

EUDAT

Building a semantic repository using B2SHARE

Herbert Schentz (herbert.schentz@umweltbundesamt.at)

Yann le Franc (ylefranc@gmail.com)

Semantic Working Group

EUDAT 3rd Conference September 2014



Overview

- About the semantic working group
- The need of a repository for ontologies and thesauri
- How to establish the repository
- Use of the repository
- Outlook: cross vocabulary searches

Semantic Working Group

- A precondition of multiple usages of data is understanding their meaning
- precondition for integration of data with different origin is their semantic interoperability.
- Semantic services are of common interest
- it is not a straightforward process to identify exactly which services in the semantic domain could be seen as common services shared by some or many communities.

Identified services of common interest

- Semantic annotation

Link data with concepts from controlled vocabularies to disambiguate them.

- Establishment of a vocabulary (ontologies and thesauri and codelist) repository

... to share vocabularies.

Why sharing vocabularies

- The aim to create concepts that are commonly used within science, in order to achieve broad comprehension is very old.
 - Periodic system of elements
 - SI units and dimensions
 - Taxonomies (Lineus)
- Digitalization of those Vocabularies has taken place.
- But there is a high need of harmonizations of different Vocabularies.

Repository of vocabularies

A repository of vocabularies will help to

- Become aware of existing vocabularies
 - for a certain scientific domain
 - for a certain research topic
 - cross scientific domains
- Lookup concepts and definitions in the vocabularies
- Reuse concepts or subgraphs of vocabularies

