

# Preventing Interoperability Problems Instead of Solving Them

Eero Hyvönen

Aalto University, School of Science and Technology, and University of Helsinki

Semantic Computing Research Group

<http://www.seco.tkk.fi>, [first.last@tkk.fi](mailto:first.last@tkk.fi)

## Abstract

A major source of interoperability problems on the Semantic Web is the different vocabularies used in metadata descriptions. This paper argues that instead of *solving* interoperability problems we should focus more effort on *avoiding* the problems in the first place, in the spirit of Albert Einstein's quote "*Intellectuals solve problems, geniuses prevent them*". For this purpose, coordinated collaborative development of open source vocabularies and centralized publication of them as public vocabulary services are proposed. Methods, guidelines, and tools to facilitate this have been developed on a national level in the Finnish FinnONTO initiative, and are now in pilot use with applications and promising first results.

## 1. Interoperability of Vocabularies

Much of the power of the Web comes from the freedom for anybody to publish and link own content on the web as the Web of Pages. When moving into the era of the Semantic Web, the Web of (Linked) Data, content is being linked on the level of ontological concepts and metadata underlying the pages<sup>1</sup> (Bizer et al., 2009) This leads to interoperability problems, especially interoperability regarding metadata schemas and vocabularies used for filling element values in the schemas. Approaches to schema interoperability include the dumb-down principle, as suggested in the Dublin Core community<sup>2</sup>, and using a shared schema ontology onto which other metadata representations can be transformed, as suggested e.g. by the CIDOC CRM and FRBR communities<sup>3</sup>. In contrast, this paper focuses on interoperability problems due to domain vocabularies (ontologies) used in annotations.

Content aggregated in semantic portals or on the web scale in the Linked Data initiative comes from actors and organizations that produce content for their own purposes and come from different disciplines, cultures, and countries. As a result, lots of different, partly overlapping vocabularies are used in metadata descriptions. To approach the interoperability problems, various techniques of ontology matching (mapping) (Euzenat, Shvaiko, 2007) are used. For example, lots of mappings based on the owl:sameAs relation have been created for the resources in the Linked Data cloud. There are, for example, mappings between the place resources of DBpedia and Geonames. A key problem here is how to deal with situations, where multiple entity names and identifiers are used for a single real world object (Bouquet et al., 2008), and where different objects have the same name or identifier. The same problem is encountered in Web 2.0 sites, where tagging using literals without identified meaning is causing more and more semantic confusions as more and more tags are being created (e.g. "jaguar" as a car model, animal, and airplane).

## 2. Coordinated Collaboration for Vocabulary Creation

The mess of meaning references on the metadata level on the Semantic Web creates lots of interesting research problems to study. Most research on interoperability issues seems to be focusing on developing methods and tools for obtaining interoperability between heterogeneous

---

<sup>1</sup> <http://linkeddata.org/>

<sup>2</sup> <http://dublincore.org/>

<sup>3</sup> <http://cidoc.ics.forth.gr/>

annotations (e.g. the datasets of the Linked Data initiative). However, from a non-academic practical viewpoint, this is a problem that should be avoided in the first place as far as possible. Obviously, more research effort should be focused on developing methods, tools, and practices by which metadata could be produced on a larger scale in an interoperable way at the time of creating it. Instead of solving interoperability problems we should rather try to prevent them by better ontology services, coordination, and collaboration in ontology development and content creation.

FinnONTO 2003-2012 is a research project and a Living Laboratory experiment (Hyvönen et al., 2008; Hyvönen, 2009), where the idea is to establish a collaboration framework for vocabulary development and services on a national level for the Semantic Web. The main goal of FinnONTO is to create an open source, national level cross-domain "content infrastructure" for the Semantic Web, aligned with international vocabularies, standards, and practices. This infrastructure and network of concepts can be paralleled, on a conceptual level, with the construction of railroad, electrical, or telephone networks in the past.

The work is based on the domain independent Semantic Web standards<sup>4</sup> of the W3C, such as RDF, SKOS, OWL, and SPARQL, but the hearth of the system is domain-specific ontologies. While standardization work at W3C focuses on defining general principles of ontological structuring and reasoning, such as subsumption and inheritance, the general goal of FinnONTO is to facilitate cross-domain interoperability of metadata descriptions on a domain-specific vocabulary level. The idea is that when content is published on the web, it should be possible to connect it semantically with other related (cross-domain) contents based on a system of mutually aligned domain ontologies.

