") (http://www.w3.org/RDF/)
SKOS/RDF	Simple Knowledge Organization System a common data model for sharing and linking knowledge organization systems via the Web. (http://www.w3.org/2004/02/skos/)
SOS	Standard, that defines a Web service interface which allows querying observations, sensor metadata, as well as representations of observed features (http://www.opengeospatial.org/standards/sos)
SPARQL	Query Language for RDF (http://www.w3.org/TR/rdf-sparql-query/)
US – LTER	Long Term Ecological Research (LTER) Network was created by the National Science Foundation (NSF) in 1980 to conduct research on ecological issues that can last decades and span huge geographical areas (http://www.lternet.edu/)
WFS	Web Feature Service specifies the behaviour of a service that provides transactions on and access to geographic features in a manner independent of the underlying data store. It specifies discovery operations, query operations, locking operations, transaction operations and operations to manage stored parameterized query expressions. (http://www.opengeospatial.org/standards/wfs)
XML	Extensible Markup Language (XML) is a simple, very flexible text format derived from SGML (ISO 8879). Originally designed to meet the challenges of large-scale electronic publishing, XML is also playing an increasingly important role in the exchange of a wide variety of data on the Web and elsewhere (http://www.w3.org/XML/)
EnvEurope	EnvEurope (Environmental quality and pressures assessment across Europe: the LTER network as an integrated and shared system for ecosystem monitoring; Life Environment Project LIFE08 ENV/IT/000399) is and European project which aimed to enhance the accessibility and discoverability of long term ecosystem data (http://www.enveurope.eu/).
ExpeER	ExpeER (Experimentation in Ecosystem Research) is an European project which aimed to bring together the major observational, experimental, analytical and modelling facilities in ecosystem science in Europe. (http://www.expeeronline.eu/)

Annex 1 EnvThes – Access and Management Interfaces

The thesaurus EnvThes (Environmental thesaurus) was developed inter-alia within the ExpeER project. The thematic and conceptual space covered by EnvThes is information gathered and related to long term ecological observation and experiments. The main application of EnvThes is as source for metadata keyword on the one hand and the annotation of data content (e.g. parameter naming) on the other. Envthes is publically available and the development is coordinated within a joint expert group on information management of LTER Europe, ILTER and AnaEE.

EnvThes is currently hosted at the Centre of Ecology and Hydrology (CEH, UK) and is managed by Herbert Schentz (EAA) and Nic Bertrand (CEH) as administrators.

EnvThes is available via two different interfaces addressing different target audiences. These two interfaces are described in the following section.

Viewing interface for interface

The *View interface* is targeted to the end users who view, discover and access the terms and structure of EnvThes using the web interface.

Link: <http://vocabs.lter-europe.net/>
 Development version: <http://vocabs.lter-europe.net/EnvThesDev.html>
 Production version: <http://vocabs.lter-europe.net/EnvThes.html>

Management interface for EnvThes

The *Management interface* is targeted to the editors and administrators of the thesaurus to maintain and further develop the thesaurus.

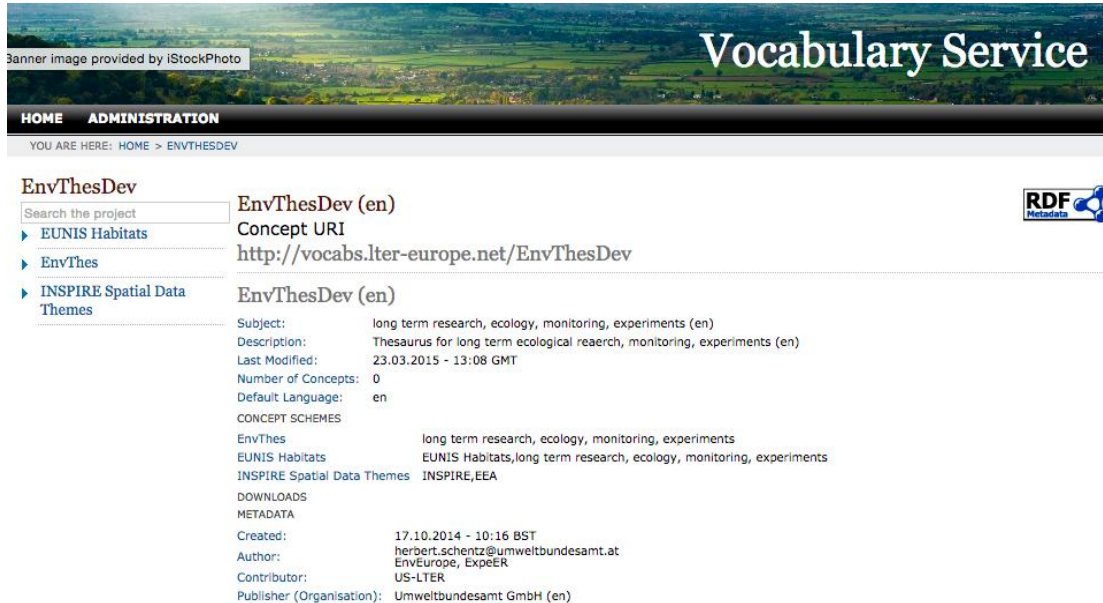
Link: <http://vocabs.lter-europe.net/PoolParty/>
 Development version: <http://vocabs.lter-europe.net/PoolParty/EnvThesDev.html>
 Production version: <http://vocabs.lter-europe.net/PoolParty/EnvThes.html>

The Management interface and the view interface include open *SPARQL endpoints* to query the thesaurus using SPARQL query language.

Link: <http://vocabs.lter-europe.net/PoolParty/sparql/>
 Development version: <http://vocabs.lter-europe.net/PoolParty/sparql/EnvThesDev/>
 Production version: <http://vocabs.lter-europe.net/PoolParty/sparql/EnvThes>

User Guide for EnvThes

The EnvThes thesaurus is exposed via a web-interface to the end users. It can be accessed via the link <http://vocabs.lter-europe.net/EnvThesDev>. This end user web-interface allows the user to **search**, **navigate and view details** on the structure and terms of the thesaurus.



Vocabulary Service

HOME ADMINISTRATION

YOU ARE HERE: HOME > ENVTHESDEV

EnvThesDev

Search the project

- ▶ EUNIS Habitats
- ▶ EnvThes
- ▶ INSPIRE Spatial Data Themes

EnvThesDev (en)

Concept URI
<http://vocabs.lter-europe.net/EnvThesDev>

EnvThesDev (en)

Subject: long term research, ecology, monitoring, experiments (en)
 Description: Thesaurus for long term ecological research, monitoring, experiments (en)
 Last Modified: 23.03.2015 - 13:08 GMT
 Number of Concepts: 0
 Default Language: en

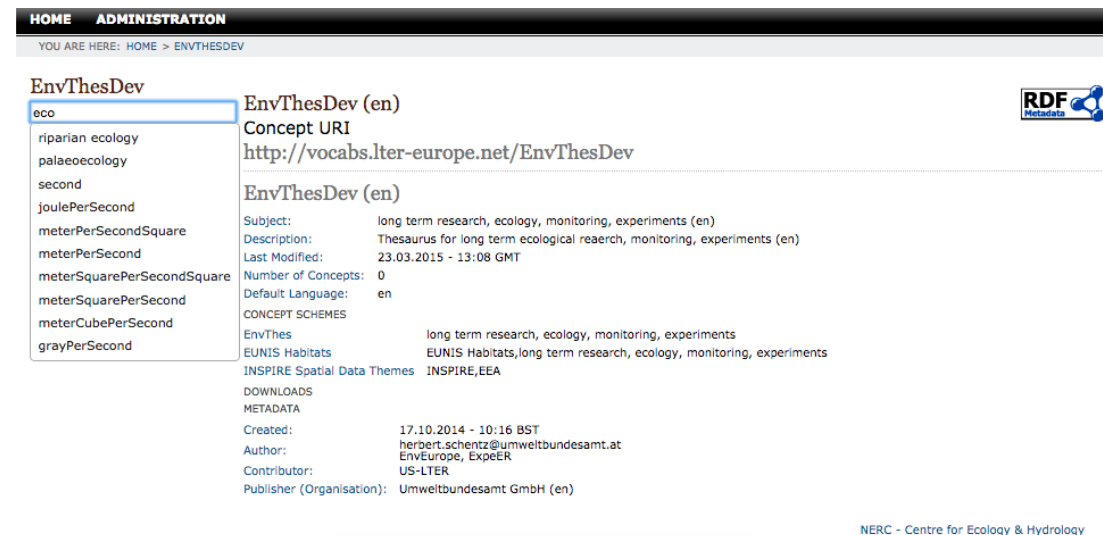
CONCEPT SCHEMES

EnvThes long term research, ecology, monitoring, experiments
 EUNIS Habitats EUNIS Habitats, long term research, ecology, monitoring, experiments
 INSPIRE Spatial Data Themes INSPIRE, EEA

DOWNLOADS
 METADATA

Created: 17.10.2014 - 10:16 BST
 Author: herbert.schentz@umweltbundesamt.at
 Contributor: EnvEurope, ExpeER
 Publisher (Organisation): Umweltbundesamt GmbH (en)

Discovery of terms - For this the autocomplete search can be used to search for any concept within the thesaurus.



