

Semantic “Sampo” Portals for Digital Humanities Based on a National Linked Open Data Infrastructure

Eero Hyvönen

University of Helsinki, Helsinki Centre for Digital Humanities (HELDIG), and
Aalto University, Semantic Computing Research Group (SeCo)
<http://heldig.fi>, <http://seco.cs.aalto.fi>

Abstract. This paper presents and overviews the “Sampo” model and series of Linked Open Data services and semantic portals, based on a 1) shared model and framework and a 2) National Semantic Web Infrastructure for Digital Humanities in Finland. Sampo portals, such as CultureSampo, BookSampo, WarSampo, BiographySampo, NameSampo, and WarVictimSampo 1914–1922, have had millions of end users on the Web in total, suggesting feasibility and sustainability of the approach developed. Several new Sampos are currently under development.

1 Sampo Model

Three generations of semantic portals can be identified during the past 20 years [9]: First, research and development focused on data harmonization, aggregation, and data exploration based on search and browsing. Second, the rise of Digital Humanities (DH) research [3] started to shift the focus to providing the user also with integrated data-analytic tools for solving research problems. Third, future portals not only provide tools for the human to solve problems but are used for finding research problems in the first place, for addressing them, and even for solving and explaining the results using techniques of Artificial Intelligence and Computational Creativity [1].

The “Sampo model”¹ [8], has been used to support development of especially 1. and 2. generation systems, but arguably also features of the 3. generation [13]. The Sampo model includes three parts that specify 1) the “business model” for creating and publishing heterogeneous, distributed Linked Data, 2) the idea of providing the end user with multiple application perspectives to the contents, and 3) how the application perspectives can be used in two basic steps using faceted search [19] for filtering out data of interest and applying data-analytics tools on it. These three ideas have been explored and developed before in different contexts. For example: The idea of collaborative content creation by data linking is a fundamental idea behind the Linked Data movement² and has been developed also in various other settings, e.g., in ResearchSpace³; Multiple perspectives to a single data service have been used in many other portals, too; The two step usage model is used in prosopographical research [21] (without the faceted search

¹ “Sampo” is according to the Finnish epic Kalevala a mythical machine giving riches and fortune to its holder, a metaphor of ancient advanced technology.

² <http://linkeddata.org/>

³ <https://www.researchspace.org/>

component). The novelty of the Sampo model comes from the idea of combining the three ideas into a whole. Furthermore, the Sampo model includes also reusable data infrastructure and software tools for implementing the portals, especially the shared ontologies and datasets of the national LODI4DH infrastructure [7], and the Sampo-UI framework for implementing user interfaces⁴.

2 Sampo Series of Semantic Portals

To develop, test, and demonstrate the model, a series of “Sampo” portals have been created and are in use, including: **CultureSampo – Finnish Culture on the Semantic Web 2.0**⁵ (online since 2009) [5,17], demonstrates how CH content of tens of different kinds can enrich each other, including a semantic model of the Kalevala epic narrative at the center. **BookSampo – Finnish Fiction Literature on the Semantic Web**⁶ (online since 2011) [16] publishes metadata about virtually all Finnish fiction literature as a knowledge graph on top of which a portal was created. BookSampo data was originally part of CultureSampo, but is today maintained as a separate portal by the Public Libraries of Finland. The portal was used by ca. 2 million users in 2019. **WarSampo – Finnish World War II on the Semantic Web**⁷ (online since 2015) [10] is a popular Finnish service that has had 630 000 users. It provides information about the ca. 100 000 casualties and significant soldiers of the WW2 in Finland and various datasets, such as 160 000 photographs from the fronts, war diaries, maps etc. **BiographySampo – Finnish Biographies on the Semantic Web**⁸ (online since 2018) [12] is yet another popular service with tens of thousands of users. It is based on mining out a large knowledge graph (over 120 million triples) from ca. 13 100 Finnish biographies of the Finnish Literature Society, authored by 1000 scholars. The data is interlinked and enriched internally and by some 16 external datasources. **NameSampo – A Linked Open Data Infrastructure and Workbench for Toponomastic Research**⁹ (online since 2019) [14] publishes data about over 2 million place names and places in Finland with old maps. It soon attracted tens of thousands of users on the Web. **WarVictimSampo**¹⁰ on Finnish wars 1914–1922 [18] (published in 2019) gathered 20 000 users in two weeks. **Mapping Manuscript Migrations** portal¹¹ [11] (published in 2020), aggregating hundreds of thousands of medieval and Renaissance manuscripts from the University of Oxford (Bodleian Library), Schoenberg Institute of the University of Pennsylvania, and the IRHT-CNRS institute in Paris is based on the Sampo model and software, too. In addition, several new Sampos are underway, such as: **FindSampo**¹² on archaeology and citizen science, **AcademySampo**¹³ on Finnish academic people 1640–1899 [15],

