SampoSampo: A Portal for Studying Enriched Data and Semantic Connections on a Cultural Heritage Linked Open Data Cloud

Eero Hyvönen[1,2][0000−0003−1695−5840], Petri Leskinen[1][0009−0008−6369−4712], 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

Abstract. This paper presents novel use cases and functionalities of the new SampoSampo data service and portal on top of it, based on a Linked Open Data cloud (LOD) of related Cultural Heritage (CH) knowledge graphs of different application domains. The portal is used for searching, exploring, and analyzing entities in the LOD cloud for globally enriched data and for finding semantic "interesting" or even "serendipitous" connections (relations) between the entities with natural language explanations.

Keywords: linked data · digital humanities · entity alignment · semantic portal · data analysis · knowledge discovery

1 Connecting Everything to Everything Else

Leonardo da Vinci (1452–1519) has said: "*Learn how to see. Realize that everything connects to everything else*". [5]. This wisdom is very true regarding Cultural Heritage (CH) data due to its rich linkages. However, exposing and learning the connections (links) between resources in CH LOD is a challenge: the data are typically available in distributed data silos, the data are heterogeneous, and different identifiers are used for the same entities in different data silos, which cuts off semantic connections. As a remedy, the SampoSampo portal addresses the following challenges: 1) How to search and link data about entities in a LOD cloud for enriched descriptions about the entities? 2) How to search for connections (relations) between entities within and across KGs to find out, for example, how people are related to places or each other [17,27,10]?

The SampoSampo portal[4] and the underlying LOD service[5] will be opened using the open MIT and CC BY 4.0 licenses in 2025. More information and publications of the project are available on the project homepage[6].

[4] SampoSampo portal: `https://samposampo.ldf.fi/`
[5] SampoSampo LOD service: `https://ldf.fi/dataset/ss/`
[6] SampoSampo project homepage: `https://seco.cs.aalto.fi/projects/ss/`

Table 1: Distribution of the 83 079 person entities of SampoSampo in the Sampo LOD cloud of KGs

#	**Sampo system**[a]	**People**	**Domain**
1	AcademySampo [24,25]	23094	Finnish Academicians (1640–1899)
2	ArtSampo [3,2]	957	Finnish paintings and art
3	BiographySampo [16,28]	21540	National biography of Finland
4	BookSampo [11,4]	9097	Finnish fiction literature
5	LetterSampo Finland [29,15]: Edelfelt letters	3319	Historical letters
6	LetterSampo Finland: Snellman letters	1342	Historical letters
7	LetterSampo Finland: Åbo Academy fonds	801	Historical letters
8	Norssit Sampo [14]	578	Short biographies of a school registry
9	OperaSampo [1]	575	Opera and music theater performances
10	ParliamentSampo [18]	2101	Parliamentary speeches
11	WarSampo [13,23]	4357	Military history of Finnish WW2

[a] Links to the homepages of all Sampo LOD services and portals [9] are available at: `https://seco.cs.aalto.fi/applications/sampo/`

2 SampoSampo Data Service and Portal

Use cases There are two main use cases for the SampoSampo system. Firstly, it is often useful to search data about entities from different data sources, because different datasets may contain complementary enriching information about the entities. For example, BiographySampo [16,28] publishes general biographical data about over 13 000 prominent Finns but there are more data available about them in other Sampos. For example, ParliamentSampo [18] publishes the speeches and activities of Parliament of Finland for 790 politicians present in BiographySampo. Aggregated data also facilitate the detection of mutual inconsistencies in the data sources. Secondly, searching for explained "interesting" implicit relations between entities within and between datasets is useful. For example: how are people related to places, say Finnish artists to France, or composer Jean Sibelius to Berlin?

SampoSampo Data Service aligns and aggregates data about 83 079 person entities from the Sampo systems [9] listed in Table 1 and several other data sources on the Web (number of aligned people in brackets): Geneanet[7] [22 407],

[7] `https://fi.geneanet.org/`

Geni.com [5980], ISNI.org [31 012], Kanto[8] [31 580], ULAN[9] [1732], VIAF.org [17 964], Wikidata.org [38 010], and Wikitree.com [5761]. In additon to people, the aligned resources include 52 161 organizations and 20 676 place entities.

More details about the data model used and how it was populated can be found in [12] and on the project homepage. The aligned LOD was published on the Linked Data Finland platform[10] [20,19] offering various LOD services, such as a SPARQL endpoint that was used to create the SampoSampo portal.