EnvThesDev

eco

- riparian ecology
- palaeoecology
- second
- joulePerSecond
- meterPerSecondSquare
- meterPerSecond
- meterSquarePerSecondSquare
- meterSquarePerSecond
- meterCubePerSecond
- grayPerSecond

EnvThesDev (en)

Concept URI
<http://vocabs.lter-europe.net/EnvThesDev>

EnvThesDev (en)

Subject: long term research, ecology, monitoring, experiments (en)
 Description: Thesaurus for long term ecological research, monitoring, experiments (en)
 Last Modified: 23.03.2015 - 13:08 GMT
 Number of Concepts: 0
 Default Language: en

CONCEPT SCHEMES

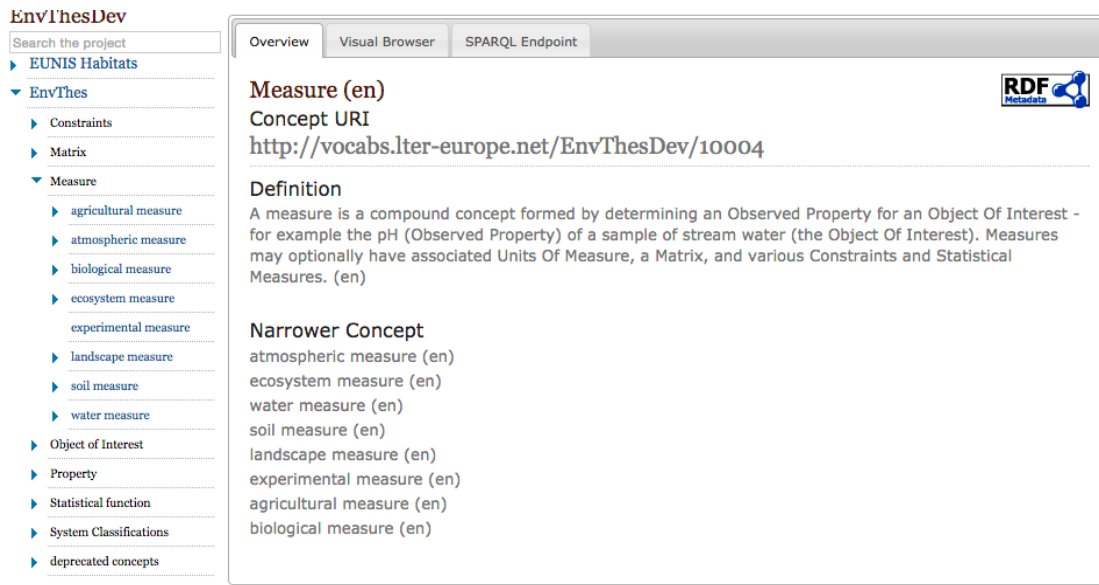
EnvThes long term research, ecology, monitoring, experiments
 EUNIS Habitats EUNIS Habitats, long term research, ecology, monitoring, experiments
 INSPIRE Spatial Data Themes INSPIRE, EEA

DOWNLOADS
 METADATA

Created: 17.10.2014 - 10:16 BST
 Author: herbert.schentz@umweltbundesamt.at
 Contributor: EnvEurope, ExpeER
 Publisher (Organisation): Umweltbundesamt GmbH (en)

NERC - Centre for Ecology & Hydrology

Navigate and access the hierarchy of the thesaurus – when using the hierarchical navigation at the left side of the window the user can navigate the hierarchical levels.



The screenshot shows the EnvThesDev interface. On the left is a sidebar with a search bar and a tree view of the thesaurus hierarchy. The main panel on the right displays details for the 'Measure (en)' concept.

EnvThesDev
Search the project
 ▶ EUNIS Habitats
 ▼ EnvThes
 ▶ Constraints
 ▶ Matrix
 ▼ Measure
 ▶ agricultural measure
 ▶ atmospheric measure
 ▶ biological measure
 ▶ ecosystem measure
 experimental measure
 ▶ landscape measure
 ▶ soil measure
 ▶ water measure
 ▶ Object of Interest
 ▶ Property
 ▶ Statistical function
 ▶ System Classifications
 ▶ deprecated concepts

Measure (en)
 Concept URI
<http://vocabs.lter-europe.net/EnvThesDev/10004>

Definition
 A measure is a compound concept formed by determining an Observed Property for an Object Of Interest - for example the pH (Observed Property) of a sample of stream water (the Object Of Interest). Measures may optionally have associated Units Of Measure, a Matrix, and various Constraints and Statistical Measures. (en)

Narrower Concept
 atmospheric measure (en)
 ecosystem measure (en)
 water measure (en)
 soil measure (en)
 landscape measure (en)
 experimental measure (en)
 agricultural measure (en)
 biological measure (en)

Viewing details of a term/concept – when selecting a term or concept from the thesaurus the details for the term are given, e.g. example for '*vegetation reflectance*'. Here also the links can be used to navigate between the related concepts as well as the broader or narrower concepts.



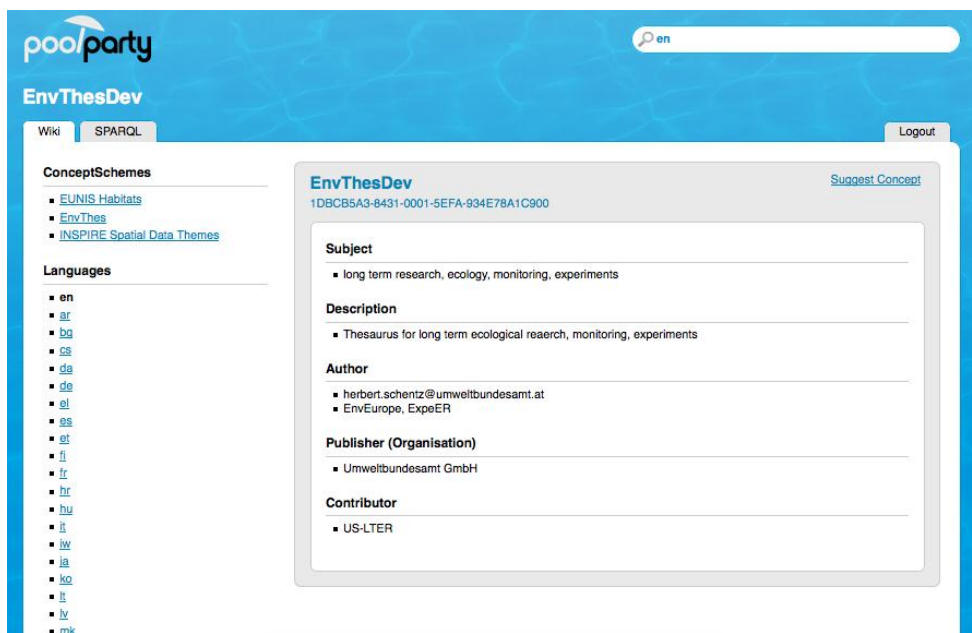
The screenshot shows the EnvThesDev interface with details for the 'vegetation reflectance (en)' concept.

vegetation reflectance (en)
 Concept URI
<http://vocabs.lter-europe.net/EnvThesDev/10118>

Broader Concept
 ecosystem measure (en)

Related Concepts
 reflectance (en)

Alternative viewer – the EnvThes thesaurus is also accessible via PoolParty browser interface at <http://onto.nerc.ac.uk/PoolParty/wiki/EnvThesDev>



The screenshot shows the PoolParty interface for the EnvThesDev thesaurus. The top navigation bar includes the PoolParty logo, a search bar with 'en' entered, and a 'Logout' button. Below the navigation bar, the 'EnvThesDev' title is displayed, along with 'Wiki' and 'SPARQL' tabs. On the left sidebar, there are sections for 'ConceptSchemes' (listing EUNIS Habitats, EnvThes, and INSPIRE Spatial Data Themes) and 'Languages' (listing various language codes including en, ar, bg, ca, da, de, el, es, et, fi, fr, hr, hu, it, ja, ko, lt, lv, and mk). The main content area displays the details for 'EnvThesDev' with the identifier '1DBC85A3-8431-0001-5EFA-934E78A1C900'. The details include:

- Subject:** long term research, ecology, monitoring, experiments
- Description:** Thesaurus for long term ecological research, monitoring, experiments
- Author:** herbert.schentz@umweltbundesamt.at, EnvEurope, ExpeER
- Publisher (Organisation):** Umweltbundesamt GmbH
- Contributor:** US-LTER

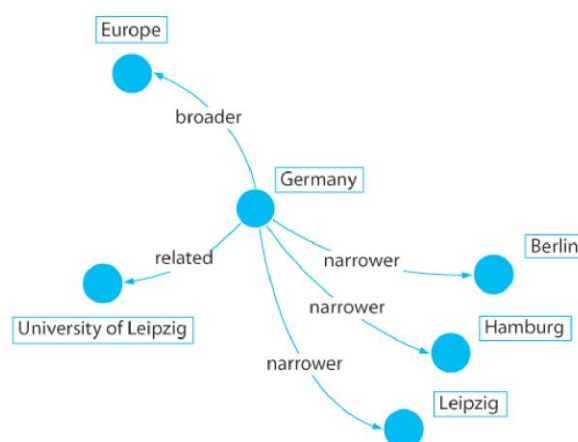
 A 'Suggest Concept' link is also visible in the top right of the main content area.

Editors Guide for EnvThes

The following text was compiled from the PoolParty user guide (Semantic Web Company¹²) in order to provide a basic guideline document for the editors (see Chapter 4.1) to work with the PoolParty interface. Parts of the text were directly copied from the original source.

A. Introduction

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.



Concepts are arranged in hierarchical or associative relationships. Hierarchical relationships are used to indicate concepts which are narrower and broader in scope, e.g. in a geographical thesaurus one expects to have concepts for countries and cities and a structure that declares which country is a super-class of which cities. An associative relationship is of a more general nature, it can be used to indicate any kind of relation between concepts. Concepts and their relations are explained in greater detail by means of examples and visualisations in the following sections.

