

Prosopographical Views to Finnish WW2 Casualties Through Cemeteries and Linked Open Data

Esko Ikkala¹, Mikko Koho¹, Erkki Heino^{1,2}, Petri Leskinen¹, Eero Hyvönen^{1,2}, and Tomi Ahoranta³

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

² HELDIG – Helsinki Centre for Digital Humanities, University of Helsinki, Finland

³ The National Archives of Finland

<http://seco.cs.aalto.fi>, <http://heldig.fi>,

<http://www.arkisto.fi/en/frontpage>

Abstract. This paper presents an application for studying the death records of WW2 casualties from a prosopographical perspective, provided by the various local military cemeteries where the dead were buried. The idea is to provide the end user with a global visual map view on the places in which the casualties were buried as well as with a local historical perspective on what happened to the casualties that lay within a particular cemetery of a village or town. Plenty of data exists about the Second World War (WW2), but the data is typically archived in unconnected, isolated silos in different organizations. This makes it difficult to track down, visualize, and study information that is contained within multiple distinct datasets. In our work, this problem is solved using aggregated Linked Open Data provided by the WarSampo Data Service and SPARQL endpoint.

1 Introduction

This paper builds upon the WarSampo project⁴ that collects data related to the Finnish WW2, and publishes this data as Linked Open Data [2] on an open SPARQL endpoint. WarSampo is to our best knowledge the first large scale system for serving and publishing WW2 LOD on the Semantic Web.⁵

Seven different user-friendly application perspectives based on the the WarSampo Data Service⁶ have been created within the semantic WarSampo Portal⁷ [8]. This paper presents a new, 8th application perspective based on the WarSampo Data Service: Cemeteries⁸. This work is based on research done in

⁴ <http://seco.cs.aalto.fi/projects/sotasampo/en/>

⁵ WarSampo was awarded with the LODLAM Challenge Open Data Prize in Venice, 2017.

⁶ <http://www.ldf.fi/dataset/warsa>, <https://datahub.io/dataset/warsampo>

⁷ <http://sotasampo.fi/en>

⁸ Preliminary version at <http://www.sotasampo.fi/en/cemeteries>

transforming Finnish WW2 casualties to Linked Data [11] and linking them to other datasets in WarSampo.

Finnish soldiers who perished in WW2 were transported back to their hometown for burial whenever it was possible [12]. Thus, the local cemetery is a natural starting point for studying the common characteristics and events of the residents of one's hometown in the turmoil of the war.

This paper presents new interactive views to the casualty data based on cemeteries and other community level groups. First, all military cemeteries in Finland can be browsed with a faceted browser, and visualized on maps to provide a global view of the cemeteries. Second, another new view is the cemetery information page that provides a community-level view presenting various kinds of information and visualizations about each cemetery and the soldiers buried there, including over 3000 photographs of the cemeteries. Third, the new functionalities of the existing casualties perspective make it possible to study various community level groups. For example, the people buried in a particular cemetery or born in a certain town form a prosopographical [5,15] group which is of interest to local historians. Visualizations of the casualty data can be created providing the end user with insights on what actually happened to these people originating from the same area. Furthermore, the group can be filtered further down into subsets of interest in versatile ways using the faceted browser and new visualizations of the existing Casualties perspective.

The principal use case of the new views is to provide local level views to the casualty data. This use case originates from the data owners at the National Archives of Finland, who emphasized the idea that the users need a local starting point for browsing the data. The existing person perspective of the WarSampo portal provides answers to the question: "What happened to my relatives in WW2" whereas the new views are meant to expose what happened to the user's or the user's relatives' hometown soldiers in WW2.

2 User Groups of WW2 Casualty Data

The WW2 Casualty database is originally collected and maintained by the National Archives of Finland. Based on their experiences, the data users can roughly be divided into three groups: academic researchers, military history enthusiasts and private citizens. The first group has the widest range of needs regarding the data, but they often have the best skills to handle and refine the data by themselves. The focus of academic research regarding this data seems to be shifting from a macro level towards individual and social aspects of war.

Military history enthusiasts usually approach the data from a military unit perspective, or they may concentrate on a certain location during a narrow time frame. They may also be searching for irregularities, such as peaks in numbers of casualties or in certain age groups within the data. One special group of data users in this segment are voluntary researchers, whose aim is to search, locate, and bring back remains of Finnish soldiers who went missing during the battles

in the area that is currently part of Russia. They use the data to identify the war victims found during their excavations at the WW2 battle sites.

Private citizens usually begin their search for information with their own relatives who were lost during the war. After finding that out, they may go on searching for similar destinies based on age group, unit, or locations (e.g. home towns or the location where their relatives lost their lives).

As mentioned above, academic researchers usually have the best skills to refine the data for their needs. At the other end are private citizens who are usually most dependent on easy-to-manage user interfaces. There have been no studies done on the data user profiles, but it seems apparent that they form the largest group. In this group are also school children who may benefit from this data during their studies as local history aspects are more and more valued in Finnish national curriculum for basic education.

