

How to Create a National Cross-domain Ontology and Linked Data Infrastructure and Use It on the Semantic Web

Eero Hyvönen

Aalto University, Department of Computer Science and
University of Helsinki, Helsinki Centre for Digital Humanities (HELDIG)
Semantic Computing Research Group (SeCo)
<https://seco.cs.aalto.fi/u/eahyvone/>

Abstract. The vision behind the Semantic Web is to build a global Web of Data (Giant Global Graph, GGG) for machines to use: based on this an interoperable and intelligent transnational WWW for humans can be created cost-efficiently. This keynote presentation for the DCMI 2001 conference addresses this grand challenge on a national level, as in practice much of the data available are often related to each other within national cultures, borders, organizations, and are represented using national languages, metadata models, vocabularies, and local conventions. This presentation overviews and discusses the vision and lessons learned in Finland on developing and deploying a cross-domain national ontology service infrastructure and Linked Open Data (LOD) publishing framework, extending the classic 5-star model to a 7-star model for better data re-usability (6. star) and quality (7. star). To test and demonstrate the infrastructure, a series of semantic portals and LOD services have been created using the “Sampo model” that has evolved gradually in 2002–2021 through lessons learned when developing and publishing the “Sampo series” of systems, including MuseumFinland (2004), HealthFinland (2009), CultureSampo (2009), BookSampo (2011), WarSampo (2015), BiographySampo (2018), NameSampo (2019), WarVictimSampo (2019), Mapping Manuscript Migrations (2020), AcademySampo (2021), as well as FindSampo, LawSampo, and ParliamentSampo underway. These systems cover a wide range of application domains and have attracted up to millions of users on the Semantic Web depending on the application, suggesting feasibility of the proposed model. This work shows a shift of focus in research on semantic portals from data aggregation and exploration systems (1. generation systems) to systems supporting research with data analytic tools (2. generation systems), and finally to automatic knowledge discovery and Artificial Intelligence (3. generation systems).

For more information see:

1. FinnONTO ontology project: <https://seco.cs.aalto.fi/projects/finnonto/>
2. Linked Data Finland project: <https://seco.cs.aalto.fi/projects/ldf/>
3. LOD Infra for Digital Humanities: <https://seco.cs.aalto.fi/projects/lodi4dh/>
4. Sampo model and portals: <https://seco.cs.aalto.fi/applications/sampo/>

Keywords: Semantic Web · Portals · Data Services · Linked Open Data