The vocabulary infrastructure has been built by transforming nationally used traditional keyword thesauri into light-weight ontologies, which makes the ontologies interoperable with already indexed content in databases. A key goal in the work is to encourage collaboration between ontology developers of different domains by proving a general FinnONTO ontology framework in which new vocabularies can be aligned already during the ontologization process, instead of afterwards. The kernel of the FinnONTO system (Hyvönen, 2008) is the General Finnish Ontology YSO developed from the widely used General Finnish Thesaurus YSA that consists of some 25,000 general concepts and that is maintained by the National Library. The corresponding ontology YSO has been extended by various domain-specific daughter ontologies, based on other national thesauri used in domains such as cultural heritage, agriculture and forestry, applied arts, places, photography, and others. These ontologies create together virtually one ontology, the Collaborative Holistic Ontology KOKO, that now has over 70,000 general concepts, excluding ontologies such as places, mammals, persons, historical events, or birds of the world.<sup>5</sup>

### 3. Commandments for Social Vocabulary Development

The key idea in the ontologization process is to aim at a system of vocabularies that are *born interoperable* with each other. To facilitate this, a new thesaurus is first matched with the general YSO top ontology in order to identify potential overlaps. The result is a Protégé-editor<sup>6</sup> project that includes YSO and the new thesaurus concepts. This structure is then corrected and maintained manually. (Alignment with other overlapping ontologies is also possible in a similar way.) In this way, the work already done in YSO can be reused in daughter ontologies and, at the same time, interoperability is enhanced by collaboration.

---

<sup>4</sup> <http://www.w3.org/standards/semanticweb/>

<sup>5</sup> <http://www.seco.tkk.fi/ontologies/>

<sup>6</sup> <http://protege.stanford.edu/>

Vocabulary work in our view is as much a social process as it is a technical challenge. The work is guided by the following ten commandments: 1) Add machine semantics. Start transforming thesauri into machine interpretable (lightweight) ontologies in order to boost their usage on the Semantic Web. 2) Think cross-domain. Consider not only your own micro world, but also cross-domain usage of concepts when making ontological decisions. 3) Establish collaboration networks of domain expert groups. Nobody masters the whole universe. 4) Reuse the others' work. 5) Maintain interoperability with the past and other ontologies. Otherwise benefits of collaboration are lost. 6) Proceed in small steps. Adding even little semantics can be very useful (and keeps e.g. the funding agencies happy). 7) Respect different ontological views. It is not possible to come up with only one ontological view of the world. 8) Accept imperfect models. The ontology will never be fully perfect. 9) Minimal ontological commitment. Keep ontological structures simple and generic in order to facilitate cross-domain reuse. 10) Coordinate the work and add new commandments if needed. This is done now by the FinnONTO research project but later, if the project is successful, by another coordinating organization.

#### 4. Vocabulary Services for Legacy Systems

Another key component of the FinnONTO infrastructure is the National Ontology Service ONKI<sup>7</sup> (Viljanen et al., 2009) hosting currently 79 ontologies and vocabularies. The idea is provide the vocabularies as a free service for both human and machine users to utilize. ONKI ontology services such as concept finding, browsing, fetching, and query expansion (Tuominen et al., 2009ab) can be integrated with legacy systems through REST, Web Service, or AJAX APIs in a way that is analogous to using Google Maps as an external service in applications. We hope that by making vocabulary services available and usable in an economically motivating way, organizations and people start using shared ONKI vocabularies and URIs, preventing interoperability problems arising from using local or depreciated vocabularies. Other ontology servers on the web with the goal of publishing and sharing ontologies in public include Cupboard (d'Aquin, 2009) and BioPortal (Musen et al., 2009).

#### 5. Evaluation

The feasibility of the FinnONTO approach is tested and demonstrated in practice by applications, such as the collaborative semantic portals MuseumFinland<sup>8</sup>, HealthFinland<sup>9</sup>, and CultureSampo<sup>10</sup> (Hyvönen et al., 2009) that makes use of the whole KOKO system aligned with some international vocabularies, such as the Getty vocabularies<sup>11</sup> AAT, TGN, and ULAN. In summer 2009, 150 organizations in Finland and abroad had been registered to use ONKI services, and new ontologized vocabularies in the system have been developed by external organizations, e.g. an ontology for maritime and literature content (KAUNO). The latter one that has been used, based on ONKI services, for annotating 55,000 pieces of Finnish novels in a Web 2.0 fashion by Finnish librarians for a semantic literature portal<sup>12</sup>. In HealthFinland metadata is being created using the ONKI ontologies and services by a variety of national health organizations, and the system is production use since 2009 (Suominen et al., 2010).