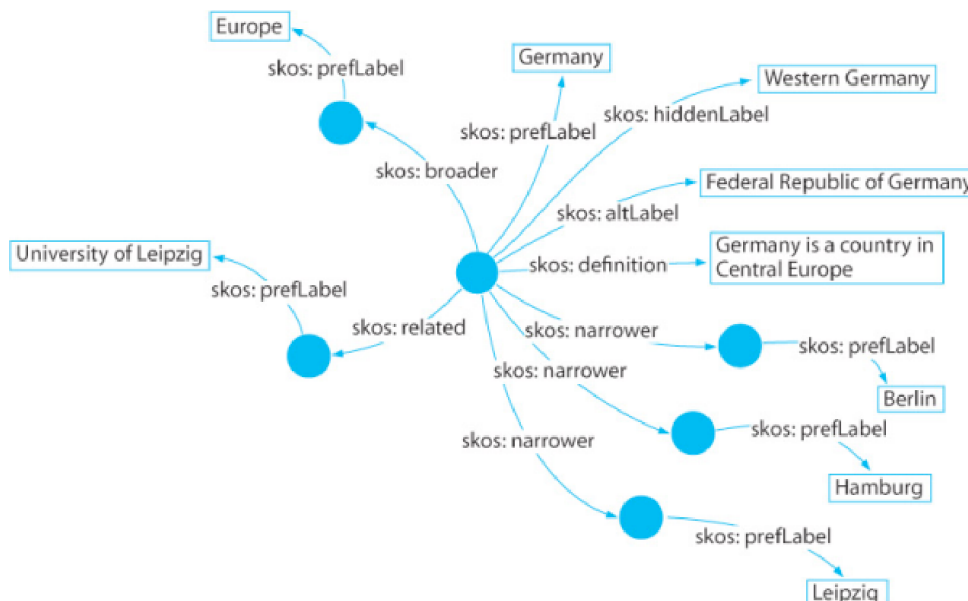
This list of the most important **SKOS** classes and properties used by PoolParty is relevant for advanced users or programmers who need to work with the RDF representation of a project. We also give a very short introduction to the Semantic Web technologies behind SKOS. For more comprehensive information on the Semantic Web, please go to W3C's **Semantic Web Activity site** and pick a presentation, which fits your interests.

SKOS is based on the Resource Description Framework (**RDF**), which is one of the fundamental Semantic Web specifications. It was developed within the W3C standards framework. RDF uses graphs as its data structure. This means that each concept is a node and edges connect these nodes to create a graph.

The image below displays data of a thesaurus presented as a graph. The graph has the concept 'Germany' as its central node. Several edges connect this node to its labels and definition as well as to other nodes. All edges are typed, thereby indicating which kind of relationships exist between the nodes they connect, e.g. the node 'Germany' is connected to the node 'Hamburg' via *skos:narrower*.

¹² See <https://grips.semantic-web.at/display/public/POOLDOKU/>

The edges are called 'properties' in RDF. They can also connect a node with a literal, in the case of SKOS these are the labels, definitions or scope notes.



Each statement in RDF encodes information as so called 'triples'. A triple always has the form **subject** > **predicate** > **object**, e.g. 'Germany' > *skos:narrower* > 'Hamburg'. Each property and class in SKOS and each concept in PoolParty has a uniform (unique) identifier, also called URI. In the case of SKOS these are always HTTP URIs, in the case of PoolParty they can either be URNs or HTTP URIs. The complete URIs of SKOS properties or classes can be formed by expanding the namespace prefix 'skos:' which is just a shortcut for the base URI <http://www.w3.org/2004/02/skos/core#>. The complete URI for e.g. *skos:narrower* is <http://www.w3.org/2004/02/skos/core#narrower>.

B. System Requirements:

Client Side

- PoolParty works with the following browsers:
- Chrome (25+)
- Firefox (3.6+)
- Safari (5+)
- PoolParty has been designed for a screen resolution of 1024x768 or higher.

It is recommended to have a DSL connection to the Internet; otherwise response times might be a bit slow.

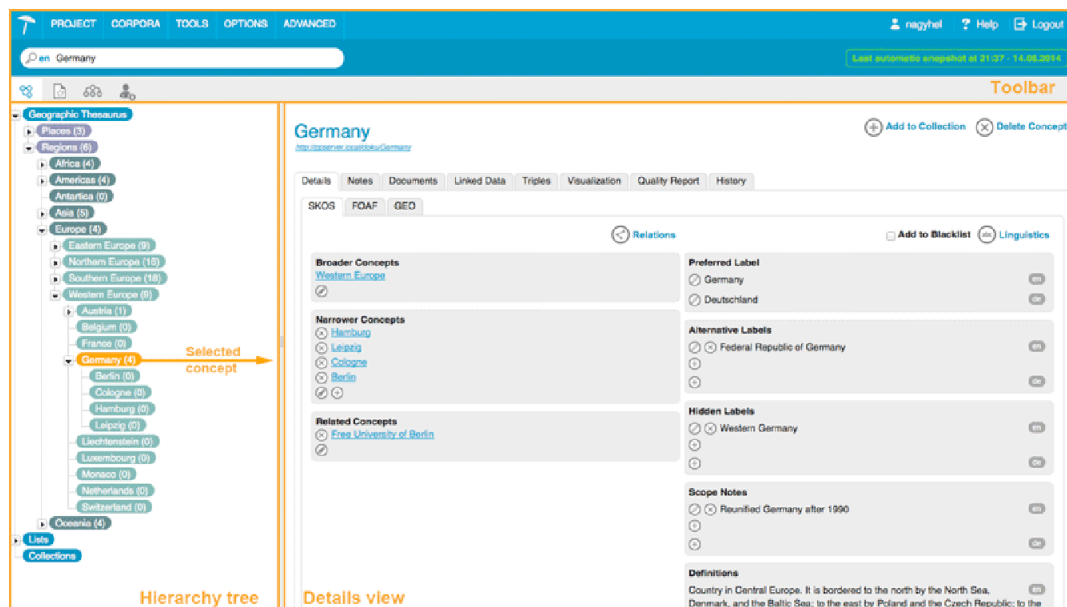
C. Edit and Administration interface of PoolParty

Once logged in, select project and load "EnvThesDev"

The graphical user interface is divided into three areas:

- **Toolbar** - Provides the main menu, a search function and the document processing icons.
- **Hierarchy tree** - Displays the hierarchical structure of your project. Click on the boxes to select a concept scheme or a concept.

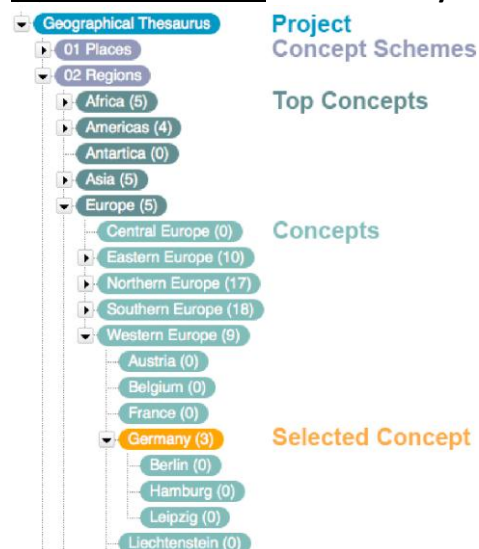
- **Details view** - Displays detailed information available for the selected project, concept scheme or concept. The available information is grouped in tabs. You can use the tab bar to switch between the different views.



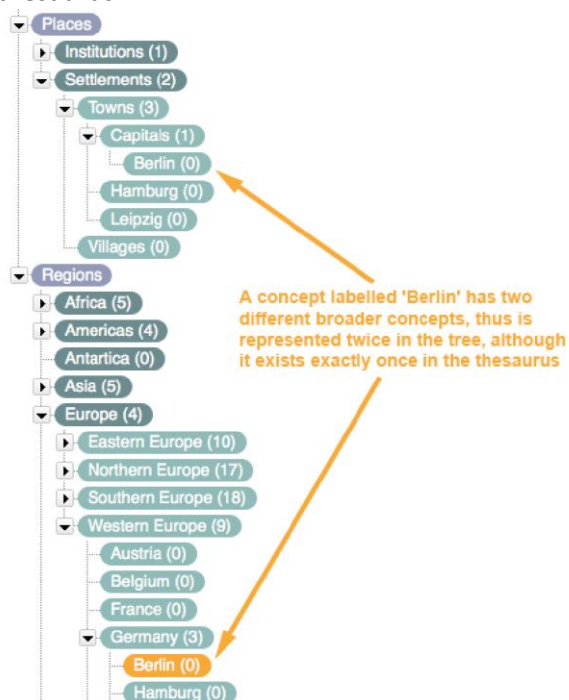
(a) TOOLBAR: The PoolParty **Toolbar** provides the following options:



(b) HIERARCHY VIEW: The **Hierarchy tree** in a project shows the hierarchy of concepts in a thesaurus:



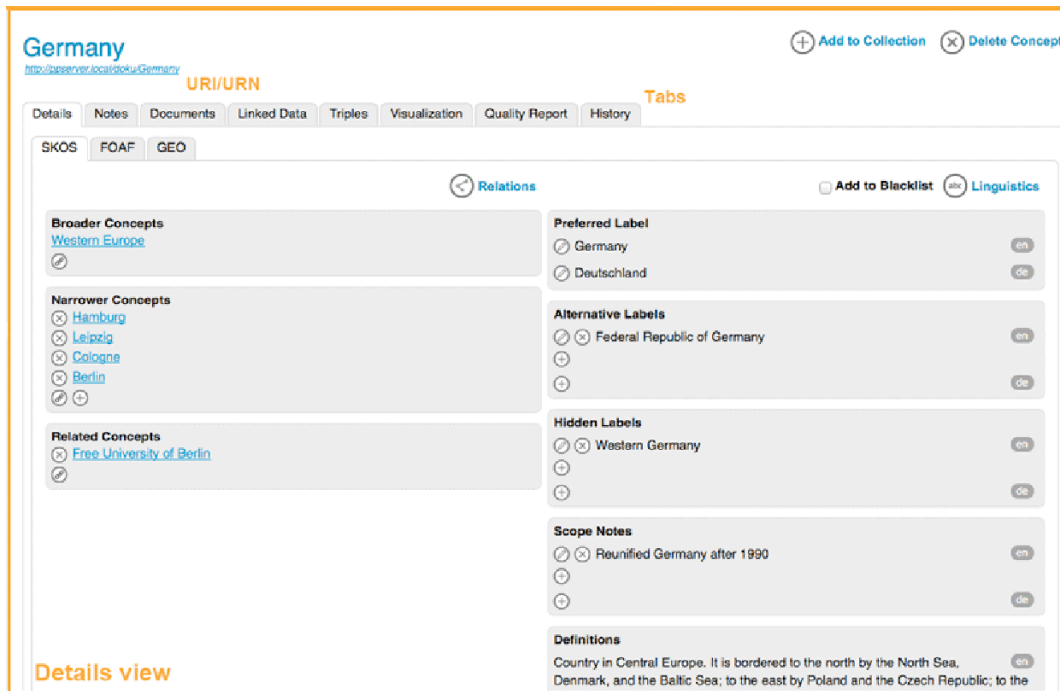
Concepts/terms in a thesaurus can be related in a poly-hierarchy, meaning that a given concept can be linked to more than one broader concept. This is also shown in the hierarchical view of the thesaurus.