SampoSampo Portal is based on the Sampo model [9] and the Sampo-UI framework [21,26] for UI design. The landing page of the portal provides access to application perspectives where the instances of KG classes (one for each perspective) can be searched using semantic faceted search whose facets correspond to the properties of the classes. After filtering results by making selections on the facets, the result set can be displayed as a table or by using a variety of data analysis tools and visualizations, such as charts, maps, and timelines. By selecting an instance from the result set, aggregated linked data related to it can be displayed, and data-analyses and visualizations pertaining to the entity instance can be shown. The demo implementation contains perspectives for searching People, Organizations, Places, Historical events, Connections between people and places, Connections between people, and Connections between entities available in the Finnish Wikipedia and Wikidata.

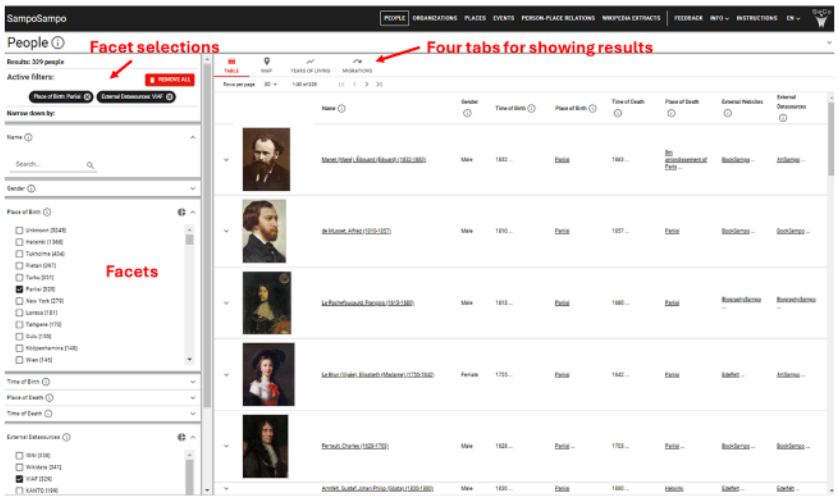

Fig. 1: Application perspective for searching and visualizing people

[8] https://finto.fi/finaf/en/
[9] https://www.getty.edu/research/tools/vocabularies/ulan/
[10] Linked Data Finland platform: https://ldf.fi

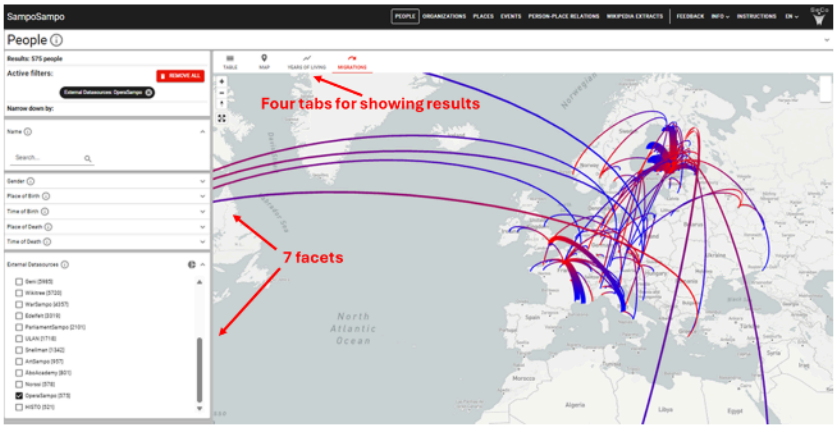

Fig. 2: Visualizing the life lines of people in OperaSampo from their place of birth (blue end of the line) to the place of death (red end)

For example, people can be searched for by the following facets illustrated in Figure 1 on the left: Name (string search facet); Gender; Place of Birth; Time of Birth; Place of Death; Time of Death; External Data Sources (where the entity appears). The user has selected Place of Birth = Paris and VIAF as the data source, and therefore people born in Paris and present in VIAF.org service are listed as a result of 329 individuals on the right.

The search result can be analyzed using the following tabs: TABLE lists the results with images and links (selected in the figure); MAP shows how people are related to places, based on events they have participated in different roles; YEARS OF LIVING illustrates life times of people on a time line; MIGRATIONS show how people have moved for their place of birth to their place of death. Figure 2 depicts the MIGRATIONS tab. In this case, OperaSampo was selected from the data source facet, and a result set of 575 people there that are linked with at least some other data sources are shown on the right. The user has selected from the tabs the MIGRATIONS tab that shows the life lines of the people on a map. One can learn from the visualization, for example, that quite a few opera people have moved to Paris and died there.

