# Using the AcademySampo Portal and Data Service for Biographical and Prosopographical Research in Digital Humanities

Petri Leskinen[1][0000−0003−2327−6942] and Eero Hyvönen[1,2][0000−0003−1695−5840]

[1] Semantic Computing Research Group (SeCo), Aalto University, Finland
[2] HELDIG – Helsinki Centre for Digital Humanities, University of Helsinki, Finland
http://seco.cs.aalto.fi, http://heldig.fi, firstname.lastname@aalto.fi

**Abstract.** This paper presents the in-use AcademySampo portal and Linked Open Data (LOD) service for biographical and prosopographical research, a new member in the Sampo series of cultural heritage applications for Digital Humanities. The portal is based on a dataset of short textual biographies about all 28 000 Finnish and Swedish academic people educated in 1640–1899 in Finland. Linked data extracted from the biography entries was enriched by internal and external data linking, and by reasoning, e.g., genealogical networks of the people mentioned. The data was published as a LOD service. This paper demonstrates how to use the AcademySampo data in Digital Humanities research by faceted search integrated seamlessly with data analytic tools of the AcademySampo portal, as well as by using the LOD service directly via a SPARQL editor and by Python scripting using Google Colab and Jupyter notebooks.

**Keywords:** Biography · Prosopography · Linked Data · Digital Humanities
**Demo paper**

## 1 Introduction

Biographical research is "concerned with the reconstruction of life histories and the constitution of meaning based on biographical narratives and documents"[3] regarding individual persons, while "prosopography is an investigation of the common characteristics of a group of people, whose individual biographies may be largely untraceable"[4] [8]. This demo paper concerns application on Semantic Web technologies and Linked Data in biographical and prosopographical research. We demonstrate how to use the new AcademySampo LOD service and semantic

---

[3] https://en.wikipedia.org/wiki/Biographical_research; accessed Aug 3, 2021
[4] https://en.wikipedia.org/wiki/Prosopography; accessed Aug 3, 2021

portal [5,4,6] for these research purposes, based on the Finnish registries "Ylioppi-lasmatrikkeli" 1640–1899[5] that contain short biographical descriptions of 28 000 students of the University of Helsinki, formerly the Royal Academy of Turku. These registries cover a significant part of the history of Finland and the Finnish university institution, since the University of Helsinki was the only university in Finland during the time frame in focus.

## 2 Using AcademySampo Portal

The AcademySampo portal[6], based on the Sampo model[7], is used by first filtering results by faceted search, and then applying data analytic tools to the result set. The user can search for people, places, occupations, and student's clubs in separate application perspectives. Data analytic tools are available for both 1) individuals (biography) and 2) groups of them (prosopography). For example, Fig. 1 illustrates the family relations of the poet Johan Ludvig Runeberg (1804–1877) extracted and reasoned from the mentions of him in the biographies [6].



**Fig. 1.** Family relations of J. L. Runeberg (1804–1877) visualized in AcademySampo

In prosopographical research, the user can first filter out a group of people and then analyse them. For example, the students of the Swedish Småland student

---

club in a certain time period can be selected and their migrations visualized from the place of birth to the place of death on a map [4].

## 3   Using the SPARQL Endpoint for Data Analysis

Alternatively, the Linked Open Data service of AcademySampo[8] can be accessed directly for customized analyses. For example, the YASGUI[9] interface for SPARQL querying and visualizing the results can be used, or Python scripting with notebooks in Google Colab[10] and Jupyter[11].

For instance, Fig. 2 illustrates the distribution of the most common vocational groups of people in the data during different time periods. This chart shows how in the 17th and early 18th century the religious vocations have been most dominant. However, during the three centuries the proportion of religious occupations has decreased from over 50 to mere 15 per cent. Respectively, the fields of public administration, and education, have had an increasing growth during the observed time period. This analysis and visualization was created using Google Colab.
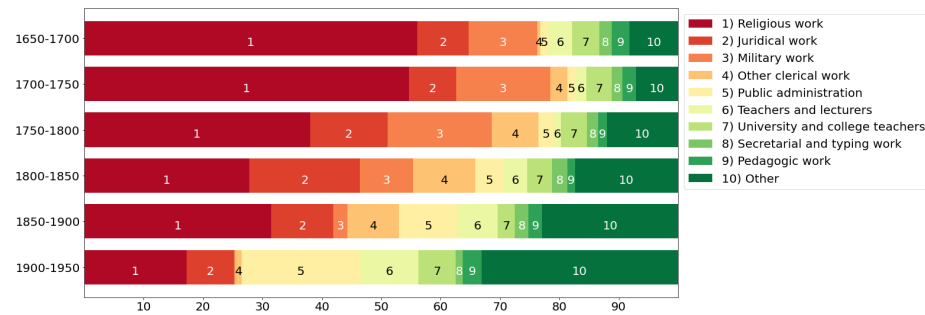


