

Deploying National Ontology Services: From ONKI to Finto

Osma Suominen¹, Sini Pessala¹, Jouni Tuominen², Mikko Lappalainen¹, Susanna Nykyri¹, Henri Ylikotila¹,
Matias Frosterus^{1,2}, Eero Hyvönen²

¹National Library of Finland

²Semantic Computing Research Group (SeCo), Aalto University, Finland

The vision: a national ontology service

In Finland, a major research initiative FinnONTO [3] was carried out in 2003–2012 with the goal of providing a national level semantic web ontology infrastructure based on centralized ontology services. Since 2008, a prototype of such a system, the ONKI Ontology Service [8,7] has been used in a living laboratory experiment with more than 400 daily human visitors and over 400 registered domains using its web services, including the ONKI mash-up widget for annotating content in legacy systems and semantic query expansion. The FinnONTO infrastructure also includes the notion of creating and maintaining a holistic Linked Open Ontology Cloud KOKO that covers different domains, is maintained in a distributed fashion by expert groups in different domains, and is provided as a national centralized service.

In 2013 the Ministry of Education and Culture and the Ministry of Finance decided to finance the deployment of ONKI and its key ontologies into a sustainable free national service Finto, created and maintained by the National Library of Finland. Finto was opened in public in January 2014, and the API services of ONKI were redirected to Finto in June. This paper summarizes, from a technical standpoint, major ideas and components underlying Finto and lessons learned during the deployment process. Issues encountered in ontology engineering regarding, e.g., concept analysis and linguistic aspects have been discussed in a separate paper [4].

Deploying ontology services

ONKI/Finto supports publication of ontologies by providing a centralized place for finding, accessing, and utilizing ontologies—the key functions of ontology libraries [1]. Using ontologies and integrating them in applications is made easier because different ontologies can be accessed via the same user interfaces and APIs.

User groups Finto ontologies are used by both humans and machines. We identified three human client groups, with slightly different needs in the user interface: full-time annotators (indexers working at, e.g., libraries and museums), users performing annotation as part of their other duties (e.g., journalists publishing articles), and ontology developers. The Finto user interface is designed to serve them all. For machine use and application developers, the service provides a variety of APIs and documentation for their usage.

Finto service utilizing Skosmos software The development of a successor system for ONKI started within the FinnONTO Project. The first step was a creation of ONKI Light [5], a lightweight prototype for an ontology browser on top of a SPARQL endpoint. The software has since evolved into a production system called Skosmos¹, a thesaurus and vocabulary browser using SKOS and SPARQL, developed at the National Library. Skosmos provides a multilingual user interface for browsing and searching the data and for visualizing concept hierarchies. The user interface has been developed by analyzing the results of repeated usability tests. The Finto service² is set up as a specific installation of Skosmos, but Skosmos can be used to provide an ontology service anywhere. Finto currently serves more than 400 human visitors per day, of which 200 are returning users who use the service regularly.

Skosmos relies on a SPARQL endpoint (Apache Jena Fuseki with the jena-text index) as its back-end and is written mainly in PHP. The main benefits of using a SPARQL endpoint is that the data provided by the service is always up to date. This allows fast update cycles in vocabulary development. Vocabularies are pre-processed using Skosify [6] to ensure that they are valid SKOS. The source code is available under the MIT license.

Machine access to concepts The ONKI system offers machine access to ontologies not only by publishing Linked Data, but also custom APIs more suited for integration to e.g. document and collection management systems. ONKI provides three main APIs: a SOAP API, a HTTP API, and a JavaScript widget [7]. These have been integrated to systems used in museums, archives and libraries. For Finto and Skosmos, a new native REST API³ providing RDF/XML, Turtle or JSON-LD serializations was developed and API wrapper code implemented to support the ONKI APIs. The ONKI system is still available for browsing ontologies, but API calls to ONKI were redirected to Finto in June 2014. The Finto API serves more than 100,000 accesses on a busy day, while the old APIs receive around 10,000 hits.

¹ <https://github.com/NatLibFi/Skosmos>

² <http://finto.fi>

³ <http://api.finto.fi>

Ontologies in ONKI and Finto At the heart of ONKI/Finto lies the General Finnish Ontology YSO. The National Library took over the development of YSO in 2013. Since then, special attention has been paid to making YSO intuitive and user-friendly without losing the benefits of machine-readable semantic data. The top-level ontology has been reworked, multilingual aspects have been refined, and work is under way to link YSO to the Library of Congress Subject Headings (LCSH). The original OWL representation was changed to SKOS, with some extensions mainly from ISO 25964⁴.

