

WarVictimSampo 1914–1922: A Semantic Portal and Linked Data Service for Digital Humanities Research on War History

Heikki Rantala¹, Esko Ikkala¹, Ilkka Jokipii^{2,3}, Mikko Koho¹, Jouni Tuominen^{1,2}, and Eero Hyvönen^{1,2}

¹ Aalto University, Semantic Computing Research Group (SeCo), Finland

² University of Helsinki, Helsinki Centre for Digital Humanities (HELDIG), Finland

³ The National Archives of Finland

Abstract. This paper presents the semantic portal and Linked Open Data (LOD) service WARVICTIMSAMPO 1914–22 about the war victims, battles, and prisoner camps in the Finnish Civil and other wars. The system is based on a database of the National Archives of Finland and related data compiled during the project. The system contains detailed information about some 40 000 deaths extracted from several data sources, and data about prisoner camps and over 1000 battles of the Civil War. A key novelty of WARVICTIMSAMPO 1914–22 is the integration of ready-to-use Digital Humanities tooling with the data service, which allows, e.g., studying information about wider prosopographical groups in addition to individual victims. We demonstrate how the tools of the portal, as well as the underlying SPARQL endpoint, can be used to explore and analyze war history in flexible and visual ways. WARVICTIMSAMPO 1914–22 is a new member in the series of "Sampo" model based semantic portals. It was published in late 2019 and got 20 000 users in two weeks.

Keywords: Linked Data, Semantic Web, War History

1 Introduction

This paper presents the semantic portal and Linked Open Data (LOD) service WARVICTIMSAMPO 1914–22 about the war victims, battles, and prisoner camps in the Finnish Civil and other wars. The tools offered by the service help researchers and general public to better access the historical data. The main focus of the service is the Finnish War Victims 1914–1922 database that includes some 40 000 victims and is maintained by the National Archives of Finland. Most, over 90 percent, of the deaths recorded here are related to the Finnish Civil War in 1918, and the rest are related to the other wars of the period. Fig. 1 depicts the distribution of death dates in the data during 1918 as shown by the service. The data includes people who have died in Finland and abroad.

In this paper we show how LOD and modern web technologies can be used to enhance and update an old data service. The paper also demonstrates how modern tools can be used with LOD to analyze the data for Digital Humanities

research. The original War Victims data was recorded in 1999–2003 as a government project [10] and includes 39 931 deaths. The original data was converted into LD form and updated with 1590 new previously unknown victims and some new information concerning the old records [8]. AN important contribution is making the access to the data easier.

The old database is not directly open for public access. There is an old web application⁴ in use for exploring the data with simple search functionality and a homepage for each person. The person’s homepage includes basic information about the victim, but many pieces of information are not shown, even though they would be available in the underlying database. The end users of the system have deemed the search interface fairly inflexible with too few options to choose from. Also some means of exporting the data from the database was asked for.

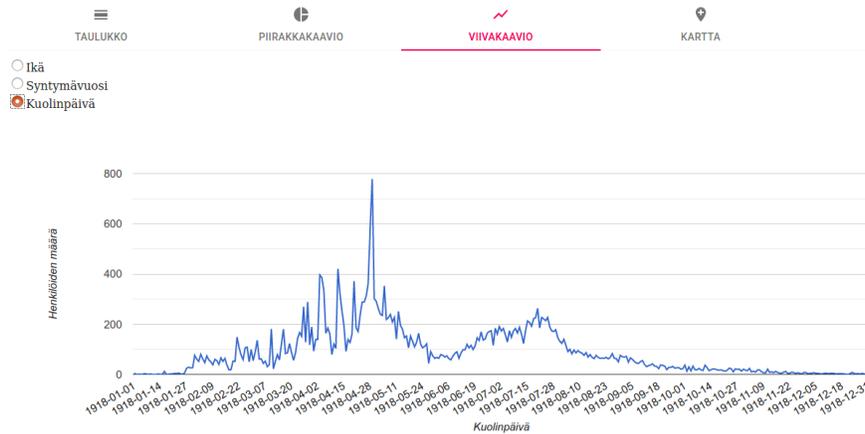


Fig. 1. Distribution of death dates during 1918 as shown in WARVICTIMSAMPO 1914–22