(c) DETAIL VIEW

In the details view you can find, add and edit detailed information for the selected node (project, concept scheme, concept) in the **hierarchy tree**. All data you enter in PoolParty is **validated** and you get immediate feedback when you try to enter data that is not allowed for the respective field type.

Located at the top of the details view underneath the name of the selected node you can find its identifier. The uniform identifier is the URI or URN of the selected project, concept scheme or concept. The **type and form of the identifier** is defined when creating the project. For projects with URIs it is displayed as a link to the respective concept schemes and concepts in the **Linked Data Frontend**. All data you enter in PoolParty in the details view or via other options is validated to prevent inconsistencies.



D. Creating concepts

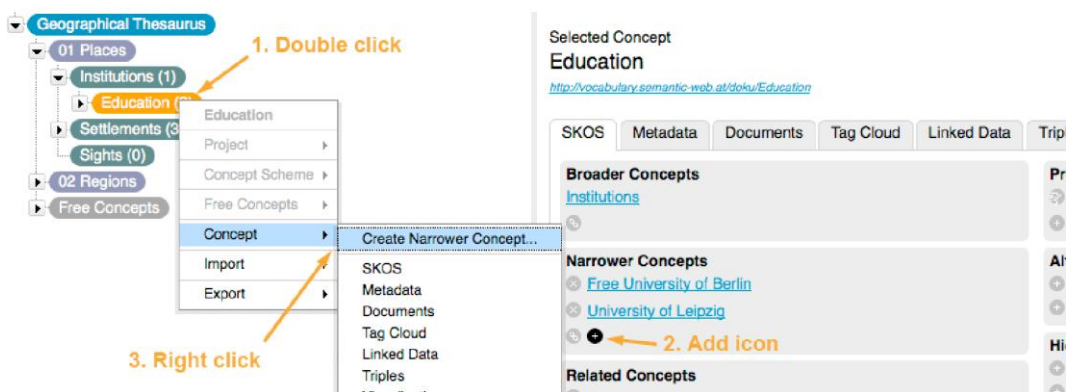
Your work as editor is to create new concepts and edit details of an existing concept. For example to edit an existing concept, select the concept you wish to edit and edit in the Details View.

All information for a selected concept scheme or concept can be edited in the **details view**. You have different ways of adding, editing or deleting details depending on the type of property.

1. To edit labels or documentation properties click on the **pen** icon to activate inline editing.
2. Clicking the **+** icon lets you add new labels or documentation properties or relations.
3. Click the **link** icon to add relations to your concept scheme or concept. Via **autocomplete** you can search for existing concepts.
4. Click the **x** icon to delete labels, documentation properties or relations to your concept.

How to Create a new Concept:

1. There are three simple ways to create new concepts:
2. By double clicking a concept's node in the hierarchy tree.
3. By clicking the **+** icon in the 'Narrower' area of a selected concept.
4. Through a context menu which can be accessed by right clicking any node in the tree.



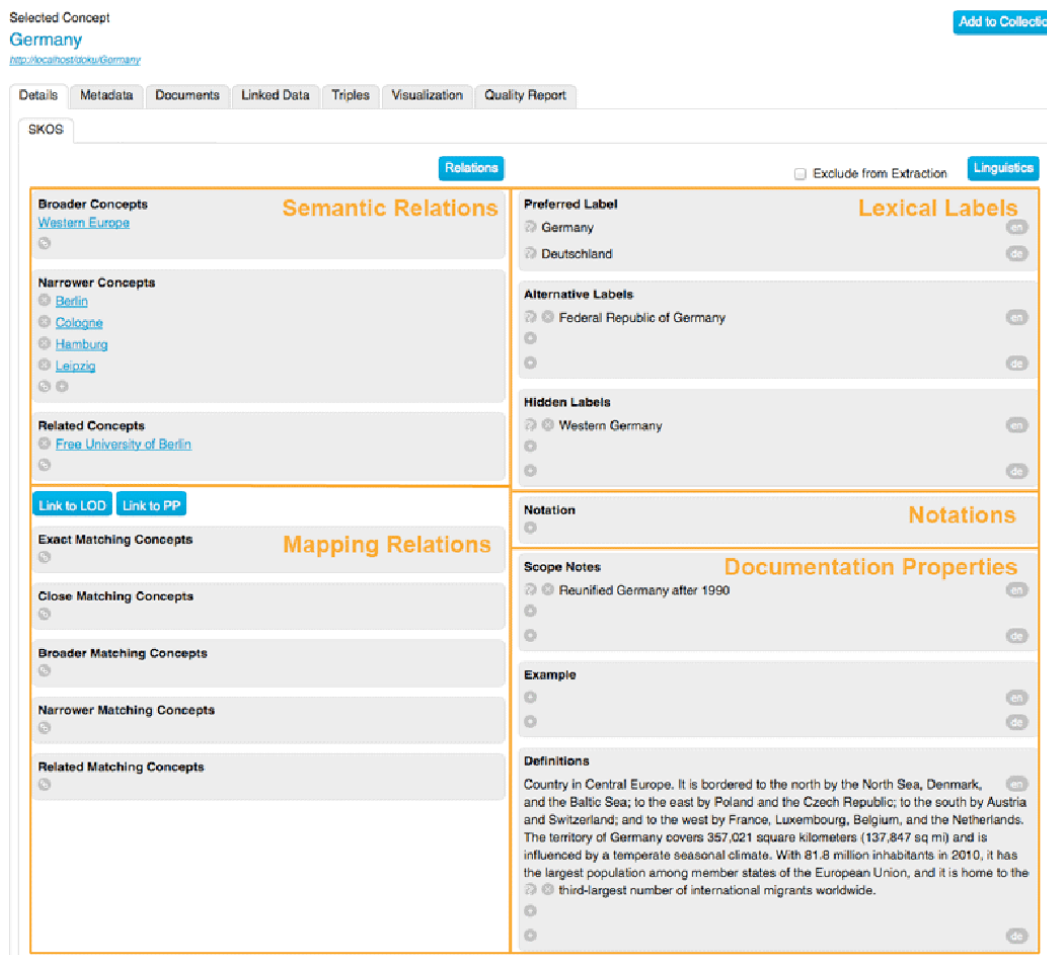
After selecting one way to create a concept, the **New Concept** dialog appears. You have to enter the name (preferred label) for the concept, clicking **Create Concept** completes the procedure.



The dialog box titled "New Concept" contains a text input field labeled "Preferred Label:". Below the input field are two buttons: "Create Concept" and "Cancel".

E. Viewing and Editing Concepts

The image below shows the default tab of the details view for a concept which is the **SKOS** tab. For the **SKOS** tab a **Basic** view and an **Advanced** view is available. In the **Basic** view only the basic SKOS properties for a concept are available. You can switch between the **Basic** and **Advanced** view via the **Option > SKOS View** menu. In this example we see the **Advanced** view for the concept 'Germany'. The main view for a concept is divided into four areas containing different information:



The screenshot shows the "Selected Concept" details for "Germany" (http://localhost:8080/Germany). The interface includes tabs for "Details", "Metadata", "Documents", "Linked Data", "Triples", "Visualization", and "Quality Report". The "SKOS" tab is active, displaying a "Relations" view with a "Linguistics" button and an "Exclude from Extraction" checkbox.

The main content area is divided into four sections:

- Semantic Relations:** Contains "Broader Concepts" (Western Europe), "Narrower Concepts" (Berlin, Cologne, Hamburg, Leipzig), and "Related Concepts" (Free University of Berlin).
- Mapping Relations:** Contains "Exact Matching Concepts", "Close Matching Concepts", "Broader Matching Concepts", "Narrower Matching Concepts", and "Related Matching Concepts".
- Lexical Labels:** Contains "Preferred Label" (Germany, Deutschland), "Alternative Labels" (Federal Republic of Germany), and "Hidden Labels" (Western Germany).
- Notations and Documentation Properties:** Contains "Notation", "Scope Notes" (Reunified Germany after 1990), "Example", and "Definitions" (Country in Central Europe. It is bordered to the north by the North Sea, Denmark, and the Baltic Sea; to the east by Poland and the Czech Republic; to the south by Austria and Switzerland; and to the west by France, Luxembourg, Belgium, and the Netherlands. The territory of Germany covers 357,021 square kilometers (137,847 sq mi) and is influenced by a temperate seasonal climate. With 81.8 million inhabitants in 2010, it has the largest population among member states of the European Union, and it is home to the third-largest number of international migrants worldwide).

The list below shows all SKOS properties available in the **SKOS** tab. The properties available in the **Basic** view are marked with *.

Semantic relations

*Broader Concepts / Narrower Concepts** - The 'broader' area of a concept's details displays its parent(s), while the 'narrower' area lists all the child concepts.

*Related** - All concepts related (in some way) to the current one.

Exact Matching Concepts - Concepts of different concept schemes (thesauri) or linked data which are identical. Default setting for resources mapped to the concept via the **linking open data feature** of PoolParty.

Close Matching Concepts - Concepts of different concept schemes (thesauri) or linked data which are nearly identical.

Broader Matching Concepts / Narrower Matching Concepts - Concepts of different concept schemes (thesauri) or linked data which have a broader / narrower meaning.

Related Matching Concepts - Concepts of different concept schemes (thesauri) or linked data which are related in some way.

Lexical Labels

There may be only one preferred label in each language of a concept, while multiple alternative and hidden labels are allowed.

*Preferred Label** - The main word or phrase that is used to identify the concept. Each concept must have a preferred label, but not more than one for each language.

*Alternative Label** - Other words or phrases used to refer to this concept. This includes synonyms, acronyms, abbreviations, spelling variants and irregular plural/singular forms.

*Hidden Label** - Labels that should be hidden for display purposes of the concept, but which still will be used for search operations. These are misspellings or otherwise deprecated terms for a concept.

Notation

Notation - A classification code uniquely identifying the concept in the concept scheme.

