EARTh vs INSPIRE:
an Environmental Thesaurus in the Linked Open Data Cloud

S. Di Franco[a],P. Plini[a], R. Albertoni[b], M. De Martino[b]

[a]CNR-IIA-EKOLab, Roma, Italy
[b]CNR-IMATI, Genova, Italy

Although different directives (e.g. INSPIRE) and policy communications (e.g. SEIS) have been launched at European-scale with the objective of improving the management of heterogeneous environmental data sources, an effective sharing of these resources is still part of the desiderata due to the intrinsic multicultural and multilingual nature of the environmental domain.

Thesauri are widely employed as common ground enabling communication among the different communities working in environment-related domains: they allow users to share and agree upon scientific/technical terms in the target domain and to express them in multiple languages. In the recent years several structured controlled vocabularies (thesauri) have been deployed by different communities having a large spectrum of competencies. They have been created embodying different points of view based on different ways of conceptualization. Their development reflects different scopes and implies quite a range of levels of abstraction and detail.

Nowadays networked information access to heterogeneous environmental data sources requires interoperability of controlled vocabularies. The Linked Data paradigm jointly with Simple Knowledge Organization System provides a promising framework to face with the aforementioned problems: it allows representing and publishing distinct thesauri and their interlinks as a whole enabling to switch among them.

This paper presents EARTh LOD, the latest release of EARTh[1], the Environmental Applications Reference Thesaurus (ver. Linked Data 1.3) that takes advantage of this framework providing a SKOS dataset recently included in the Linked Open Data Cloud. Compared to other environmental thesauri available as Linked Data like AGROVOC[2], EUNIS[3], Geological Survey of Austria (GBA) Thesaurus[4], EARTh provides a more general purpose and thematically neutral terminological support. Compared to the GEneral Multilingual Environmental Thesaurus (GEMET)[5], namely the *de facto* general purpose thesaurus standard, EARTh provides a minor multilingual support, but it extends GEMET with more than 9000 concepts arranged into a completely revised hierarchical structure. Being one of the largest general purpose and structured environmental terminological resources available in the LOD cloud, EARTh is expected in close future to become a good linking point serving like a kind of bridge in the integration of other terminological resources dealing with environmental topics. Currently, EARTh LOD already includes more than 12000 links towards thesauri such as GEMET, AGROVOC, EUROVOC and UMTHES enabling in the traditional thesaurus-based indexing of digital resources, as well as the use of digital resources across multi-thesauri applications and platforms.

---

[1] http://datahub.io/dataset/environmental-applications-reference-thesaurus
[2] http://aims.fao.org/website/AGROVOC-Thesaurus/sub
[3] http://eunis.eea.europa.eu/
[4] http://thedatahub.org/dataset/geological-survey-of-austria-thesaurus
[5] http://www.eionet.europa.eu/gemet

Currently a work in progress is performed within the European eENVplus project (http://www.eenvplus.eu/) aiming at extending and creating a multilingual thesaurus framework for the environment (De Martino et al. 2010) and to implement services for its exploitation for metadata indexing and data discovery. One of the main actions is making EARTh compliant with the INSPIRE directive and to become with GEMET, thanks to their synergy, effectively the hub for the bridging to other environmental related vocabularies for the Environment. It includes improvement in terms of EARTh content as well as its Linked Data publication and consuming.

Concerning the content, an overall revision of the thesaurus structure and content is currently undergoing as consequence of the recent publication of ISO 25964-1:2011 (ISO, 2011) and ISO 25964-2:2013 (ISO, 2013). The number of concepts and lexical correspondents of EARTh are expected to increase also as a consequence of a larger adoption by environmental communities: it will conceptually cover the main concepts of the 34 Annex I,II and III data themes of INSPIRE. In terms of linguistic equivalents it would be improved by exploiting its synergy with GEMET multilingual vocabulary

Concerning the Linked Data publication and consumption, novel releases are expected to overcome current limitations:

- Further materializations (e.g., entailments of skos:semanticRelation) will be made available in the future releases; RT properties, which have been indistinctly mapped into skos:related in order to avoid the adoption of user-defined RDF vocabularies, will be differentiated as in the original version of EARTh;

- Semantic interoperability of EARTh with other thesauri will be strengthened by exploiting EARTh LOD within the Thesaurus Framework in eENVplus project: (i) harmonization of EARTh and GEMET in order to bring an added value to the general purpose multilingual thesaurus for the environment; (ii) identification of interlinking with the other environment-related target thesauri (e.g., SnowTerm, a specific domain thesaurus on polar and mountain environment; ONEGEOLOGY.) according to the environment-related issues coming from the INSPIRE data themes; (iii) semantic interoperability of EARTh with other thesauri dealing with geographical information tools (e.g EOSterm, a thesaurus on Remote Sensing and GIS terminology).

- Advanced semantic explorative services to consume EARTh LOD and its interlinking to search for dataset will be developed by shifting from on thesaurus to another: (i) services for the metadata editing to data provider, (ii) services to exploit the thesaurus knowledge for end-user purpose (e.g. digital-content description, and search as well as translation support), (iii) semantics explorative service to navigate the thesauri.

Reference

(M. De Martino, et al, 2011) M. De Martino and R. Albertoni, A multilingual/multicultural semantic-based approach to improve Data Sharing in a SDI for Nature Conservation, International Journal of Spatial Data Infrastructures Research, vol.6, 2011, ISSN 1725-0463, pp. 206-233.

(ISO, 2011) ISO 25964-1 Information and documentation - Thesauri and interoperability with other vocabularies - Part 1: Thesauri for information retrieval, 2011

(ISO, 2013) ISO 25964-2 Information and documentation - Thesauri and interoperability with other vocabularies - Part 2: Interoperability with other vocabularies , 2013