

IC_10 Domain extension of existing thesauri

1. Background

1.1 Short description

The multilingual SKOS/RDF thesaurus EnvThes integrates important terms used in long term ecological monitoring, research and experiments. It is based on US LTER Controlled Vocabulary, and has been extended by QDTY units and dimensions, EUNIS Habitats, INSPIRE spatial data themes and special concepts, needed by the EnvEurope and ExpeER communities. It links to other existing thesauri and vocabularies like GEMET, EArth, AgroVoc, EuroVoc and Wikipedia, EEA vocabularies. In this implementation case we want to extend EnvThes for the marine domain. Possible inputs can come from LifeWatch Italy, whose Interactions Thematic Centre developed a Phytoplankton Traits Thesaurus and is working on the Fish, Macrozoobenthos and Microzooplankton Traits Thesauri. Other RIs such as SeaDataNet, EMBRC could contribute to this extension as they are interested in linking with the biological community.

1.2 Contact

Background	Contact Person	Organization	Contact email
_ICT	e-Infrastructure]	Barbara Magagna	Umweltbundesamt GmbH Barbara.magagna@umweltbundesamt.at

1.3 Use case type

Implementation case

1.4 Scientific domain and communities

Scientific domain

[Atmosphere | Biosphere | hydrosphere | geosphere] – all of them

Community

Data Publication/Data Service Provision

Behavior

Semantic Harmonisation/Data Discovery and Access and Service Description/ServiceRegistration

Role:

Semantic Service Provider

2. Detailed description

Objective and Impact

Set steps to harmonize the conceptual worlds of the different RI thus enhancing semantic interoperability, a precondition for sharing metadata and data.

Challenges

- (1) Synonyms and exact matching terms: how to harmonize contradicting conceptual approaches in different communities
- (2) How to completely detect homonyms
- (3) How to deal with contradicting views of different communities.

Detailed scenarios

tbd

Technical status and requirements

EnvThes is an existing thesaurus used by the ecological monitoring and experimental communities (LTER, ANAEE) . It is based SKOS/RDF and is implemented using Topbraid thesaurus server with a Browser user interface, linkedData interface and SPARQL endpoint.
The technical status of the other vocabularies has to be evaluated (see below)
In case other formal languages than SKOS/RDF are used, translating machines will be needed.

Implementation plan and timetable

- 1) timeline: tbd
- 2) milestones:
 - (1) Find vocabularies that shall be integrated
 - (2) Analyze formal language of found vocabularies (e.g. SKOS/RDF) and related services. (e.g. SPARQL endpoint)
 - (3) Automatically search for matching strings
 - (4) Manually revise the concepts that have been found automatically in the different communities
 - (5) Develop a decision guidance across RIs
 - (6) Depending on organizational decisions establish links between vocabularies, and/or import concepts to vocabularies
- 3) involved RIs: LTER, EMBRC, LifeWatch Italy,...
- 4) *links to ENVRIPlus work packages / tasks*: 5.3, 8.2, 8.3
- 5) allocation of resources: tbd

Expected output and evaluation of output

1. Links between concepts in different vocabularies
2. Eventually merged vocabularies

External Links

1. IC_10 notebook: [+](https://envriplus.manageprojects.com/projects/wp9-service-validation-and-deployment-1/notebooks/634+)<https://envriplus.manageprojects.com/projects/wp9-service-validation-and-deployment-1/notebooks/634+>