Figure 3 illustrates how a perspective for discovering and searching implicit semantic connections is used. Here data from our earlier relational search systems [17,27] was re-used to search semantic connections. The perspective is based on the class of connections whose instances represent connections between people and places with properties for them as well as an explanation about the connection. The facets on the left are: Person, Occupation or title of the person, Place, and Type of the connection, based on the events through which the connections have been established. Ten connection types have been used, such as getting an honorary price in a place, creating a painting depicting a place, or writing a book about it. The user has selected from the Occupation facet category "Au-

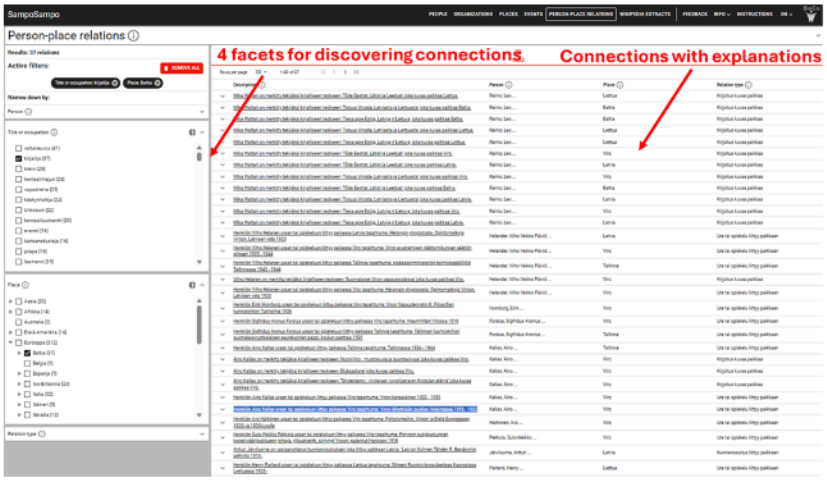

Fig. 3: Findig connections with explanations between people and places

thor" and "Baltic countries" from the Place facet. On the right, 37 connections between people and places with their explanations are listed, such as "Relating to the studies or career of the person Aino Kallas, she was the spouse of the ambassador of Estonia in Finland in 1918 - 1922" (In Finnish: "Henkilön Aino Kallas uraan tai opiskeluun liittyy paikassa Viro tapahtuma: Viron lähettilään puoliso Helsingissä 1918 - 1922".)

3 Related Works and Contributions

SampoSampo portal is based on a new entity alignment LOD service. Our work was inspired by related systems, such as the Linked Open Data Cloud[11], VIAF[12] [8], works of ontology mapping, ontology services [30,6], Linked Open Vocabularies[13], and the proxy data model of Europeana [22]. Previously, entity alignment services have been used as *tools* for aggregating and enriching data in new applications. A major contribution of SampoSampo is to present the novel idea and approach of using an entity alignment service *itself*, based on a LOD cloud and related datasets, for developing applications to find and study enriched global entity data and their implicit semantic connections. Such an application can be used for research in Digital Humanities [7].

Acknowledgments This research is part of the Finnish FIN-CLARIAH research infrastructure project, funded by the Research Council of Finland and the European Union NextGenerationEU instrument. Computing services provided by the CSC – IT Center for Science were used.

[11] Linked Open Data Cloud: `https://lod-cloud.net/`
[12] Virtual International Authority File system: `https://viaf.org`
[13] Linked Open Vocabularies: `https://lov.linkeddata.es/dataset/lov/`