**Fig. 2.** Most common vocational groups visualized in Google Colabs

## 4   Discussion

**Related Works** Analyzing biographical data has grown into a new research and application field, reported, e.g., in the Biographical Data in Digital World workshops BD2015 [2], BD2017 [3], and BD2019. Data analyses related to those of this paper have been made for dictionaries of biography of U.K. [9], Ireland [1], and Finland [7], but not for the new AcademySampo data.

---

[8] The AcademySampo LOD service is available at https://www.ldf.fi/dataset/yoma.

[9] https://yasgui.triply.cc

[10] https://colab.research.google.com/notebooks/intro.ipynb

[11] https://jupyter.org

**Contributions** Developing AcademySampo demonstrates, how textual biographies can be transformed into linked data and be enriched with related datasets as well as by reasoning new relations in the data [6]. The resulting LOD, published on the Linked Data Finland platform[12] can be used by users without programming skills via the ready-to-use tools integrated seamlessly with faceted search and exploration in the portal. Alternatively, the SPARQL endpoint can be uses flexibly for versatile data-analyses and visualizations with little knowledge about SPARQL and programming.

# References

1. Bhreathnach, Ú., Burke, C., Fhinn, J.M., Cleircín, G.Ó., Raghallaigh, B.Ó.: A quantitative analysis of biographical data from Ainm, the Irish-language biographical database (2019), http://doras.dcu.ie/23774/1/Ainm%20BD%20FINAL.docx.pdf, presented at the 3rd Conference on Biographical Data in a Digital World (BD 2019).
2. ter Braake, S., Anstke Fokkens, R.S., Declerck, T., Wandl-Vogt, E. (eds.): BD2015, Biographical Data in a Digital World 2015. CEUR Workshop Proceedings, Vol-1399 (2015), http://ceur-ws.org/Vol-1272/.
3. Fokkens, A., ter Braake, S., Sluijter, R., Arthur, P., Wandl-Vogt, E. (eds.): BD2017 Biographical Data in a Digital World 2015. CEUR Workshop Proceedings, Vol-1399 (2017), http://ceur-ws.org/Vol-2119/.
4. Hyvönen, E., Leskinen, P., Rantala, H., Ikkala, E., Tuominen, J.: Akatemiasampo-portaali ja -datapalvelu henkilöiden ja henkilöryhmien historialliseen tutkimukseen (AcademySampo portal and data service for biographical and prosopographical research). Informaatiotutkimus 40(2), 28–56 (2021), https://journal.fi/inf/article/view/102656.
5. Leskinen, P., Hyvönen, E.: Linked open data service about historical Finnish academic people in 1640–1899. In: DHN 2020 Digital Humanities in the Nordic Countries. Proceedings of the Digital Humanities in the Nordic Countries 5th Conference. pp. 284–292. CEUR Workshop Proceedings, Vol. 2612 (2020), http://ceur-ws.org/Vol-2612/short14.pdf.
6. Leskinen, P., Hyvönen, E.: Reconciling and using historical person registers as linked open data in the AcademySampo knowledge graph. In: Proceedings of the 20th International Semantic Web Conference (ISWC 2021). Springer (2021), https://seco.cs.aalto.fi/publications/2021/leskinen-hyvonen-reconciling-2021.pdf, in press.
7. Tamper, M., Leskinen, P., Hyvönen, E., Valjus, R., Keravuori, K.: Analyzing biography collection historiographically as linked data: Case national biography of Finland (2021), https://seco.cs.aalto.fi/publications/2021/tamper-et-al-bs-2021.pdf, submitted.

---

[12] https://ldf.fi

[13] https://intavia.eu/

[14] https://nexuslinguarum.eu/the-action

8. 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). https://doi.org/1854/8212.
9. Warren, C.: Historiography's two voices: Data infrastructure and history at scale in the oxford dictionary of national biography (ODNB). Journal of Cultural Analytics (2018). https://doi.org/DOI:10.22148/16.028.