The death records contain basic information of the people (e.g., name, place of birth, date of birth, date of death), socioeconomic information (e.g., occupation, marital status), and war related information (e.g., military rank, military organization, time of imprisonment). To publish the data we use the “Sampo” publishing model⁵ [4]. The data in LOD form is loaded into a SPARQL endpoint hosted at the Linked Data Finland platform⁶ [3], where it can be queried using SPARQL. The semantic portal makes queries to this publicly open endpoint, and a researcher can also query the endpoint for her own purpose.

⁴ <http://vesta.narc.fi/cgi-bin/db2www/sotasurmaetusivu/main>

⁵ <https://www.europenowjournal.org/2019/09/09/linked-data-in-use-sampo-portals-on-the-semantic-web/>

⁶ <http://www.ldf.fi/dataset/viso>

2 WarVictimSampo 1914–22 Semantic Portal

A semantic portal was developed to allow different user groups to access the data easily. Potential user groups include researchers, students, and the wider public interested in either the Finnish Civil War in general or the fates of their relatives. Even though the data can be accessed by anyone with SPARQL queries, that can be too technically demanding for many user groups. On the other hand, even a researcher who is able to create her own SPARQL queries may find it useful to have an easy way to explore the data and to create simple visualizations quickly. Visualization are hopefully useful for both finding new data and educating the public about history. These tools should not be expected to fully replace manual research and close reading. They should only be used to spot interesting things that require more detailed analysis.

The user interface of the semantic portal is implemented as a full stack JavaScript web application, using the Sampo-UI framework⁷. The user interface is built around the concept of faceted search [9]. With faceted search, the user can easily narrow the search step by step by making selections based on predetermined orthogonal hierarchies of property values called facets. Facets also show the number of available items with each possible selection. This allows the user to immediately see the number of solutions of each possible selection. Combined with selections on other facets like occupation, party, and age, the user may also draw interesting conclusions by observing the hit distributions on the facets. Thus faceted search can be used to find individuals, such as relatives, that fit certain criteria, but it can also be used to find information about the distributions of different kind of the casualties. This kind of search paradigm is an example of exploratory search [6].

The user interface currently includes two main perspectives for exploring the underlying knowledge graph: 1) The main perspective is based on searching and exploring the casualties. 2) There is also a perspective based on the battles of the Finnish Civil War, covering currently 1182 geo-coded battles. Other views may be added later in the same way as in other “Sampo” series semantic portals⁸.

For the both perspectives there are multiple tabs to view the data in different ways. Currently the data can be shown as table or downloaded as CSV file, and it can be visualized with pie charts, line charts and maps. All the visualizations are dynamic, reflecting the selections made with the facets. For example Fig. 2 shows how people in the data from the Vaasa province clearly have died mainly in the area of Tampere for some reason that can be explored further.

Both the line chart and pie chart visualizations can be used for multiple different properties. For example the line chart can be used to visualize age distributions, birth years, or for death dates as shown in Fig. 1.

The WARVICTIMSAMPO 1914–22 semantic portal and data service were opened to public on 20 November 2019 and has had tens of thousands users.

⁷ <https://github.com/SemanticComputing/sampo-ui>

⁸ <https://www.europenowjournal.org/2019/09/09/linked-data-in-use-sampo-portals-on-the-semantic-web/>

