

Authors:

Matias Frosterus, information systems specialist

Mikko Lappalainen, information specialist

from

the National Library of Finland

## **National Library as a Public Sector Metadata Curator - The Finto Ontology Infrastructure**

The National Library of Finland has been given a central role in ensuring the semantic interoperability of Finnish public sector metadata. Metadata about various resources ranging from traditional library material to public administration documents to scientific data is being recorded using linked lightweight ontologies and thesauri published openly in the Finto service of the National Library. The main idea is to offer a semantic network of pathways between knowledge bases that is continuously evolving based on real-world user needs. A concentrated effort in building a national ontology infrastructure for Finland began in a large-scale university research project in the early 2000s. Many annotation thesauri from various domains were modelled into OWL ontologies with the idea of making them compliant with the principles of the Semantic Web. The emphasis on concepts as opposed to terms allowed for more precise linking between the vocabularies and explicit relationships between concepts enabled simple machine-reasoning.

The OWL ontologies were later simplified into concept-based lightweight ontologies depicted in SKOS as the more advanced features offered by OWL could not be efficiently utilized in actual in-use systems. Care was however taken to retain the benefits of the concept-based modelling and thus allowing for linking and machine processability. Finally, in order to facilitate the use of these linked lightweight ontologies, a common publication platform was realized under the Finto project from 2013 onwards.

The idea of the Finto project is to establish a centralized national-level ontology infrastructure based on the results of prior research. The project - overseen by the Ministry of Finance and the Ministry of Education and Culture together with the National Library - develops both the linked ontology cloud KOKO, and the Finnish Thesaurus and Ontology Service Finto (finto.fi). KOKO collects together all the major public sector lightweight annotation ontologies and provides a network of metadata that connects knowledge from many key services such as ministry web pages, news articles of the Finnish Broadcasting Company Yle, scientific dataset repositories, and library and museum collections.

In the heart of KOKO lies the General Finnish Ontology YSO that is being developed by the experts at the National Library. YSO, a trilingual lightweight ontology of approximately 30 000 concepts, is based on the General Finnish Thesaurus YSA that was originally developed in the 1980's mainly for book indexing. YSO is built upon open linked data and semantic web principles, and the idea is that YSO can be used as a building block for next generation information systems in various public organizations from libraries to administration agencies. All domain ontologies, such as the public administration ontology JUHO, the business ontology LIITO and the Cultural Studies ontology KULO, are being developed according to principles built up by the National Library. Domain ontologies extend the core established by YSO and through YSO they also link to other domain ontologies. Strong mappings between ontologies from different domains using linked data principles allow for relatively easy interoperability between datasets from various sources.

The network of ontologies grows dynamically and in co-operation with domain experts from various organizations. The National Library coordinates this development and provides tools and guidance. Datasets connected through semantic metadata allow for new interoperability.

For example, public sector data and scientific data, annotated using concepts from linked ontologies, can be searched and shared together while ensuring that the semantics of the descriptive metadata remain consistent.

The Finnish Thesaurus and Ontology Service Finto provides a centralized publication platform for the linked ontologies and other vocabularies, with modern APIs to integrate them into various systems. Finto also allows browsing and searching of the vocabularies with various views into the hierarchy or group structures present in a vocabulary. A shared platform for publication allows for easier integration into legacy systems thus making them capable of consuming and producing linked data. The vocabularies are offered under open licenses and using Finto is free. Finto works on top of Skosmos, an open source SKOS vocabulary browser and publication platform that is being developed in The National Library. Widening the scope of services offered to new domains outside the National Library's typical focus comes with a new set of challenges. The vocabularies need to be offered in formats that are used more widely outside the library domain. Also, usability of the services needs to cater to a wider variety of user expertise. Since Finto is being developed in-house as an open source project, agility and continuous usability reviews are important in getting feedback on how to improve the service for different user groups.

Libraries have always been at the forefront of producing high-quality descriptive metadata. Finto is an example of how also other domains can benefit from the expertise of libraries in curating metadata. Building bridges between datasets allows for easier integration as well as serendipitous discoveries.