LusTRE: a Linked Thesaurus Framework for Environment

Riccardo Albertoni\textsuperscript{1}, Monica De Martino\textsuperscript{1}, Paola Podestà\textsuperscript{1}, Paolo Plini\textsuperscript{2}

\textsuperscript{1}CNR-IMATI, Via De Marini, 6, Torre di Francia, 16149 Genova, Italia
{\text{name.surname}@ge.imati.cnr.it}

\textsuperscript{2}CNR-IIA-EKOLab - Via Salaria Km 29,300 C.P. 10,I-00016 Monterotondo stazione RM, Italia, plini@iia.cnr.it

W3C-IT LOD 2014, Rome, Italy, 20-21 Feb 2014
INSPIRE implementation rules

- recommend the adoption of (multilingual) thesauri when compiling metadata for data/services

However

Different thesauri have been developed, and may be deployed for cataloguing the geographical, e.g.,

AGROVOC    EARTH    GEMET    THiST    ...

Thesauri are heterogeneous wrt thematic coverage, multilinguality, granularities, popularity in certain communities

Heterogeneity is precious!!!
LusTRE: Enabling Thesauri Joint Exploitation

Design Principle (NatureSDIPlus 2009-2011)

- **Modularity**: To add new KOS as a new module plugged in the set of thesauri in the TF
- **Openness**: To easily extend each KOS keeping separated the original one
- **Interlinking**: Linking among the terms referring to the same concepts in more than one thesaurus in order to harmonize their usage.
- **Exploitability**: To encode in a standard and flexible format in order to encourage the adoption and its enrichment from third party system

Simple Knowledge Organization System (SKOS) to encode the thesaurus content

Linked Data best practices to publish the thesaurus in machine understandable format

TF Extension (eENVplus 2013-2015)

- Publication of further thesauri not yet exposed as Linked data
- Interlinking with well-known LOD environmental related thesauri
- Services to access LusTRE and cross-walking from a thesaurus to another


Common Thesaurus Framework (TF)
INSPIRE Themes: Habitat and biotopes, Species distribution, Biogeographical regions, Protect sites

What is available as result of NatureSDIPlus?

<table>
<thead>
<tr>
<th>KOS</th>
<th>Sources</th>
<th>Number of terms</th>
<th>Languages</th>
<th>Data Theme</th>
</tr>
</thead>
<tbody>
<tr>
<td>EARTH</td>
<td>CNR</td>
<td>14340</td>
<td>en, it</td>
<td>General level</td>
</tr>
<tr>
<td>IUCN classification</td>
<td>IUCN</td>
<td>8</td>
<td>en</td>
<td>Protected sites</td>
</tr>
<tr>
<td>Habitat types</td>
<td>EUNIS DB</td>
<td>5431</td>
<td>es, nl, el, pt, it, fr, da, fi, sv, de</td>
<td>Habitats and biotopes</td>
</tr>
<tr>
<td>Nature 2000 A I</td>
<td>EUNIS DB</td>
<td>183447</td>
<td>la, hu, et, pt, es, fr, ro, it, mt, fi, sq, da, no, sv, it, lv, cs, sk, sl, hr, pl, nl, lb, en, bg, ca, de, gr, el, ga, mk, nb, ru, sr, lt, uk</td>
<td>Species distribution</td>
</tr>
<tr>
<td>Main threats to biodiversity by biogeographic region</td>
<td>EEA</td>
<td>12</td>
<td>en</td>
<td>Bio-geographical regions</td>
</tr>
<tr>
<td>Digital Map of European Ecological Regions</td>
<td>CMEER</td>
<td>About 68</td>
<td>en</td>
<td>Bio-geographical regions</td>
</tr>
</tbody>
</table>
**Survey on Environmental Thesauri**

<table>
<thead>
<tr>
<th>Terminologies in the LOD to be interlinked</th>
<th>Terminologies to be included in the TF</th>
</tr>
</thead>
<tbody>
<tr>
<td>• AGROVOC,</td>
<td>• ShowTerm,</td>
</tr>
<tr>
<td>• Eurovoc,</td>
<td>• EOSterm,</td>
</tr>
<tr>
<td>• SoilThes,</td>
<td>• UMTHES,</td>
</tr>
<tr>
<td>• NERC,</td>
<td>• ThIST</td>
</tr>
<tr>
<td>• ThesSoz,</td>
<td>• Inspire IFCD and INSPIRE Glossary</td>
</tr>
<tr>
<td>• Geological Survey of Austria (GBA)-thesaurus,</td>
<td>• EEA-EIONET Data Dictionaries</td>
</tr>
<tr>
<td>• EnvThes</td>
<td>• EEA-EIONET AQ pollutants</td>
</tr>
<tr>
<td></td>
<td>• One Geology, IUGS- CGI vocabularies</td>
</tr>
</tbody>
</table>

*M. De Martino (CNR-IMATI), R. Albertoni (CNR-IMATI), P.Podestà(CNR-IMATI), C. Cipolloni (ISPRA) P. Plini (CNR-IIA), D 4.1 – Survey on environmental thesauri, eENVplus, December 2013*
Server re-engineering
we moved from D2R + Mysql to Virtuoso + Pubby to have
+ More flexibility (Named Graph)
+ Better performance
+ Materialization of SKOS entailments

EARTh included into the Linked Open Data Cloud:
http://datahub.io/it/dataset/environmental-applications-reference-thesaurus

Albertoni R., De Martino M., Di Franco S., De Santis V., Plini P.:
EARTh: An Environmental Application Reference Thesaurus in the Linked Open Data cloud.
Future actions

Further thesauri and linksets  (end of the year)

Evaluation of provided content

• Quality of thesauri included in LusTRE, e.g., deploying qSKOS
  

• Quality and usefulness of Linkset
  • We are developing in-house quality measures for Linksets extending
... Waiting for next LusTRE’s release ... We invite you:

- To take a look at the Thesaurus Framework at: http://linkeddata.ge.imati.cnr.it:2020
- To check if a term is contained in TF at: http://linkeddata.ge.imati.cnr.it:8890/fct/
- To access the SPARQL ENDPOINT at: http://linkeddata.ge.imati.cnr.it:8890/sparql
  - To querying thesaurus concepts, relationships thesaurus
  - To mappings between EARTH to GEMET, AGROVOC, UMTHES...
  - To build your own services/application on the TF
- To interlink your vocabularies/thesauri with LusTRE’s thesauri

For info and support contact albertoni@ge.imati.cnr.it  demartino@ge.imati.cnr.it
Further in-house technology

SSONDE,
• a Open Source Framework,
• providing an instance similarity
• enabling in a detailed comparison and ranking of resources through the comparison of their RDF ontology driven metadata
• code available at https://code.google.com/p/ssonde/