Document properties

*Scope Note** - A description to clarify the usage and scope of a concept, i.e. information about what is or is not included within its meaning.

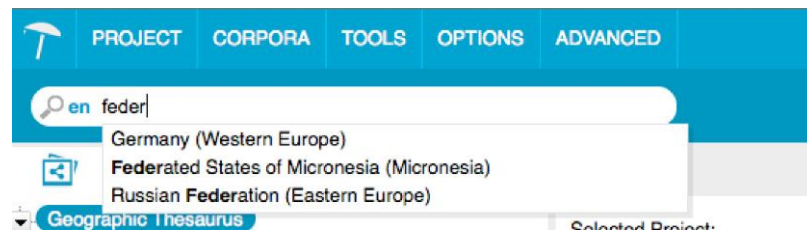
Example - An example to illustrate the usage and scope of a concept.

*Definition** - A description of the meaning of the concept.

F. Searching for Concepts

To find a concept quickly within an opened project, begin to enter a search string in the **search bar**. The system will look for concepts that contain that string in any label (also in alternative labels etc.) and will suggest matching ones. The example displayed in the image below shows that three concepts were found when 'feder' was entered as a search string. 'Germany' is being found also

because it has an alternative label 'Federal Republic of Germany'. Selecting a concept from the drop-down list presented by autocomplete will expand the containing branch of the **hierarchy tree** and also will display the **concept's details view**.



Alternatively it is possible to hit the Enter key after having entered a search string. In that case a search is done in the project and a list of all concepts in which the search string matches a preferred, alternative or hidden label is displayed in the **Concepts > Concept Search** tab of the project node. Via the filter section you can filter the list by label (all (default), only preferred, only alternative, only hidden). Here you can also determine whether the search string should be at the beginning of the concept label or should only contain it, and also if it should not be contained by any label at all.

Selected Project: **Geographic Thesaurus**
bb2d1837-7654-4c89-8994-d2315468b56b

Project Linking LD Frontend PP Browser

Metadata & Statistics Concepts Triples Log SPARQL Autopopulate Quality Report

Concept Search Concept Index

Filter By

Any label contains en feder Search

Results

3 matching concepts!

Concept	Alternative Labels	Broader Concepts	# Narrower Concepts	Delete
Federated States of Micronesia		Micronesia	0	
Germany	Federal Republic of Germany	Western Europe	4	
Russian Federation		Eastern Europe	0	

1

In addition to the guidelines defined for the work with EnvThes a number of actions should be avoided even if possible.

DO NOT's

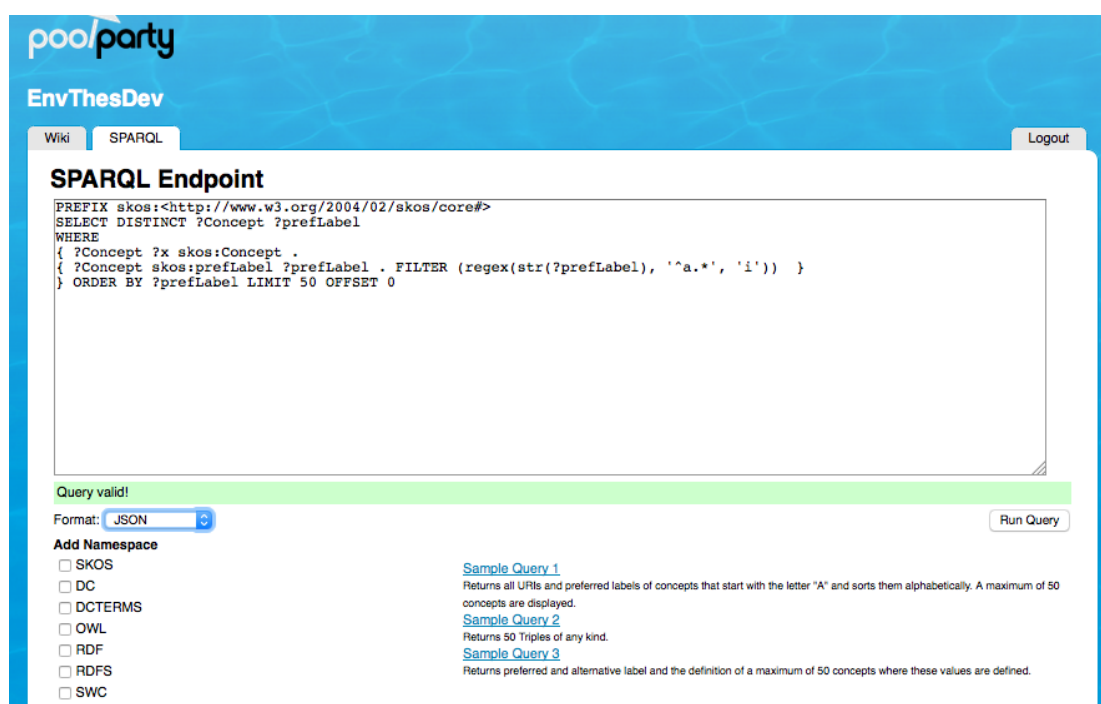
- Delete an existing Concept Scheme
- Create a new concept Scheme
- Create a new Top Concept
- Drag and drop Higher level concepts in other parts of the hierarchy

SPARQL Endpoint

The Management interface PoolParty also provides a publically available and accessible SPARQL endpoint allows querying the thesaurus using the SPARQL query language. In the following a number of Example queries shall demonstrate some of the numerous possibilities to select content of the thesaurus. First the query is shown and afterwards an extraction of the result.

SPARQL is the query language for the semantic web and any application that offers a SPARQL endpoint, allows establishing simple or sophisticated queries. The SPARQL endpoints of EnvThes development version (<http://vocabs.lter-europe.net/PoolParty/sparql/EnvThesDev>) and of EnvThes productive version (<http://vocabs.lter-europe.net/PoolParty/sparql/EnvThes>) are freely accessible.

Links Main link: <http://vocabs.lter-europe.net/PoolParty/sparql/EnvThesDev>
 Alternative link: <http://onto.nerc.ac.uk/PoolParty/sparql/EnvThesDev>



The screenshot shows the PoolParty EnvThesDev SPARQL Endpoint interface. It features a blue header with the 'poolparty' logo and 'EnvThesDev' text. Below the header, there are tabs for 'Wiki' and 'SPARQL', and a 'Logout' button. The main area is titled 'SPARQL Endpoint' and contains a text box with a sample SPARQL query:

```
PREFIX skos:<http://www.w3.org/2004/02/skos/core#>
SELECT DISTINCT ?Concept ?prefLabel
WHERE
{
  ?Concept ?x skos:Concept .
  { ?Concept skos:prefLabel ?prefLabel . FILTER (regex(str(?prefLabel), '^a.*', 'i')) }
} ORDER BY ?prefLabel LIMIT 50 OFFSET 0
```

Below the query box, there is a green bar indicating 'Query valid!'. A 'Format:' dropdown menu is set to 'JSON', and a 'Run Query' button is visible. On the left, there is an 'Add Namespace' section with checkboxes for SKOS, DC, DCTERMS, OWL, RDF, RDFS, and SWC. On the right, there are three sample queries with descriptions:

- [Sample Query 1](#): Returns all URIs and preferred labels of concepts that start with the letter "A" and sorts them alphabetically. A maximum of 50 concepts are displayed.
- [Sample Query 2](#): Returns 50 Triples of any kind.
- [Sample Query 3](#): Returns preferred and alternative label and the definition of a maximum of 50 concepts where these values are defined.

EXAMPLE 1: Get for all concept/terms the English preferred labels and definitions

This example query gets all English preferred labels as strings (without language tag) and English definitions as strings, when they exist. If no definition exists, only the label is listed, using the OPTIONAL clause.

Example query:

```
select distinct ?concept (str(?prefLab) as ?label) (str(?def) as ?definition)
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.
FILTER(lang(?prefLab) = 'en').
OPTIONAL {?concept <http://www.w3.org/2004/02/skos/core#definition> ?def.
FILTER (lang(?def)='en')}.
}
```

```
order by ?label
LIMIT 2000
```

Example output

?concept	?label	?definition
http://vocabs.lter-europe.net/EnvThesDev/43	"C horizon"	"[Agren & Andersson 2012] Mineral horizon largely unaffected by soil development."
http://vocabs.lter-europe.net/EnvThesDev/EnvEU_249	"C/N ratio"	
http://vocabs.lter-europe.net/EnvThesDev/EUUnits_52	"C/kg"	
http://vocabs.lter-europe.net/EnvThesDev/EUUnits_53	"C/m²"	
http://vocabs.lter-europe.net/EnvThesDev/EUUnits_54	"C/m³"	
http://vocabs.lter-europe.net/EnvThesDev/EnvEU_309	"CH4 fluxes"	"the flows of CH4 particles in a given space and per time unit"
http://vocabs.lter-europe.net/EnvThesDev/EnvEU_260	"CO2 concentration"	" A concentration is the amount of a substance in a given volume, thus CO2 concentration is a measure of the percentage of the atmosphere that is composed of CO2 (e.g., 380 ppm). "
http://vocabs.lter-europe.net/EnvThesDev/EnvEu_126	"CO2 flux per soil horizon"	
http://vocabs.lter-europe.net/EnvThesDev/EnvEU_310	"CO2 fluxes"	"the flows of CO2 particles in a given space and per time unit"
http://vocabs.lter-europe.net/EnvThesDev/EnvEu_127	"CO2 surface flux"	" For the exchange of chemicals (water,CO2, etc.) a surface flux is defined as the amount per unit volume of a quantity that passes through a horizontal unit area on the surface"
http://vocabs.lter-europe.net/EnvThesDev/85	"COLTEST"	"Test in order to show what could be done within the catalogue of life"
http://vocabs.lter-europe.net/EnvThesDev/USLterCV_733	"CTD"	"the primary tool for determining essential physical properties of sea water. It gives scientists a precise and comprehensive charting of the distribution and variation of water temperature, salinity, and density that helps to understand how the oceans affect life."
http://vocabs.lter-europe.net/EnvThesDev/INSPIRE_6	"Cadastral parcels"	"Areas defined by cadastral registers or equivalent."