References

1. Ahola, A., Hyvönen, E., Rantala, H., Kauppala, A.: Historical opera and music theatre performances on the semantic web: Operasampo 1830-1960. In: Knowledge Graphs in the Age of Language Models and Neuro-Symbolic AI. Proceedings of the 20th International Conference on Semantic Systems, 17-19 September 2024, Amsterdam, The Netherlands. pp. 386–402. IOS Press (2024). https://-doi.org/10.3233/SSW240031
2. Ahola, A., Peura, L., Leal, R., Rantala, H., Hyvönen, E.: Using generative AI and LLMs to enrich art collection metadata for searching, browsing, and studying art history in digital humanities. In: Proceedings, 2nd International Conference on Data & Digital Humanities Generative Artificial Intelligence for Text and Multimodal Data 12th - 13th December 2024, University of Minho, Braga, Portugal (November 2024), `https://seco.cs.aalto.fi/publications/2024/ahola-et-al-genai-2024.pdf`, accepted, forth-coming
3. Ahola, A., Rantala, H., Hyvönen, E.: ArtSampo – Finnish art on the Semantic Web. In: The Semantic Web: ESWC 2024 Satellite Events, Hersonissos, Crete, Greece, May 26 - 30, 2024, Proceedings. Springer (May 2024). https://doi.org/10.1007/978-3-031-78952-6_18
4. Annastiina Ahola, Telma Peura, E.H.: Using linked data for data analytic literary research: Case BookSampo – Finnis fiction literature on the Semantic Web (2025). https://doi.org/10.1002/asi.24984
5. Askins, L.: Learn how to see. Realize that everything connects to everything else. -Leorardo da Vinci. Independent publication (2021), `https://www.amazon.com/Realize-everything-connects-Leonardo-Notebook/dp/B09L9Y5CP3`
6. 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). https://doi.org/10.1504/IJMSO.2015.073879
7. Gardiner, E., Musto, R.G.: The Digital Humanities: A Primer for Students and Scholars. Cambridge University Press, New York, NY, USA (2015), `https://doi.org/10.1017/CBO9781139003865`
8. Hickey, T.B., Toves, J.A.: Managing ambiguity in VIAF. DLib Magazine **20**(7/8) (2014). https://doi.org/doi:10.1045/july2014-hickey
9. Hyvönen, E.: Digital humanities on the Semantic Web: Sampo model and portal series. Semantic Web **14**(4), 729–744 (2022). https://doi.org/10.3233/SW-223034
10. 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** (2025). https://doi.org/10.1016/j.websem.2024.100852
11. Hyvönen, E., Ahola, A., Ikkala, E.: BookSampo fiction literature knowledge graph revisited: Building a faceted search interface with seamlessly integrated data-analytic tools. In: Theory and Practice of Digital Libraries (TPDL 2022), Accelerating Innovations Track, Padova, Italy. pp. 506—511. Springer–Verlag (2022), `https://doi.org/10.1007/978-3-031-16802-4_54`
12. Hyvönen, E., Ahola, A., Leskinen, P., Rantala, H., Tuominen, J.: How to create a portal for digital humanities research using a linked open data cloud of cultural heritage knowledge graphs: Case SampoSampo. In: Proceedings: SemDH 2025 Second International Workshop of Semantic Digital Humanities, co-located with ESWC 2025, Portoroz, Slovenia. CEUR Workshop Proceedings (2025), `https://seco.cs.aalto.fi/publications/2025/hyvonen-et-al-samposampo-semdh-2025.pdf`, forth-coming

13. 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: Sack, H., Blomqvist, E., d'Aquin, M., Ghidini, C., Ponzetto, S.P., Lange, C. (eds.) The Semantic Web – Latest Advances and New Domains (ESWC 2016). pp. 758–773. Springer–Verlag (May 2016), https://doi.org/10.1007/978-3-319-34129-3_46
14. Hyvönen, E., Leskinen, P., Heino, E., Tuominen, J., Sirola, L.: Reassembling and enriching the life stories in printed biographical registers: Norssi High School Alumni on the semantic web. In: Proceedings, Language, Data and Knowledge (LDK 2017). pp. 113–119. Springer–Verlag (June 2017), https://link.springer.com/chapter/10.1007/978-3-319-59888-8_9
15. Hyvönen, E., Leskinen, P., Poikkimäki, H., Rantala, H., Tuominen, J., Drobac, S., Koho, O., Pikkanen, I., Paloposki, H.L.: LetterSampo Finland (1809–1917) data service and portal: Searching, exploring, and analyzing historical letters and their underlying networks. In: Proceedings of ESWC 2025, supplement, poster and demo papers. pp. Accepted, forthcoming. Springer-Verlag (2025), https://seco.cs.aalto.fi/publications/2025/hyvonen-et-al-lettersampo-finland-poster-2025.pdf
16. 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: The Semantic Web. 16th International Conference, ESWC 2019. pp. 574–589. Springer–Verlag (2019). https://-doi.org/10.1007/978-3-030-21348-0_37
17. Hyvönen, E., Rantala, H.: Knowledge-based relational search in cultural heritage linked data. Digital Scholarship in the Humanities (DSH) **36**, 155–164 (2021). https://doi.org/https://doi.org/10.1093/llc/fqab042
18. Hyvönen, E., Sinikallio, L., Leskinen, P., Drobac, S., Leal, R., Mela, M.L., Tuominen, J., Poikkimäki, H., Rantala, H.: Publishing and using parliamentary linked data on the semantic web: ParliamentSampo system for Parliament of Finland. Semantic Web pp. Pre–Press, 1–23 (2024). https://doi.org/10.3233/SW-243683
19. Hyvönen, E., Tuominen, J.: 8-star linked open data model: Extending the 5-star model for better reuse, quality, and trust of data. In: Posters, Demos, Workshops, and Tutorials of the 20th International Conference on Semantic Systems (SEMANTiCS 2024). vol. 3759. CEUR Workshop Proceedings (September 2024), https://ceur-ws.org/Vol-3759/paper4.pdf
20. 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–Verlag (May 2014). https://doi.org/10.1007/978-3-319-11955-7_24
21. Ikkala, E., Hyvönen, E., Rantala, H., Koho, M.: Sampo-UI: A full stack JavaScript framework for developing semantic portal user interfaces. Semantic Web **13**(1), 69–84 (2022). https://doi.org/10.3233/SW-210428
22. Isaac, A.: Europeana data model primer. Tech. rep., Europeana (2023), https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Primer_130714.pdf
23. Koho, M., Ikkala, E., Leskinen, P., Tamper, M., Tuominen, J., Hyvönen, E.: WarSampo knowledge graph: Finland in the Second World War as linked open data. Semantic Web **12**(2), 265–278 (January 2021). https://doi.org/10.3233/SW-200392, https://doi.org/10.3233/SW-200392