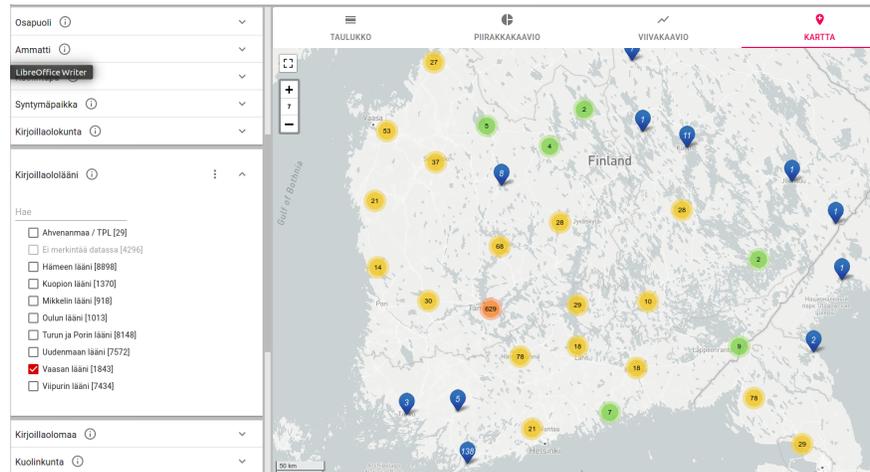


Fig. 2. A map visualization of death places for the people registered in Vaasa province.

3 Related Work

WARVICTIMSAMPO 1914–22 is a follow up project of WarSampo [5], which uses LD to present and publish information related to the Second World War in Finland, including death records. The novelty of WARVICTIMSAMPO 1914–22 lays in the idea of developing new data-analytic tooling for research in war history, as well as in creating, cleaning, extending, and publishing the former War Victims 1914–22 database for open use on the the Semantic Web. Despite both being wars, the Finnish Civil War and the Finnish involvement in the Second World War are quite different phenomena. The data related to the Civil War can, for example, potentially be used to analyze the underlying reasons for the war, which may not make sense in the context of the Second World War. There is also much more uncertainty and potential for interpretation due to the nature of the event.

There have been several projects publishing linked data about the World War I on the web, such as Europeana Collections 1914–1918⁹, 1914–1918 Online¹⁰, WW1 Discovery¹¹, Out of the Trenches¹², CENDARI¹³, Muninn¹⁴, and WW1LOD [7]. In addition to WarSampo, there are a few works that use the

⁹ <http://www.europeana-collections-1914-1918.eu>

¹⁰ <http://www.1914-1918-online.net>

¹¹ <http://ww1.discovery.ac.uk>

¹² <http://www.canadiana.ca/en/pcdhn-lod/>

¹³ <http://www.cendari.eu/research/first-world-war-studies/>

¹⁴ <http://blog.muninn-project.org>

Linked Data approach to WW2, such as [2, 1], Open Memory Project¹⁵ on holocaust victims, and the Dutch project Netwerk Orlogsbronnen¹⁶.

Acknowledgements Thanks to Päivi Happonen, Vili Haukkovaara, Markku Mäenpää, and Jarmo Nieminen for fruitful discussions and collaboration in the project. Our research was funded by the Prime Minister’s Office. Thanks to CSC – IT Center for Science, Finland, for computational resources.

References

1. de Boer, V., van Doornik, J., Buitinck, L., Marx, M., Veken, T.: Linking the kingdom: enriched access to a historiographical text. In: Proc. of the 7th International Conference on Knowledge Capture (KCAP 2013). pp. 17–24. ACM (2013)
2. Collins, T., Mulholland, P., Zdrahal, Z.: Semantic browsing of digital collections. In: Proceedings of the 4th International Semantic Web Conference (ISWC 2005). pp. 127–141. Springer–Verlag (November 2005)
3. 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)
4. 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
5. 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–Verlag (2016)
6. Marchionini, G.: Exploratory search: from finding to understanding. *Communications of the ACM* **49**(4), 41–46 (2006)
7. 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 (nov 2017). <https://doi.org/10.1007/s00799-016-0186-2>
8. 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: Proceedings of Digital Humanities in Nordic Countries (DHN 2020), Riga. CEUR Workshop Proceedings (March 2020), accepted
9. Tunkelang, D.: Faceted search. *Synthesis Lectures on Information Concepts, Retrieval, and Services*, Morgan & Claypool, Palo Alto, California (2009)
10. Westerlund, L. (ed.): Sotaoloissa vuosina 1914–22 surmansa saaneet. Tilastoraportti. Valtioneuvoston kanslia (2004)

¹⁵ http://www.bygle.net/wp-content/uploads/2015/04/Open-Memory-Project_3-1.pdf

¹⁶ <https://www.oorlogsbronnen.nl>