EXAMPLE 2: Get English preferred labels and definitions for all concept/terms from the INSPIRE spatial data themes and EUNIS Habitats

This example shows the usage of the fact, that we used special identifiers for content that is imported from other vocabularies. In the case of EUNIS habitats, imported from the EEA, all identifiers start with "EUNIS" and in the case of the INSPIRE spatial data themes, imported from the EIONET, all identifiers start with "INSPIRE"

Example query:

```
select distinct ?concept (str(?prefLab) as ?label) (str(?def) as ?definition)
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.
FILTER(lang(?prefLab) = 'en').
OPTIONAL {?concept <http://www.w3.org/2004/02/skos/core#definition> ?def.
FILTER (lang(?def)='en')}.
FILTER (CONTAINS (STR(?concept), "INSPIRE") || (CONTAINS (STR(?concept), "EUNIS")) ) }
order by ?label
LIMIT 2000
```

Example output

?concept	?label	?definition
http://vocabs.lter-europe.net/EnvThesDev/EUNIS_D3	"Aapa, palsa and polygon mires"	
http://vocabs.lter-europe.net/EnvThesDev/EUNIS_E4_3	"Acid alpine and subalpine grassland"	
http://vocabs.lter-europe.net/EnvThesDev/EUNIS_H3_1	"Acid siliceous inland cliffs"	
http://vocabs.lter-europe.net/EnvThesDev/EUNIS_H2_5	"Acid siliceous screes of warm exposures"	
http://vocabs.lter-europe.net/EnvThesDev/EUNIS_G1_8	"Acidophilous [Quercus]-dominated woodland"	
http://vocabs.lter-europe.net/EnvThesDev/EUNIS_J3_2	"Active opencast mineral extraction sites, including quarries"	
http://vocabs.lter-europe.net/EnvThesDev/EUNIS_J3_1	"Active underground mines"	
http://vocabs.lter-europe.net/EnvThesDev/EUNIS_H6_1	"Active volcanic features"	
http://vocabs.lter-europe.net/EnvThesDev/EUNIS_J6_4	"Agricultural and horticultural waste"	
http://vocabs.lter-europe.net/EnvThesDev/EUNIS_J2_4	"Agricultural constructions"	
http://vocabs.lter-europe.net/EnvThesDev/EUNIS_E2_6	"Agriculturally-improved, re-seeded and heavily fertilised grassland, including sports fields and grass lawns"	
http://vocabs.lter-europe.net/EnvThesDev/INSPIRE_5	"Addresses"	"Location of properties based on address identifiers, usually by road name, house number, postal code."
http://vocabs.lter-europe.net/EnvThesDev/INSPIRE_4	"Administrative units"	"Units of administration, dividing areas where Member States have and/or exercise jurisdictional rights, for local, regional and national governance, separated by administrative boundaries."
http://vocabs.lter-europe.net/EnvThesDev/INSPIRE_22	"Agricultural and aquaculture facilities"	"Farming equipment and production facilities (including irrigation systems, greenhouses and stables)."

EXAMPLE 3: Hierarchy Levels with and without definition

The example query shows a way to provide flat views of hierarchies. Sometimes it may be desirable to get a "flattened" view of a hierarchy. The following query produces such a flattened list for up to 5 hierarchy levels by the establishment of 5 subqueries, which are joined by the fact that there is a "broader" relation from the lower level to the upper level. (e.g. *?gen2* <<http://www.w3.org/2004/02/skos/core#broader>> *?gen1*). In order to also get the results for less than 5 hierarchy levels, the OPTIONAL clause is used. The same clause is used in order to get results, independently of the fact whether there is a definition or not.

Example query:

```

PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
PREFIX dc:<http://purl.org/dc/elements/1.1/>
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl:<http://www.w3.org/2002/07/owl#>
PREFIX skos:<http://www.w3.org/2004/02/skos/core#>
Select (str(?gen1label) as ?gen1lab) (str(?defin1) as ?def1)
      (str(?gen2label) as ?gen2lab) (str(?defin2) as ?def2) (str(?gen3label) as
?gen3lab) (str(?defin3) as ?def3) (str(?gen4label) as ?gen4lab) (str(?defin4) as
?def4) (str(?gen5label) as ?gen5lab) (str(?defin5) as ?def5)
WHERE
{?gen1      <http://www.w3.org/2004/02/skos/core#broader>      <http://vocabs.lter-
europe.net/EnvThesDev/10005>.
?gen1      <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
<http://www.w3.org/2004/02/skos/core#Concept>.
?gen1 <http://www.w3.org/2004/02/skos/core#prefLabel> ?gen1label.
FILTER (lang(?gen1label) = 'en').

```

```

OPTIONAL{?gen1 <http://www.w3.org/2004/02/skos/core#definition> ?defin1.
FILTER (lang(?defin1) = 'en')}
  OPTIONAL {
    ?gen2 <http://www.w3.org/2004/02/skos/core#broader> ?gen1.
    ?gen2 <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
    <http://www.w3.org/2004/02/skos/core#Concept>.
    ?gen2 <http://www.w3.org/2004/02/skos/core#prefLabel> ?gen2label.
    FILTER (lang(?gen2label) = 'en').
    OPTIONAL{?gen2 <http://www.w3.org/2004/02/skos/core#definition> ?defin2.
    FILTER (lang(?defin2) = 'en')}.
    OPTIONAL{
      ?gen3 <http://www.w3.org/2004/02/skos/core#broader> ?gen2.
      ?gen3 <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
      <http://www.w3.org/2004/02/skos/core#Concept>.
      ?gen3 <http://www.w3.org/2004/02/skos/core#prefLabel> ?gen3label.
      FILTER (lang(?gen3label) = 'en').
      OPTIONAL{?gen3 <http://www.w3.org/2004/02/skos/core#definition>
      ?defin3.
      FILTER (lang(?defin3) = 'en')}.
      OPTIONAL{
        ?gen4 <http://www.w3.org/2004/02/skos/core#broader> ?gen3.
        ?gen4 <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
        <http://www.w3.org/2004/02/skos/core#Concept>.
        ?gen4 <http://www.w3.org/2004/02/skos/core#prefLabel>
        ?gen4label.
        FILTER(lang(?gen4label) = 'en').
        OPTIONAL{?gen4 <http://www.w3.org/2004/02/skos/core#definition>
        ?defin4.
        FILTER (lang(?defin4)= 'en')}.
        OPTIONAL{
          ?gen5 <http://www.w3.org/2004/02/skos/core#broader>
          ?gen4.
          ?gen5 <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
          <http://www.w3.org/2004/02/skos/core#Concept>.
          ?gen5 <http://www.w3.org/2004/02/skos/core#prefLabel>
          ?gen5label.
          FILTER(lang(?gen5label) = 'en').
          OPTIONAL{?gen5
          <http://www.w3.org/2004/02/skos/core#definition> ?defin5.
          FILTER (lang(?defin5)= 'en')}.
          }
        }
      }
    }
  }
}

```

Example output

ecosystems"	"[Henderson's] n. community of different spe-cies interdependent on each other, together with their non-living environment, which is relatively self-contained in terms of energy flow, and is distinct from neigh-bouring communities. Different types of ecosystem	"terrestrial ecosystems"	"[GEMET] Any terrestrial environment, from small to large, in which plants and animals interact with the chemical and physical features of the environment."	"forest ecosystems"	"[GEMET] Any forest environment, in which plants and animals interact with the chemical and physical features of the environment, in which they live."	"forests"	"[GEMET] A vegetation community dominated by trees and other woody shrubs, growing close enough together that the tree tops touch or overlap, creating various degrees of	"clearcuts"	"Area where clearcutting has been applied; the area is subject of fast succession processes."
-------------	--	--------------------------	--	---------------------	--	-----------	---	-------------	---

	are defined by the collection of organisms found within them, e.g. forest, grassland. Continuous ecosystems covering very large areas, such as the northern coniferous forest or the steppe grassland, are known as biomes. "						shade on the forest floor. It may produce benefits such as timber, recreation, wildlife habitat, etc."		
--	---	--	--	--	--	--	--	--	--

EXAMPLE 4: Get all predicates of a concept

The query example, generally speaking demands: "tell me all that can be known about a certain concept", in our case "abundance"

Example query

```
select ?s ?p ?o
where {?s ?p ?o.
?s <http://www.w3.org/2004/02/skos/core#prefLabel> ?prefLab.
FILTER (STR(?prefLab)='abundance') .
FILTER (lang(?prefLab)='en') }
```

Example output

s	p	o
http://vocabs.lter-europe.net/EnvThesDev/USLterCV_6	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.w3.org/2004/02/skos/core#Concept
http://vocabs.lter-europe.net/EnvThesDev/USLterCV_6	http://www.w3.org/2004/02/skos/core#prefLabel	"abundance"@en
http://vocabs.lter-europe.net/EnvThesDev/USLterCV_6	http://www.w3.org/2004/02/skos/core#narrower	http://vocabs.lter-europe.net/EnvThesDev/USLterCV_462
http://vocabs.lter-europe.net/EnvThesDev/USLterCV_6	http://www.w3.org/2004/02/skos/core#scopeNote	"US LTER controlled vocabulary"@en
http://vocabs.lter-europe.net/EnvThesDev/USLterCV_6	http://www.w3.org/2004/02/skos/core#exactMatch	http://linkeddata.ge.imati.cnr.it:2020/resource/EARTh/177160
http://vocabs.lter-europe.net/EnvThesDev/USLterCV_6	http://www.w3.org/2004/02/skos/core#exactMatch	http://linkeddata.ge.imati.cnr.it:2020/resource/EARTh/177160
http://vocabs.lter-europe.net/EnvThesDev/USLterCV_6	http://www.w3.org/2004/02/skos/core#notation	"[controlled by] U Grandin, 2013-08-22"
http://vocabs.lter-europe.net/EnvThesDev/USLterCV_6	http://www.w3.org/2004/02/skos/core#notation	"[source of definition] Begon, M., et al. (1996). Ecology Individuals, Populations and Communities. Cambridge, Blackwell Science."
http://vocabs.lter-europe.net/EnvThesDev/USLterCV_6	http://www.w3.org/2004/02/skos/core#definition	"The number of organisms in a population, guild, life form, etc., combining 'intensity' (density within inhabited areas) and 'prevalence' (number and size of inhabited areas)"@en
http://vocabs.lter-europe.net/EnvThesDev/USLterCV_6	http://purl.org/dc/terms/hasVersion	"2"^^
http://vocabs.lter-europe.net/EnvThesDev/USLterCV_6	http://purl.org/dc/terms/contributor	"kerteszmiklos@okologia.mta.hu"
http://vocabs.lter-europe.net/EnvThesDev/USLterCV_6	http://www.w3.org/2004/02/skos/core#broader	http://vocabs.lter-europe.net/EnvThesDev/10019