24. Leskinen, P., Hyvönen, E.: Biographical and prosopographical analyses of Finnish academic people 1640–1899 based on linked open data. In: Proceedings of the Biographical Data in a Digital World 2022 (BD 2022), Tokyo. Institute of Cultural History, ZRC SAZU, Ljubljana, Slovenia (2024). https://doi.org/10.3986/9789610508120$_7$
25. Leskinen, P., Hyvönen, E.: Biographical and prosopographical analyses of finnish academic people 1640–1899 based on linked open data. In: Proceedings of the Biographical Data in a Digital World 2022 (BD 2022), Tokyo. Institute of Cultural History, ZRC SAZU, Ljubljana, Slovenia (January 2024), https://doi.org/10.3986/9789610508120_7
26. Rantala, H., Ahola, A., Ikkala, E., Hyvönen, E.: How to create easily a data analytic semantic portal on top of a SPARQL endpoint: introducing the configurable Sampo-UI framework. In: VOILA! 2023 Visualization and Interaction for Ontologies, Linked Data and Knowledge Graphs 2023. vol. 3508. CEUR Workshop Proceedinbgs (2023), https://ceur-ws.org/Vol-3508/paper3.pdf
27. Rantala, H., Hyvönen, E., Leskinen, P.: Finding and explaining relations in a biographical knowledge graph based on life events: Case BiographySampo. In: Joint Proceedings of the ESWC 2023 Workshops and Tutorials co-located with 20th European Semantic Web Conference (ESWC 2023). vol. 3443. CEUR Workshop Proceedings (2023), https://ceur-ws.org/Vol-3443/ESWC_2023_SEMMES_relations.pdf
28. Tamper, M., Leskinen, P., Hyvönen, E., Valjus, R., Keravuori, K.: Analyzing biography collections historiographically as linked data: Case National Biography of Finland. Semantic Web **14**(2), 385–419 (2023), https://doi.org/10.3233/SW-222887
29. Tuominen, J., Koho, M., Pikkanen, I., Drobac, S., Enqvist, J., Hyvönen, E., La Mela, M., Leskinen, P., Paloposki, H.L., Rantala, H.: Constellations of Correspondence: a linked data service and portal for studying large and small networks of epistolary exchange in the Grand Duchy of Finland. In: DHNB 2022 The 6th Digital Humanities in Nordic and Baltic Countries Conference. pp. 415–423. CEUR Workshop Proceedings, Vol. 3232 (March 2022), http://ceur-ws.org/Vol-3232/paper41.pdf
30. Xia, W., Jiménez-Ruiz, E., Cross, V.V.: Using BioPortal as a repository for mediating ontologies in ontology alignment. In: Workshop on Semantic Web Applications and Tools for Life Sciences (2015), https://api.semanticscholar.org/CorpusID:37359417