Connecting Everything to Everything Else in a Cloud of Cultural Heritage Knowledge Graphs: SampoSampo Data Linking Service and Semantic Portal (Abstract)

Eero Hyvönen[1,2][0000−0003−1695−5840], Petri Leskinen[1][0000−0003−2327−6942], Annastiina Ahola[1][0009−0008−6369−4712], Heikki Rantala[1][0000−0002−4716−6564], and Jouni Tuominen[3,2,1][0000−0003−4789−5676]

[1] Semantic Computing Research Group (SeCo), Aalto University, Finland
[2] Helsinki Centre for Digital Humanities (HELDIG), University of Helsinki, Finland
[3] Helsinki Institute for Social Sciences and Humanities (HSSH), University of Helsinki

Enriching data by linking and reasoning is the great underlying promise of the Semantic Web and Linked Open Data (LOD) [2], as promoted by the 5th star in Tim Berners-Lee 5-star model[4]. Cultural Heritage (CH) data are typically richly interlinked not only within one application domain (say in a museum collection) but also between different fields of cultural data (in libraries, archives, and galleries). However, exposing and learning the connections (links) in CH LOD for researchers and the public to study is a challenge [3]: 1) The data are typically available in distributed data silos. 2) The data are heterogeneous based on different data models, media types, and natural languages. 3) Different identifiers for entities are used for the same entities in different data silos, which cuts off connections.

As a remedy, this paper presents a new LOD alignment service SampoSampo for a cloud of a dozen interlinked CH KGs and dozen other data services of different CH domains. Data alignment services, such as Viaf.org[5] and LOD Cloud[6], are traditionally used to enrich data in DH applications across datasets. In our case, SampoSampo has already been used for enriching data in the LetterSampo Finland system[7] [6]. This paper contributes to the state-of-the-art by arguing and demonstrating following additional use cases, based on a new semantic portal *SampoSampo – Connecting Everything to Everything Else* on top of the alignment service: 1) Searching, exploring, and analyzing entities that are shared by several datasets in a global cloud, based on 97 000 people, 55 000 organizations, and 27 000 places. 2) Finding data quality issues based on semantic disagreements between local primary data sources [7]. 3) Discovering "interesting" or even serendipitous connections between the entities with natural language ex-

[4] 5-star model: `https://5-star.info`
[5] The Virtual International Authority File: `https://viaf.org`
[6] International LOD cloud: `https://lod-cloud.net`
[7] Lettersampo Finland: `https://seco.cs.aalto.fi/projects/coco/`

planations, using relational search [5] and neuro-symbolic use of KGs, Large Language Models, and Generative AI [1].

The SampoSampo portal[8] together with the underlying LOD service[9] and data dumps in Zenodo[10] are available openly CC BY 4.0. The portal software is in Github[11] opened using the open MIT license. SampoSampo constitutes a component of the Finnish DARIAH-FI research infrastructure[12] [4]. Although SampoSampo focuses on a particular CH LOD cloud, the methods and tools presented are more general and can be applied to other similar cases.

More information and publications of the project and related works are available on the project homepage[13].

References

1. Ahola, A., Leskinen, P., Rantala, H., Tuominen, J., Hyvönen, E.: Using large language models for searching explainable relations in a cloud of cultural heritage knowledge graphs: Samposampo as a neuro-symbolic system. In: Proceedings of DHNB 2026 (October 2025), https://seco.cs.aalto.fi/publications/2025/ahola-et-al-samposampo-2025.pdf, forth-coming
2. Heath, T., Bizer, C.: Linked Data: Evolving the Web into a Global Data Space. Morgan & Claypool, Palo Alto, California (2011), http://linkeddatabook.com/editions/1.0/
3. Hyvönen, E.: Publishing and using cultural heritage linked data on the Semantic Web. Morgan & Claypool, Palo Alto, California (2012). https://-doi.org/10.2200/S00452ED1V01Y201210WBE003
4. Hyvönen, E.: How to create a national cross-domain ontology and linked data infrastructure and use it on the semantic web. Semantic Web (2024), https://doi.org/10.3233/SW-243468, dOI: 10.3233/SW-243468
5. Hyvönen, E.: Serendipitous knowledge discovery on the web of wisdom based on searching and explaining interesting relations in knowledge graphs. Journal of Web Semantics **85**, 100852 (2025). https://doi.org/10.1016/j.websem.2024.100852
6. Hyvönen, E., Leskinen, P., Poikkimäki, H., Rantala, H., Tuominen, J., Drobac, S., Koho, O., Pikkanen, I., Paloposki, H.L.: Searching, exploring, and analyzing historical letters and the underlying networks: LetterSampo Finland (1809–1917) data service and semantic portal. Digital Humanities in the Nordic and Baltic Countries Publications, post-proceedings (2026), https://seco.cs.aalto.fi/publications/2025/hyvonen-et-al-lettersampo-finland-2025.pdf, forth-coming
7. Leskinen, P., Ahola, A., Rantala, H., Tuominen, J., Hyvönen, E.: Consistency checking in a cloud of interlinked cultural heritage knowledge graphs – first results of using the samposampo data service and portal. In: Proceedings of DHNB 2026 (2026), https://seco.cs.aalto.fi/publications/2025/leskinen-et-al-samposampo-2025.pdf, forth-coming

[8] SampoSampo portal: https://samposampo.fi
[9] SampoSampo LOD service: https://ldf.fi/dataset/ss/
[10] Zenodo data dumps: https://zenodo.org/records/18188936
[11] Portal software: https://github.com/SemanticComputing/samposampo-web-app
[12] LOD part of the FIN-CLARIAH/DARIAH infrastructure: https://seco.cs.aalto.fi/projects/fin-clariah/
[13] SampoSampo project homepage: https://seco.cs.aalto.fi/projects/ss/