SLterCV 6		
http://vocabs.lter-europe.net/EnvThesDev/USLterCV_6	http://www.w3.org/2004/02/skos/core#related	http://vocabs.lter-europe.net/EnvThesDev/10022
http://vocabs.lter-europe.net/EnvThesDev/USLterCV_6	http://www.w3.org/2004/02/skos/core#related	http://vocabs.lter-europe.net/EnvThesDev/10026
http://vocabs.lter-europe.net/EnvThesDev/USLterCV_6	http://www.w3.org/2004/02/skos/core#related	http://vocabs.lter-europe.net/EnvThesDev/10018
http://vocabs.lter-europe.net/EnvThesDev/USLterCV_6	http://www.w3.org/2004/02/skos/core#related	http://vocabs.lter-europe.net/EnvThesDev/USLterCV_52
http://vocabs.lter-europe.net/EnvThesDev/USLterCV_6	http://www.w3.org/2004/02/skos/core#related	http://vocabs.lter-europe.net/EnvThesDev/10021
http://vocabs.lter-europe.net/EnvThesDev/USLterCV_6	http://www.w3.org/2004/02/skos/core#related	http://vocabs.lter-europe.net/EnvThesDev/10020
http://vocabs.lter-europe.net/EnvThesDev/USLterCV_6	http://www.w3.org/2004/02/skos/core#narrower	http://vocabs.lter-europe.net/EnvThesDev/10085
http://vocabs.lter-europe.net/EnvThesDev/USLterCV_6	http://www.w3.org/2004/02/skos/core#related	http://vocabs.lter-europe.net/EnvThesDev/10089
http://vocabs.lter-europe.net/EnvThesDev/USLterCV_6	http://www.w3.org/2004/02/skos/core#related	http://vocabs.lter-europe.net/EnvThesDev/10106
http://vocabs.lter-europe.net/EnvThesDev/USLterCV_6	http://purl.org/dc/terms/modified	"2014-12-17T15:48:02Z"^^

Annex 2 – Documentation of work

Workshops

1. ExpeER EnvThes consolidation kick off, 12th-13th February 2014

Location: **Umweltbundesamt/ Environment Agency Austria**, Brigittenauer
Lände 58, A – 1200 Vienna, Austria
Meeting Room BGL 5

Objectives of the meeting: Setting up the team of EnvThes Editors within ExpeER
Definition of the objectives by end of 2014
Definitions of tasks to achieve the objectives

	Part one “Introduction and expectations” Wednesday, 12 th January, 2014
13:00 - 14:00	Welcome and coffee
14:00 - 15:00	Introduction of participants and EnvThes Objectives of the meeting
15:00 - 18:00	Expectations and probable key factors of the project (incl. Coffee Break around 16:30)

	Part two “Activities on units and parameters” Thursday, 13 th January, 2014
9:00 – 10:30	Ideas and discussion about activities on units
10:30 – 10:45	<i>Coffee Break</i>
10:45 – 12:15	Ideas and discussion about activities on parameters
12:15 – 13:15	<i>Lunch Break</i>

	Part three “collaboration and tasks” Thursday, 13 th January, 2014
13:15 – 14:30	Ways of collaboration and exchange of data and content
14:30 – 15:30	Definition of tasks
15:30 – 16:00	Conclusion and next steps
16:00	End of Meeting

Task list from ExpeER EnvThes consolidation kick off, 12th-13th February 2014:

Number	Task	Who ? (responsible persones in capital letters)
1	Mining units, dimensions and parameters from existing DBs	HERBERT, Alessandro Oggioni, Tomas

2	Structuring the list of mining; Find relations between Unit and Parameter	MIKI, MARK, HERBERT
3	Coordination of units, parameter with ANAEE-France	NIC, Damien
4	Concepts from experimental platforms	MARK, DARIO (Laura, Carlo), DAMIEN
5	Synchronize EnvThes with ECOPAR and DEIMS	MARK (content) HERBERT, NIC (technic)
6	Revising rules for naming parameter	GÜNTER
7	Review of existing parameters in EnvThes	HERBERT, GÜNTER, Miki
8	Clarify the technical feasibilities	HERBERT, NIC
9	Check option EnvThes/ECOPAR	HERBERT, NIC
10	Review of methods in EnvThes	MIKI
11	Collaboration with ILTER	NIC, HERBERT (technic)
12	Organizing a technical workshop inviting others in June 2014	NIC

2. EnvThes in ExpeER, 9th to 10th December 2014

Location: **Umweltbundesamt/ Environment Agency Austria**, Brigittenauer
Lände 58, A – 1200 Vienna, Austria
Meeting Room (BGL 1 / BGL VIII)

Objectives of the meeting:

- Clarify and decide structure of parameters
- Assign tasks

	Part one Tuesday, 9 th December 2014
12:00 - 13:00	Welcome and coffee
13:00 – 13:10	Objectives of the meeting, Adoption of Agenda (Günter Pfaff)
13:10 – 13:30	Overview on what has been done so far (Herbert Schentz)
13:30 – 14:00	Proposed basic structure (John Watkins or Nic Bertrand)
14:00 – 14:30	ECOPAR (Mark Frenzel)

14:30 - 15:00	Agronomic terms (Francesco Danuso)
15:00 – 15:30	Needs for Statistics (?)
15:30 - 16:00	<i>Coffee Break</i>
16:00 – 17:30	Establishment of main elements (structures and concepts for all parameters) (all together)

	Part two” Wednesday, 10 th December, 2014
9:00 – 13:00	Distribution of work for entering the structure, terms and definitions. (all together) <i>incl. Coffee Break around 10:30</i>

Task list from EnvThes in ExpeER, 9th to 10th December 2014

#	Task	Who	When
1	Access to EnvThes – AnaEE/AnaEE-France for Francesco, Dario	Günter	10.12.2014
2	Check ANAEE-France terms in viewing mode and tell Herbert, which should be implemented in EnvThes	Dario, Francesco	Open
3	Check top level concepts “events” and “processes” as the exist in EnvThes3 to see, which can be used in EnvThes	Dario, Francesco	Open
4	Check with Damien about the status and planning of the vocabulary of the ANAEE-France thesaurus	Nic	12.12.2014
5	Grant access for Francesco and Dario to the editing mode of EnvThes	Nic	11.12.2014
6	Editing “processes” in the editing mode of EnvThes	Francesco, Dario	31.1.2015 (tbc)
7	Develop a (video)-tutorial about the use / good practices of EnvThes (editing GUI)	Nic	15.12.2014
8	Edit measures from EnvThes_Parameters directly in EnvThes and document it in Google spreadsheet (ENVTHES_CONCEPTS_MEASURES_Work on Google Drive).	Miki	5.1.2015
9	Edit measures from ECOPAR (terrestrial) directly in EnvThes and document it in Google spreadsheet (ENVTHES_CONCEPTS_MEASURES_Work on Google Drive).	Miki & Ulf	5.1.2015
10	Edit measures from ECOPAR (rivers) directly in EnvThes and document it in Google spreadsheet	Mark (& Stefan	12.1.2015

	(ENVTHES_CONCEPTS_MEASURES_Work on Google Drive).	Stoll)	
11	Edit measures from ECOPAR (lakes & marine) directly in EnvThes and document it in Google spreadsheet (ENVTHES_CONCEPTS_MEASURES_Work on Google Drive).	John (& Alessandro, Mihai)	12.1.2015
12	Check status of definitions in EnvThesDev	Herbert	17.12.2014
13	Organise Skype-meeting for Thursday 18 th December to discuss progress, common understanding	Günter	11.12.2014

The ILTER workshop

SETUP

For the Workshop the following languages, which match the outlook of the ILTER community have been supported

Default Language: en

Other Languages: zh ro vi no hu de fi sv fr zh-tw sk da zh-cn it iw sr ms ko ar cs pl pt es ja

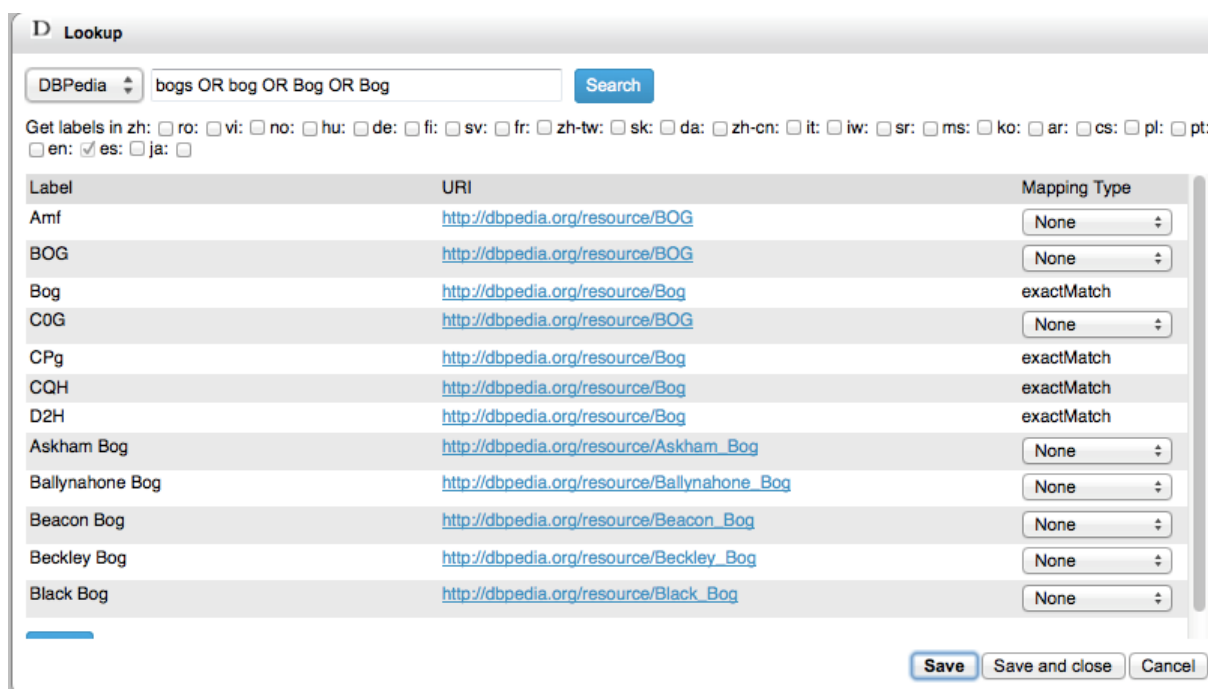
(for the developmentversion of EnvThes (EnvThesDev) the following languages have been added:

bg da el et hr mk nl ru sl sq tr)

