

Building and Using a National Linked Open Data Infrastructure for Digital Humanities: The Finnish Approach

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 overviews the vision, activities, and results in Finland since 2003 for building a national semantic web infrastructure for publishing and using Cultural Heritage Linked data in Digital Humanities. The infrastructure is heavily used in the "Sampo" series of Linked Open Data services and semantic portals that have had millions of end users on the Web.

1 A National Vision since 2003

Linked Open Data Infrastructure for Digital Humanities in Finland (LODI4DH) is a joint initiative of Aalto University, Semantic Computing Research Group (SeCo)¹, and University of Helsinki (UH), Helsinki Centre for Digital Humanities² (HELDIG), for creating a national data infrastructure and Linked Data services for open science. LODI4DH continues the work of the national FinnONTO initiative³ (2003–2012) on creating a shared national ontology infrastructure and services [4], and its continuation projects Linked Data Finland⁴ (2012-2014) and Linked Open Data Science Service⁵ (2015-2017) focusing of linked data publishing with seamlessly integrated Digital Humanities (DH) tooling. LODI4DH is based on a large national collaboration network of over 50 organizations in Finland who have provided the underlying data and domain expertise in the work, and also funded the projects during the past years.

Practical results of this work include 1) shared centralized ontology services that can be used in legacy systems and linked data applications via APIs, 2) a platform and "7-star" linked data model for publishing linked datasets conforming to the best practices of W3C, 3) the "Sampo" model and software framework for implementing semantic portals, and 4) a series of interlinked semantic Sampo portals for Digital Humanities. LODI4DH is under development but also in use: in 2019, the ontology services, deployed by the National Library of Finland, served 32 million of API calls for annotating metadata. Many museums in Finland, e.g., Espoo City Museum, AKSELI Consortium

¹ <http://seco.cs.aalto.fi>

² <https://heldig.fi>

³ <https://seco.cs.aalto.fi/projects/finnonto/>

⁴ <https://seco.cs.aalto.fi/projects/ldf/>

⁵ <https://seco.cs.aalto.fi/projects/lodsci/>

of eight museums, and the new national MuseumPlus cataloging system, make use of the FinnONTO ontologies and the Finto service. The Sampo portals have had millions of users on the Web. The Finnish national approach is arguably first of its kind in the world.

2 Ontology Infrastructure with Ontology Services

The FinnONTO project created a LOD cloud of interlinked ontologies, based on thesauri, gazetteers, and terminologies used in Finland [4]. For serving the ontologies, the ONKI.fi ontology service [15] was developed in a Living Lab environment, and finally part of it was deployed as the Finto.fi service by the National Library [14].

LODI4DH continues the work of the FinnONTO initiative on creating a shared ontology infrastructure. Data from collaborating organizations is aggregated into shared open domain ontologies, including:

1. **Historical persons** The work focuses on BiographySampo [8], AcademySampo [11] and related biographical and prosopographical systems.⁶
2. **Historical places and maps** The focus is on the Finnish Ontology Service of Historical Places and Maps (Hipla.fi)⁷ [6].
3. **Times** LODI4DH works with the AriadnePlus⁸ project aiming at a Finnish time period ontology aligned with PeriodO⁹.
4. **Events** The focus is on applications, such as BiographySampo, WarSampo¹⁰, and WarVictimSampo [13] on Finnish war history¹¹, and the general Finnish History Ontology HISTO.¹²
5. **Keyword concepts** Here the work on the Finnish LOD ontology cloud KOKO [2] goes on in areas such as archaeology¹³, law¹⁴, and historical professions¹⁵ [10].

3 Linked Data Finland and Sampo Portals

As for the LOD publishing platform, the “7-star” Linked Data Finland model and platform LDF.fi¹⁶ is used and developed further with additional services for DH data production, publishing, data analysis, validation, and visualization. LDF.fi extends Tim Berners-Lee’s famous 5-star model¹⁷ by two additional stars: the 6th star is given, if

⁶ <https://seco.cs.aalto.fi/projects/biographies/>

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

⁸ <https://ariadne-infrastructure.eu/>

⁹ <http://perio.do>

¹⁰ <https://seco.cs.aalto.fi/projects/sotasampo/en>

¹¹ <https://seco.cs.aalto.fi/projects/sotasurmat-1914-1922/en/>

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

¹³ <https://seco.cs.aalto.fi/projects/sualt/>

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

¹⁵ <https://seco.cs.aalto.fi/ontologies/ammo/>

¹⁶ <http://ldf.fi>

¹⁷ <https://5stardata.info/en/>

the dataset is published with the schemas it conforms to. The 7th start is given if an analysis of the quality of the data with respect to the schemas is provided, too [9].