⁴ <https://github.com/SemanticComputing/sampo-web-app>

⁵ <https://seco.cs.aalto.fi/applications/kulttuurisampo/>

⁶ <https://seco.cs.aalto.fi/applications/kirjasampo/>

⁷ <https://seco.cs.aalto.fi/projects/sotasampo/en/>

⁸ <https://seco.cs.aalto.fi/projects/biografiasampo/en/>

⁹ <https://seco.cs.aalto.fi/projects/nimisampo/en/>

¹⁰ <https://seco.cs.aalto.fi/projects/sotasurmat-1914-1922/>

¹¹ <https://seco.cs.aalto.fi/projects/mmm/>

¹² <https://seco.cs.aalto.fi/projects/sualt/>

¹³ <https://seco.cs.aalto.fi/projects/yo-matrikkelit/>

LawSampo on Finnish legislation and case law, and **ParliamentSampo**¹⁴ on the open data of the Parliament of Finland.

3 Ontology Infrastructure with Ontology Services

Sampo portals reuse the *Linked Open Data Infrastructure for Digital Humanities in Finland (LODI4DH)*¹⁵, an initiative that continues the work of the national FinnONTO projects¹⁶ (2003–2012) on creating a shared national ontology infrastructure and services [4]. Practical results of this work include 1) shared centralized ontology services [20] that can be used in legacy systems and linked data applications via APIs, 2) the "7-star" linked data model and platform Linked Data Finland [6] for publishing linked datasets. The FinnONTO project created a LOD cloud of interlinked ontologies, based on thesauri used in Finland [4]. In 2019, the ontology services, deployed by the National Library of Finland as the national Finto.fi service, had 2 million page visits and served 32 million of API calls for annotating metadata. LODI4DH continues the work with foci on ontologies of Historical persons, Historical places and maps, Times, Events, and Keyword concepts, extending the Finnish LOD ontology cloud KOKO [2].

References

1. Boden, M.A.: Computer models of creativity. *AI Magazine* **30**(3), 23–34 (2009)
2. Frosterus, M., Tuominen, J., Pessala, S., Hyvönen, E.: Linked open ontology cloud: managing a system of interlinked cross-domain light-weight ontologies. *International Journal of Metadata, Semantics and Ontologies* **10**(3), 189–201 (2015)
3. Gardiner, E., Musto, R.G.: *The Digital Humanities: A Primer for Students and Scholars*. Cambridge University Press, New York, NY, USA (2015)
4. Hyvönen, E., Viljanen, K., Tuominen, J., Seppälä, K.: Building a national Semantic Web ontology and ontology service infrastructure – The FinnONTO approach. In: *Proceedings of the ESWC 2008, Tenerife, Spain*. pp. 95–109. Springer (2008)
5. Hyvönen, E., Mäkelä, E., Kauppinen, T., Alm, O., Kurki, J., Ruotsalo, T., Seppälä, K., Takala, J., Puputti, K., Kuittinen, H., Viljanen, K., Tuominen, J., Palonen, T., Frosterus, M., Sinkkilä, R., Paakkari, P., Laitio, J., Nyberg, K.: *CultureSampo – Finnish culture on the Semantic Web 2.0. Thematic perspectives for the end-user*. In: *Museums and the Web 2009. Archives & Museum Informatics, Toronto* (2009)
6. Hyvönen, E., Tuominen, J., Alonen, M., Mäkelä, E.: Linked Data Finland: A 7-star model and platform for publishing and re-using linked datasets. In: *The Semantic Web: ESWC 2014 Satellite Events*. pp. 226–230. Springer (May 2014)
7. Hyvönen, E.: Linked Open Data infrastructure for Digital Humanities in Finland. In: *Proc. of the Digital Humanities in the Nordic Countries (DHN 2020)*. CEUR WS Proceedings (2020), forth-coming
8. Hyvönen, E.: "Sampo" model and semantic portals for Digital Humanities on the Semantic Web. In: *Digital Humanities in Nordic Countries (DHN 2020)*, Proceedings, accepted (2020), <https://seco.cs.aalto.fi/publications/2020/hyvonen-samos-dhn-2020.pdf>