Adding definitions and translations to multilingual thesaurus represent a significant human effort. During the workshop, we exploited some of PoolParty Linked Data capabilities to describe exact matches to terms present in dbpedia and import definitions and translations. (See screenshots below)

Establishment of Links

Linking to DBPedia within PoolParty Thesaurus Management System



D Lookup

DBPedia

Get labels in zh: ☐ ro: ☐ vi: ☐ no: ☐ hu: ☐ de: ☐ fi: ☐ sv: ☐ fr: ☐ zh-tw: ☐ sk: ☐ da: ☐ zh-cn: ☐ it: ☐ iw: ☐ sr: ☐ ms: ☐ ko: ☐ ar: ☐ cs: ☐ pl: ☐ pt: ☐ en: ☒ es: ☐ ja: ☐

Label	URI	Mapping Type
Amf	http://dbpedia.org/resource/BOG	None
BOG	http://dbpedia.org/resource/BOG	None
Bog	http://dbpedia.org/resource/Bog	exactMatch
C0G	http://dbpedia.org/resource/BOG	None
CPg	http://dbpedia.org/resource/Bog	exactMatch
CQH	http://dbpedia.org/resource/Bog	exactMatch
D2H	http://dbpedia.org/resource/Bog	exactMatch
Askham Bog	http://dbpedia.org/resource/Askham_Bog	None
Ballynahone Bog	http://dbpedia.org/resource/Ballynahone_Bog	None
Beacon Bog	http://dbpedia.org/resource/Beacon_Bog	None
Beckley Bog	http://dbpedia.org/resource/Beckley_Bog	None
Black Bog	http://dbpedia.org/resource/Black_Bog	None

Copying / Linking concepts to dbPedia

Selected Concept
bogs
http://vocabs.lter-europe.net/iltervocabdemo/USLterCV_70

[Remove as Top Concept](#)

SKOS Metadata Documents Tag Cloud Linked Data Triples Visualization

DBPedia LCSH Umbel Sindice Wordnet Dmoz

[Link to DBPedia](#) [Copy Data from DBPedia](#) [Delete DBPedia Data](#)

Linked Resources

<http://www.w3.org/2004/02/skos/core#exactMatch>
<http://dbpedia.org/resource/Bog>

Copied Resources

<http://www.w3.org/2004/02/skos/core#altLabel>

- Moor (de)
- Pantano (es)
- 酸沼 (zh)
- Torbiera (it)
- ボグ (ja)
- Mosse (sv)
- Torfowisko wysokie (pl)
- Suo (fi)
- Tourbière (fr)
- Myr (no)
- Myr (sv)
- Bog (en)

Exact matches to other vocabularies

ILTER Controlled Vocabulary Demo
Search the project

Overview Visual Browser SPARQL Endpoint

bogs (en)
Alternative Labels *bog (en)*
Concept URI
http://vocabs.lter-europe.net/iltervocabdemo/USLterCV_70

Broader Concept
wetlands (en)

Show 10 entries

Search:

Data Source	Linked Data	Relation
http://dbpedia.org	Moor (de)	Alternative Labels
http://dbpedia.org	Pantano (es)	Alternative Labels
http://dbpedia.org	酸沼 (zh)	Alternative Labels
http://dbpedia.org	Torbiera (it)	Alternative Labels
http://dbpedia.org	ボグ (ja)	Alternative Labels
http://dbpedia.org	Mosse (sv)	Alternative Labels
http://dbpedia.org	Torfowisko wysokie (pl)	Alternative Labels
http://dbpedia.org	Suo (fi)	Alternative Labels
http://dbpedia.org	Tourbière (fr)	Alternative Labels
http://dbpedia.org	Myr (no)	Alternative Labels

Showing 1 to 10 of 29 entries

NERC - Centre for Ecology & Hydrology

Translations

During the workshop in addition to pulling definitions and translations using PoolParty's links to DBpedia, attendees from Korea, Japan, China and Taiwan provided comprehensive translation for terms already captured in the US LTER vocabulary (US_LTER_CV).

The process of linking defining and translating terms could be further automated by using tools such as SILK (<http://silk.semwebcentral.org/>) and LOD-Manager .

Skype conferences

Skype-Meeting on characteristics 1, 7th July 2014

Decisions:

Number	Topic	Decision	Who ? (responsible persones in capital letters)
1	Naming & Rules	To avoid misunderstandings, the term parameter will be replaced by "characteristics", including then both statistical parameters and observed parameters	-

Information

Number	Topic	Information	Who ? (eg. For obtaining more detailed information)
1	Identification of references in ANAEE-France	Characteristics in AnaEE-France marked with "\$" as prefix indicate the reference who set the entry in the database	Damien

Next Steps / To Do:

What	When	Who (responsible persones in capital letters)
Look into AGROVOC	Until 17 th July	ALL
Look into GEMET	Until 17 th July	ALL
Check hierarchy from ECOPAR for use of EnvThes	Until 17 th July	MARK, HERBERT, Miki,
Send latest version of ANAEE-France thesaurus to Herbert	Until 10 th July	DAMIEN
Send IC-FAR Thesaurus to Herbert	Until 10 th July	FRANCESCO

Skype-Meeting on characteristics 2, 17th July 2014

Decisions

Number	Topic	Decision	Who ? (responsible persones in capital letters)

1	Naming & Rules	<p>Principles:</p> <ol style="list-style-type: none"> 1. Keep out conceptual relationships (is part of the ontology, to be developed later) 2. Make a simple structure 3. Concentrate on question, how things can be found in the Thesaurus 4. Main question: do we need that concept and its definition? 5. Keep terms have as atomic as possible 	-
2	Naming & Rules	<p>Categories for 1st level of hierarchy (parent terms):</p> <ul style="list-style-type: none"> • physical characteristics • chemical characteristics • biological characteristics • economic characteristics 	-

Information

Number	Topic	Information	Who ? (eg. For obtaining more detailed information)
1	Tools for linking Thesauri	SILK http://silk.wbsg.de/	Günter, Herbert
2	Tools for linking Thesauri	Openrefine http://openrefine.org/	Christian, Damien
3	Tools for linking Thesauri	Topbraid	John
4	Other Thesauri / Ontologies	CAST http://onto.nerc.ac.uk/CAST.html	John

Next Steps / To Do:

What	When	Who (responsible persons in capital letters)
Put Excel-File onto Google Drive	Until 17 th July	JOHANNES
Structure characteristics in Excel-File	ASAP after 17 th July	ALL
Organise / chair next Skype-Meeting	8 th August, 9:30-11:00	HERBERT / MARK

Skype-Meeting on additional sources, 8th August 2014

What	When	Who (responsible persones in capital letters)
Find bugs in EnvThes3	ASAP	CEH
Create EnvThes Instance without COL	ASAP	HERBERT
Make COL an own project	After creation of EnvThes without COL	HERBERT / CEH

Skype-Meeting on parameters and structure, 18th December 2014

1

Number	Topic	Decision	Who ? (responsible persones in capital letters)
1	Singular / plural	Concepts will be expressed in singular. Even expressions of USLter will be changed to singular	-
2	Upper case / lower case	Concepts will be written in lower case except of Names. Existing terms will be checked	
3	Compound terms	Terms should be expressed in their entirety except of compound terms with taxa	
4	Terms with taxa	Miki will establish examples with compound terms containing taxa (occurrence, abundance, diversity)	MIKI
5	Missing definitions	An EXCEL spreadsheet containing concepts, where definitions are missing shall help to find them in poolParty and edit them	HERBERT
6	USLTER terms	An EXCEL spreadsheet containing all terms coming from US LTER shall help to see, which terms should not be changed	HERBERT
7	matrix	Should be extent by plant tissue	done
8	interfaces	Vegetation – air will not be defined, as they only concern data (ontology!) and not terms	
9	reflectance	Not decided	
10	Events	The branch “events” of EnvThes3 will be put bellow an upper term “Event” in EnvThesDev, replacing process	HERBERT
11	Process	The branch process of EnvThes3 will be cleaned from double occurrence and entered below object of interest	HERBERT

Skype Meeting parameters and properties, 9th January 2015

Number	Topic	Decision / open issue
1	Plants parameters / properties	Open issue: Plants parameter and properties come together (sometimes with hierarchy sometimes without) Plants in a meadow Plants in a forest Plants in a lake Population dynamics Population changes over time
2	Object of interest / entity	Open issue: object of interest and entity come together
3	Place names	place names should not be below object of interest
4	monohierarchies / multihierarchies	we do not create multihierarchies we just have monohierarchies and related terms

Skype-Meeting Envthes and ANAEE-France, 9th January 2015

Number	Topic	Decision / open issue	Who ? (responsible personnes in capital letters)
1	Concepts in AnaEE-France Thesaurus	Check AnaEE-France Thesaurus for concepts, that should go to EnvThes	Damien Dario Laura
2	event	Event and lower concepts like human triggered event and nature triggered event should be entered	all