During the FinnONTO project, many YSO-based domain ontologies were created in collaboration with expert organizations. YSO is used as the central hub or glue relating all the domain ontologies to one another while minimizing the number of direct links between them. The aim here is to facilitate distributed development of the domain ontologies in the expert organizations while allowing the domain ontology developers to worry only about links and changes to one other ontology. Their content is aggregated into the single unified ontology KOKO [2].

The work on this harmonized cloud of interlinked ontologies was begun during FinnONTO and has now been continued at the National Library. The ultimate goal is to use KOKO to relate the annotations in the datasets of the various organizations facilitating interoperability and breaking down silos. KOKO has been in pilot use as an annotation vocabulary in, e.g., various museums and in the National Broadcasting Company YLE – i.e., in organizations that potentially deal with material from all possible domains.

In addition to YSO, KOKO and the YSO-based domain ontologies, a number of thesauri, classifications and other controlled vocabularies have been published in Finto, including Medical Subject Headings⁵, Iconclass⁶, and Lexvo⁷ language codes. As of July 2014, 27 vocabularies are available and more are being prepared for publishing.

Experiences during deployment

The process of transitioning users from ONKI to Finto has generally been smooth. The ONKI name has already been used for several related but distinct initiatives, so the name Finto was chosen for the new service to avoid confusion. The Finto user interface has undergone multiple rounds of usability testing and is already well liked by users, but the implementation could be further improved in terms of speed and scalability. The bottleneck is usually the SPARQL endpoint, especially for non-exact label matches where the text index cannot always be used effectively.

The transition from ONKI APIs to Finto was managed by testing the new wrapper implementations both internally and among major user organizations well in advance of their deployment. Nevertheless, there were some problems that had to be resolved after the transition, e.g., browser cache and SSL issues. Since Finto hosts newer versions of some ontologies including YSO and KOKO, users were simultaneously introduced to new versions, which caused some compatibility issues. The change from OWL to SKOS caused some API queries to break, and the new version of KOKO was also different both in terms of content and modeling. However, the transitioning problems were generally resolved in less than a week. In future, we expect similar migration issues as ontologies evolve and new versions are published, but the changes are likely to be incremental in nature.

In future, we expect some growth in the usage of the Finto service, as well as more systems integrated to use the API services. We also expect other organizations to deploy their own Skosmos instances for publishing their own vocabularies.

References

1. d'Aquin, M., Noy, N.F.: Where to publish and find ontologies? a survey of ontology libraries. *Web Semantics: Science, Services and Agents on the World Wide Web* 11, 96–111 (2012)
2. Frosterus, M., Tuominen, J., Pessala, S., Seppälä, K., Hyvönen, E.: Linked Open Ontology Cloud KOKO—managing a system of cross-domain lightweight ontologies. In: *The Semantic Web: ESWC 2013 Satellite Events*. pp. 296–297. Springer-Verlag, Berlin Heidelberg (May 26–30 2013)
3. Hyvönen, E., Viljanen, K., Tuominen, J., Seppälä, K.: Building a National Semantic Web Ontology and Ontology Service Infrastructure—The FinnONTO Approach. In: *Proc. of the European Semantic Web Conf. (ESWC 2008)*. Springer-Verlag (2008)
4. Lappalainen, M., Frosterus, M., Nykyri, S.: Reuse of library thesaurus data as ontologies for the public sector. In: *IFLA WLIC 2014 (August 16–22 2014)*
5. Suominen, O., Johansson, A., Ylikotila, H., Tuominen, J., Hyvönen, E.: Vocabulary services based on SPARQL endpoints: ONKI Light on SPARQL. In: *Poster proceedings of the 18th International Conference on Knowledge Engineering and Knowledge Management (EKAW 2012) (October 2012)*
6. Suominen, O., Mader, C.: Assessing and improving the quality of SKOS vocabularies. *J. on Data Semantics* 3(1), 47–73 (2014)
7. Tuominen, J., Frosterus, M., Viljanen, K., Hyvönen, E.: ONKI SKOS Server for Publishing and Utilizing SKOS Vocabularies and Ontologies as Services. In: *Proc. of the European Semantic Web Conf. (ESWC 2009)*. Springer-Verlag (2009)
8. Viljanen, K., Tuominen, J., Hyvönen, E.: Ontology libraries for production use: The Finnish ontology library service ONKI. In: *Proc. of the European Semantic Web Conf. (ESWC 2009)*. Springer-Verlag (2009)

⁴ <http://www.niso.org/schemas/iso25964/>

⁵ <http://www.nlm.nih.gov/mesh/>

⁶ <http://iconclass.org>

⁷ <http://lexvo.org>