ONKI/Finto and Linked Data Finland services have already had a wide user base demonstrating the need for the LODI4DH infrastructure. Applications based on them have also made their way from academic research into real use, especially the “Sampo” series of semantic portals [5] for Digital Humanities. For example, the system “Book-Sampo – Finnish Fiction Literature on the Semantic Web”, based originally on LDF.fi and maintained now by the Finnish public libraries (Kirjastot.fi), had 2 million visitors in 2019, and the semantic portal “WarSampo – Finnish World War II on the Semantic Web” (2015) has had 620 000 distinct users. BiographySampo, NimiSampo, and War-VictimSampo 1914–1922 have had tens of thousands of users.

In addition to the Finnish projects, there have been several joint research projects and collaborations with foreign research organizations, such as University of Oxford, Stanford University, University of Colorado Boulder, and University of Pennsylvania, and Institut de recherche et d’histoire des textes (IRHT), where the Finnish Linked Data services (LDF.fi) for DH have been used. These include WW1LOD, a data service and semantic portal based on World War I data [12], a prototype [16] hosting data about ca. 150 000 letters for the Reassembling the Republic of Letters¹⁸ initiative dealing with Early Modern correspondence data [3], and Mapping Manuscript Migrations¹⁹ data service and portal [7] for medieval pre-modern manuscripts [1].

4 Discussion: Sustainability

The end users of the LODI4DH applications have been both researchers and the public in the large, and the ontology services are used by professional catalogers in memory organizations, too. Our experiences suggest that building and maintaining a shared infrastructure is quite essential in developing applications for Digital Humanities effectively. In our own case studies, we have been able to reuse repeatedly and to develop further the FinnONTO ontologies, datasets, web services, and software components (e.g., faceted search engines, language technology tools) in novel configurations and applications. Without sustainable reuse of the infrastructure, developing, e.g., the wide range of the Sampo portals [5] would have been impossible.

References

1. Burrows, T., Hyvönen, E., Ransom, L., Wijsman, H.: Mapping manuscript migrations. Digging into data for the history and provenance of medieval and renaissance manuscripts. *Manuscript Studies* **3**(1), 249–252 (2018), <https://mss.pennpress.org/home/>
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. Hotson, H., Wallnig, T. (eds.): *Reassembling the Republic of Letters in the Digital Age*. Göttingen University Press (2019), <https://doi.org/10.17875/gup2019-1146>

¹⁸ <http://www.republicofletters.net/>

¹⁹ <http://mappingmanuscriptmigrations.org/>

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–Verlag (2008)
5. Hyvönen, E.: “Sampo” model and semantic portals for Digital Humanities on the Semantic Web. In: Proc. of the Digital Humanities in the Nordic Countries (DHN 2020). CEUR WS Proceedings (2020), forth-coming
6. Hyvönen, E., Ikkala, E., Tuominen, J.: Linked data brokering service for historical places and maps. In: Proceedings of the 1st Workshop on Humanities in the Semantic Web (WHiSe). pp. 39–52. CEUR Workshop Proceedings (May 2016), <http://ceur-ws.org/Vol-1608/>, vol 1608
7. 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/>
8. 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–Verlag (2019)
9. 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: Presutti, V., Blomqvist, E., Troney, R., Sack, H., Papadakis, I., Tordai, A. (eds.) The Semantic Web: ESWC 2014 Satellite Events. ESWC 2014. pp. 226–230. Springer-Verlag (May 2014)
10. Koho, M., Gasbarra, L., Tuominen, J., Rantala, H., Jokipii, I., Hyvönen, E.: AMMO Ontology of Finnish Historical Occupations. In: Proceedings of the The First International Workshop on Open Data and Ontologies for Cultural Heritage (ODOCH’19). vol. 2375, pp. 91–96. CEUR Workshop Proceedings (June 2019), <http://ceur-ws.org/Vol-2375/>, vol 2375
11. 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
12. Mäkelä, E., Törnroos, J., Lindquist, T., Hyvönen, E.: WW1LOD: An application of CIDOC-CRM to World War 1 linked data. *International Journal on Digital Libraries* **18**(4), 333–343 (2017). <https://doi.org/10.1007/s00799-016-0186-2>
13. 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
14. Suominen, O., Pessala, S., Tuominen, J., Lappalainen, M., Nykyri, S., Ylikotila, H., Frosterus, M., Hyvönen, E.: Deploying national ontology services: From onki to finto. In: Proceedings of the Industry Track at the International Semantic Web Conference 2014. CEUR Workshop Proceedings (2014), <http://www.ceur-ws.org/Vol-1383/>, vol 1383
15. 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–Verlag (2009)
16. Tuominen, J., Mäkelä, E., Hyvönen, E., Bosse, A., Lewis, M., Hotson, H.: Reassembling the republic of letters – a linked data approach. In: Proceedings of the Digital Humanities in the Nordic Countries 3rd Conference (DHN 2018). pp. 76–88. CEUR Workshop Proceedings (2018), <http://www.ceur-ws.org/Vol-2084/paper6.pdf>