Examples of Repositories

- Linked open vocabulary

<http://lov.okfn.org/dataset/lov/>

Limited to vocabularies with linkedData interface and/or SPARQL endpoint

- Bioportal

<http://bioportal.bioontology.org/>

Limited to life sciences

- OBO foundry: Ontobee

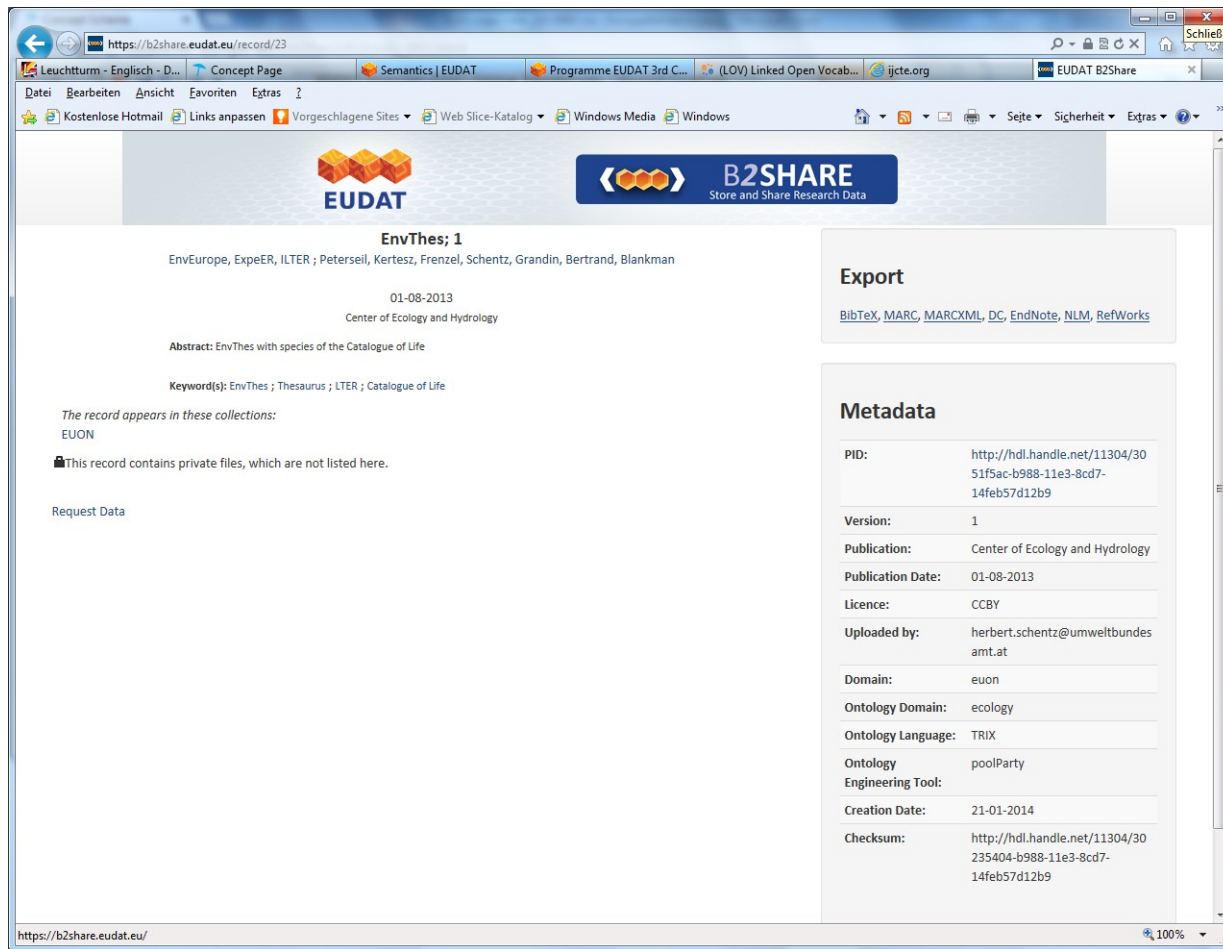
<https://github.com/ontoden/ontobee>

Limited to RDF / OWL

How to establish vocabularies

- Most vocabularies can be made available as files in specific formats:
 - OWL
 - SKOS / RDF
 - OBO
 - CYCL
 - Others
- Files can be loaded to B2SHARE and made available

Example of vocabulary in B2SHARE



The screenshot shows a web browser window displaying a record in the B2SHARE system. The browser's address bar shows the URL <https://b2share.eudat.eu/record/23>. The page header features the EUDAT logo and the B2SHARE logo with the tagline "Store and Share Research Data".

The main content area displays the following information:

- EnvThes: 1**
- EnvEurope, ExpeER, ILTER ; Peterseil, Kertesz, Frenzel, Schentz, Grandin, Bertrand, Blankman
- 01-08-2013
- Center of Ecology and Hydrology
- Abstract:** EnvThes with species of the Catalogue of Life
- Keyword(s):** EnvThes ; Thesaurus ; LTER ; Catalogue of Life
- The record appears in these collections:*
- EUON
- 🔒 This record contains private files, which are not listed here.
- Request Data

On the right side, there are two panels:

- Export**: Includes links for [BibTeX](#), [MARC](#), [MARCXML](#), [DC](#), [EndNote](#), [NLM](#), and [RefWorks](#).
- Metadata**: A table of key-value pairs:

PID:	http://hdl.handle.net/11304/3051f5ac-b988-11e3-8cd7-14feb57d12b9
Version:	1
Publication:	Center of Ecology and Hydrology
Publication Date:	01-08-2013
Licence:	CCBY
Uploaded by:	herbert.schentz@umweltbundesamt.at
Domain:	euon
Ontology Domain:	ecology
Ontology Language:	TRIX
Ontology Engineering Tool:	poolParty
Creation Date:	21-01-2014
Checksum:	http://hdl.handle.net/11304/30235404-b988-11e3-8cd7-14feb57d12b9

Use of repository



The formal
Language of the
Vocabulary is crucial:

You need an appropriate
Editor.
Therefore the engineering
Tool is mentioned too!

Outlook: cross vocabulary searches

- LinkedOpenVocabulary and bioPortal are good examples, but, unfortunately limited to one domain and/or one formal language.
- Both offer cross vocabulary search for concepts:

Example Search within LOV



LINKED OPEN VOCABULARIES

SEARCH ecosystem

11
results

wlo:EcosystemRole (wlo) 0.556
n/a (use in LOD)
<http://purl.org/ontology/wlo/EcosystemRole>
rdfs:label **Ecosystem** Role @en-gb
localName **Ecosystem**Role
rdfs:comment **Ecosystem** roles are about the part an animal or @en-gb

geosp:Omernik_Level_III_Ecoregion_Classification (geosp) 0.466
10 occurrences in 1 LOD datasets
http://rdf.geospecies.org/ont/geospecies#Omernik_Level_III_Ecoregion_Classification
dce:description reflect differences in **ecosystem** quality and

mtlo:BC12_Ecosystem (mtlo) 0.319
n/a (use in LOD)
http://www.ics.forth.gr/isl/MarineTLO/v4/marinetlo.owl#BC12_Ecosystem
localName BC12_Ecosystem

mtlo:BT7_Ecosystem_Type (mtlo) 0.254
n/a (use in LOD)
http://www.ics.forth.gr/isl/MarineTLO/v4/marinetlo.owl#BT7_Ecosystem_Type
rdfs:comment . This coherence identifies the **ecosystem** and not @en
localName BT7_Ecosystem_Type

mtlo:BC14_Ecosystem_Environment (mtlo) 0.186

Type

class (10)
property (1)

Tag

biology (10)
environment (1)

Vocabulary

mtlo (9)
geosp (1)
wlo (1)

This can help reduce the reinvention of wheels

Big challenge

- However:
 - Without limitations to a scientific domain and/or formal language, this is a big challenge:
 - The establishment of a Repository for Vocabularies within B2SHARE is a starting point.