Our own experience suggests that gaining semantic interoperability in terms of vocabularies is a very tedious task and hinders fast publication cycle from legacy databases to Web. In

---

<sup>7</sup> <http://www.onki.fi/>

<sup>8</sup> <http://www.museosuomi.fi/>

<sup>9</sup> <http://www.terveysuomi.fi/>

<sup>10</sup> <http://www.kulttuurisampo.fi/>

<sup>11</sup> [http://www.getty.edu/research/conducting\\_research/vocabularies/](http://www.getty.edu/research/conducting_research/vocabularies/)

<sup>12</sup> <http://www.kirjasampo.fi/>

CultureSampo, for example, the content is heterogeneous and harvested from tens of memory organizations of different kinds producing heterogeneous content. This vocabulary interoperability problem should in our mind definitely be addressed seriously at the time and place of content creation, not after harvesting the content, and we hope that the FinnONTO infrastructure is a step towards facilitating this in practice. Changing the practices and adapting software in legacy to use ontologies cannot happen instantly but only in time. However, we believe there is now a road ahead to go.

## References

- d'Aquin, M., Lewen, H.: Cupboard - a place to expose your ontologies to applications and the community. In: Proceedings of the 6th European Semantic Web Conference (ESWC 2009), Springer-Verlag, 2009.
- C. Bizer, T. Heath, T. Berners-Lee. Linked Data - The Story So Far. In: Journal of Semantic Web and Information Systems, Vol. 5, No. 3, 2009.
- Paolo Bouquet, Heiko Stoermer, Claudia Niederee, and Antonio Mana. Entity Name System: The Backbone of an Open and Scalable Web of Data. In Proceedings of the IEEE International Conference on Semantic Computing, ICSC 2008, number CSS-ICSC 2008-4-28-25, pages 554-561. IEEE Computer Society, August 2008.
- Eero Hyvönen, Kim Viljanen, Jouni Tuominen and Katri Seppälä: Building a National Semantic Web Ontology and Ontology Service Infrastructure - The FinnONTO Approach. Proceedings of the European Semantic Web Conference ESWC 2008, Springer, 2008.
- Eero Hyvönen, Eetu Mäkelä, Tomi Kauppinen, Olli Alm, Jussi Kurki, Tuukka Ruotsalo, Katri Seppälä, Joeli Takala, Kimmo Puputti, Heini Kuittinen, Kim Viljanen, Jouni Tuominen, Tuomas Palonen, Matias Frosterus, Reetta Sinkkilä, Panu Paakkarinen, Joonas Laitio, Katariina Nyberg: CultureSampo - Finnish Culture on the Semantic Web 2.0. Thematic Perspectives for the End-user. Proceedings, Museums and the Web 2009, Indianapolis, USA, April 15-18, 2009.
- E. Hyvönen: Developing and Using a National Cross-Domain Semantic Web Infrastructure. In P. Sheu et al. (eds), Semantic Computing. IEEE Wiley, 2010 (forth-coming).
- J. Euzenat, P. Shvaiko. Ontology Matching. Springer-Verlag, 2007.
- M. Musen, N. Shah, N. Noy., B. Dai, M. Dorf, N. Griffith, J. Buntrock, C. Jonquet, M. Montegut, and D. Rubin: BioPortal: Ontologies and Data Resources with the Click of a Mouse. In: Proc. of the Annual Symposium proceedings/AMIA Symposium, 2008.
- Jouni Tuominen, Matias Frosterus, Kim Viljanen and Eero Hyvönen: ONKI SKOS Server for Publishing and Utilizing SKOS Vocabularies and Ontologies as Services. Proceedings of the 6th European Semantic Web Conference (ESWC 2009), Springer-Verlag, 2009a.
- Jouni Tuominen, Tomi Kauppinen, Kim Viljanen and Eero Hyvönen: Ontology-Based Query Expansion Widget for Information Retrieval. Proceedings of the 5th Workshop on Scripting and Development for the Semantic Web (SFSW 2009). CEUR Workshop Proceedings, Vol. 449, <http://ceur-ws.org/>, 2009b.
- Kim Viljanen, Jouni Tuominen and Eero Hyvönen: Ontology Libraries for Production Use: The Finnish Ontology Library Service ONKI. Proceedings of the 6th European Semantic Web Conference (ESWC 2009). Springer-Verlag, 2009.