European conference of the Czech Presidency of the Council of the EU
TOWARDS eENVIRONMENT
Opportunities of SEIS and SISE: Integrating Environmental Knowledge in Europe
March 25-27, 2009 | Prague, Czech Republic
www.e-envi2009.org

Summary of W4 eEnvironment Terminology

1.   General Terminology Use Cases

The above mentioned conference showed that there are common terminology use cases along with any kind of environmental reporting such as:

- *Clarify domain specific understanding of concepts*
- *Agree on standardized classifications*

For example: What is a „water use" (in terms of m³/year)? Water abstracted for irrigation purposes? Does it include water taken from the tap? Is cooling water considered to be waste water or not? Questions like these (Nagy 2009) need to be precisely clarified to make reported data comparable. There need to be common statistic classifications along with EUROSTAT (Heidorn 2009). This is how domain experts are involved in terminology work when they try to achieve harmonisation.

Usually this is done in committees with results written down in papers which are hard to find and so cannot be referred to in a transparent manner.

This is the point where terminology experts can provide help and support sustainability of terms and definitions by two contributions:

- *Formalise explicit vocabularies*
- *Publish these vocabularies in the web*

Formal terminology structures such as Thesauri or Web Ontology (Schleidt 2009) provide means to make different meanings of apparently identical terms explicit and transparent to all users.

Based on semantic web technologies, each term is represented by a HTTP Uniform Resource Locator (URL) which resolves to a description of the term along with hyperlinks to related terms with their distinct meaning. This again allows published data to be linked to the precise terms respectively. What is left to the report suppliers is:

- *Reference these vocabularies from the data.*

(Use HTTP URL to do so!)

2. Identified Trends

The eTerminology workshop participants identified two major trends in terminology of the recent years.

2.1 Closing the gap between data, metadata and terminology

While data, metadata and terminology have been kept separately from each other in the early years, new semantic technologies are closing this gap and provide a seamless meaningful navigation (Bandholtz 2009). Based on vocabularies that are published in the Web the data providers may refer to the appropriate terms:

- *Reference terms by hyperlink to vocabulary i.e. indicators, unit of measurement*

This given, users may start on Web pages that give proper definitions of terms and navigate to related data from there, and vice versa. This is widely demonstrated by the *Linking Open Data* project as part of the W3C Semantic Web activity.
http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData

2.2 Adding semantics to vocabularies

While early vocabularies mainly consisted of simple lists of terms (either flat or hierarchically), current term collections tend to add semantics and meaningful relations:

- *Express vocabularies as rich thesauri or Web ontologies*

Rich thesauri (Fock 2009, Plini 2009) tend to carefully extend the traditional set of relations (in particular the associative-RT and equivalence-USE/UF) by selected named relations such as "effected_by" or "takes_place_in" (Reimerink 2009) in order to extend the semantic expressivity. This is influenced by artificial intelligence patterns, but it avoids any infinite complexity which would lead away from the intended usage (Morrison 2009).

While the Web ontology Language (OWL) provides a standardised formalism to express both rich thesauri and artificial intelligence, terminology experts recommend the use of light-weight Ontology that allow for simple reasoning and fast processing.

- *Differentiation between traditional knowledge ontologies and the use as rich terminology i.e. thesaurus with enhanced relations.*

3.   Requirements

Seeing these use cases and current trends in terminology, the eTerminology participants identified several requirements concerning the improvement of terminology management and application.

3.1   Terminology Governance

Each terminology applicable for usage in SEIS needs an assigned editorial board. This is based on the definitions provided by domain experts (Prüller 2009), but it needs to develop a governance strategy and practice (Legat 2009) including base lines such as:

- *Systematic collection, structuring, validation of terms*
- *Collect multiple definitions in separate space*
- *Vocabulary ownership and licensing*
- *Versioning strategy*
- *Guarantee persistence through "Cool URIs" – response depends on requested media type*

3.2   Federation of Vocabularies

There will be multiple sources and editorial boards, depending on domain or language specific communities. It is not intended nor does it appear to be realistically to harmonise a one-and-only over-all vocabulary. Therefore we need methods to build a federation of multiple vocabularies addressing issues such as:

- *Alignment of multiple vocabularies*
- *Links for mappings, cross-references*
- *Translations between language-specific concepts*
- *Separate conceptual frameworks*
- *Trust and provenance*

3.3   Specific Tools and Resources

Understood that such terminologies will incorporate the common semantic backbone of SEIS, there is a need of specific resources. Human resources and dedicated tools need to be provided by the stakeholder authorities along with the dedicated reporting activities. Industry shall be integrated based on open source communities that must not prohibit any further commercial utilization of supplied work packages (Bügel 2009).

- *Human resources for the work of editorial boards*
- *Tools for management and dissemination/distribution/referencing*
- *Support for open source initiatives*

4. References

All mentioned contributions can be downloaded:

Conference proceedings: http://www.e-envi2009.org/?proceedings
Workshop presentations: http://www.e-envi2009.org/?presentations#w4

Contact:
Thomas Bandholtz (Workshop Chair)
innoQ Deutschland GmbH
Halskestr. 17
D-40880 Ratingen
Germany

Tel: +49 178 4049387
thomas.bandholtz@innoq.com