¹⁴ <https://seco.cs.aalto.fi/projects/semparl/en/>

¹⁵ <https://seco.cs.aalto.fi/projects/>

¹⁶ <https://seco.cs.aalto.fi/projects/finnonto/>

9. Hyvönen, E.: Using the Semantic Web in Digital Humanities: Shift from data publishing to data-analysis and serendipitous knowledge discovery. *Semantic Web – Interoperability, Usability, Applicability* **11**(1), 187–193 (2020)
10. Hyvönen, E., Heino, E., Leskinen, P., Ikkala, E., Koho, M., Tamper, M., Tuominen, J., Mäkelä, E.: WarSampo data service and semantic portal for publishing linked open data about the Second World War history. In: *The Semantic Web – Latest Advances and New Domains (ESWC 2016)*. pp. 758–773. Springer (2016)
11. Hyvönen, E., Ikkala, E., Tuominen, J., Koho, M., Burrows, T., Ransom, L., Wijsman, H.: A linked open data service and portal for pre-modern manuscript research. In: *DHN 2019 Digital Humanities in Nordic Countries. Proceedings of the Digital Humanities in the Nordic Countries 4th Conference*. pp. 220–229. CEUR Workshop Proceedings, Vol-2364 (2019), <http://www.ceur-ws.org/Vol-2364/>
12. Hyvönen, E., Leskinen, P., Tamper, M., Rantala, H., Ikkala, E., Tuominen, J., Keravuori, K.: BiographySampo – publishing and enriching biographies on the semantic web for digital humanities research. In: *Proceedings of the 16th Extended Semantic Web Conference (ESWC 2019)*. pp. 574–589. Springer (2019)
13. Hyvönen, E., Rantala, H.: Knowledge-based relation discovery in cultural heritage knowledge graphs. In: *Proceedings of the 4th Digital Humanities in the Nordic Countries Conference (DHN 2019)*. CEUR Workshop Proceedings, Vol-2364 (2019)
14. Ikkala, E., Tuominen, J., Raunamaa, J., Aalto, T., Ainiala, T., Uusitalo, H., Hyvönen, E.: Namesampo: A linked open data infrastructure and workbench for toponomastic research. In: *Proceedings of the 2nd ACM SIGSPATIAL Workshop on Geospatial Humanities*. pp. 2:1–2:9. *GeoHumanities’18*, ACM, New York, NY, USA (November 2018). <https://doi.org/10.1145/3282933.3282936>, <http://doi.acm.org/10.1145/3282933.3282936>
15. Leskinen, P., Hyvönen, E.: Linked open data service about historical Finnish academic people in 1640–1899. In: *Proc. of the Digital Humanities in the Nordic Countries (DHN 2020)*. CEUR WS Proceedings (2020), forth-coming
16. Mäkelä, E., Hypén, K., Hyvönen, E.: BookSampo—lessons learned in creating a semantic portal for fiction literature. In: *Proc. of ISWC-2011*, Bonn, Germany. Springer (2011)
17. Mäkelä, E., Ruotsalo, T., Hyvönen, E.: How to deal with massively heterogeneous cultural heritage data—lessons learned in CultureSampo. *Semantic Web* **3**(1), 85–109 (2012)
18. Rantala, H., Jokipii, I., Koho, M., Ikkala, E., Tuominen, J., Hyvönen, E.: Building a linked open data portal of war victims in Finland 1914–1922. In: *Proc. of the Digital Humanities in the Nordic Countries (DHN 2020)*. CEUR WS Proceedings (2020), forth-coming
19. Tunkelang, D.: *Faceted search*. Morgan & Claypool Publishers, CA, USA (2009)
20. Tuominen, J., Frosterus, M., Viljanen, K., Hyvönen, E.: ONKI SKOS server for publishing and utilizing SKOS vocabularies and ontologies as services. In: *Proceedings of the 6th European Semantic Web Conference (ESWC 2009)*. Springer (2009)
21. Verboven, K., Carlier, M., Dumolyn, J.: A short manual to the art of prosopography. In: *Prosopography approaches and applications. A handbook*, pp. 35–70. Unit for Prosopographical Research (Linacre College) (2007)