The WarSampo project is targeted at all three user groups. The WarSampo Data Service publishes all information as Linked Open Data so that academic researchers can, at their choice, download the data and process it further, or query it directly with SPARQL. The WarSampo Portal provides user-friendly applications for all the user groups to search, browse, analyze and visualize the data.

3 Cemeteries as Linked Open Data

An important part in studying histories of communities are cemetery studies, which provide a physical record of a community's former inhabitants [14]. In this research we study cemeteries of war graves as data by linking cemeteries to the soldiers who have been buried there, and developing on-line tools and visualizations for analyzing the data.

3.1 Collecting Cemetery Data

A complete listing of all war cemeteries in Finland has not previously been available, but it has been estimated that there are about 690 of them. The Central Organization of Finnish Camera Clubs (Suomen Kameraseurojen Liitto ry, SKsL) is coordinating a project called "War Cemeteries in Finland" where local camera clubs photograph and collect data about all the war cemeteries.

The camera clubs were instructed to take five specific photographs from each cemetery. At the same time the following information will be collected: 1) official name of the cemetery, 2) year of foundation, 3) architect, 4) memorial (name, sculptor and unveiling date), 5) street address, 6) coordinates, 7) former municipality, and 8) current municipality. The cemetery and photograph data is first organized into a CSV table by a representative of the SKsL, who manually harmonizes the data entry forms sent by the camera clubs. Then the CSV table

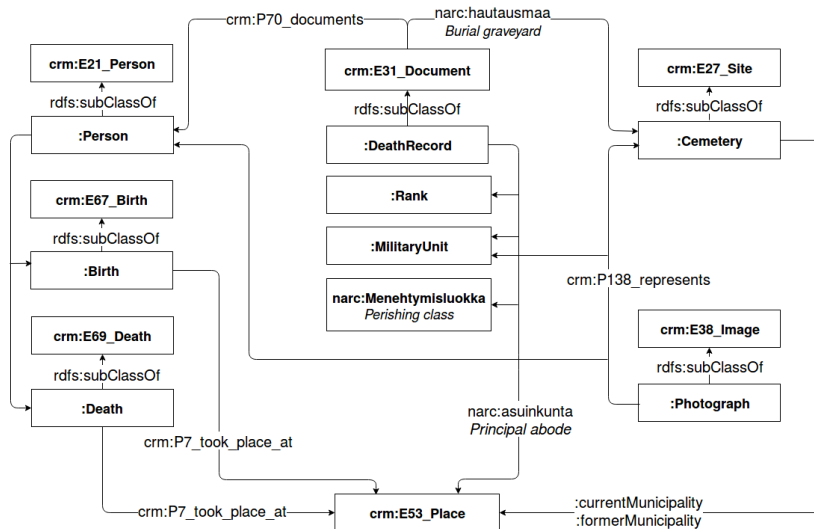


Fig. 1. The extension of WarSampo data model for cemeteries

is converted into RDF⁹ and linked to existing WarSampo datasets. Photographs and new cemetery data is then shown on the cemetery page.

3.2 Data Model and Linking War Casualties to Cemeteries

For interlinking the heterogeneous datasets of WarSampo, a wartime place ontology containing place names with different levels of granularity and types was created earlier as a pilot implementation of the “Finnish Ontology Service of Historical Places and Maps” [7]. Now, when the cemetery data is added to the place ontology, it brings out a new local level to the ontology, as there is often more than one cemetery within a municipality.

In order to add the cemetery data to the WarSampo Data Service, the event-based data model of WarSampo was extended with a new cemetery class. The part of the WarSampo data model that is relevant to this paper is presented in Figure 1. Because the WarSampo data model is founded on the CIDOC Conceptual Reference Model (CRM) [4], the cemetery class is a subclass of the CRM class E27_Site. The cemetery data is linked to the person instances via death records and to place ontology via municipalities.

The namespaces `:`, `:crm`, and `:narc` used in Figure 1 refer to WarSampo schema (<http://ldf.fi/schema/warsa/>), CRM, and Casualties schema (<http://ldf.fi/schema/narc-menehtyneet1939-45/>) respectively. Using the CRM model is not straightforward, and there are hardly any best practices

⁹ At the time of writing the information collection is still in progress and preliminary data and photographs related to 218 cemeteries have been converted into RDF and published on the WarSampo Data Service.

available. A previous modeling decision was that when a class containing custom properties that were not available in the CRM is needed, it will be created as a subclass of a corresponding CRM class, because otherwise we would have to add invalid properties to a CRM class.

Another justification for using subclasses is that the usual way of defining types of instances in Linked Data is with `rdf:type` pointing to a `rdfs:Class`, but the CRM model, originating from outside the Linked Data world, uses another way of specifying types. According to CRM the way to include type information is to use the `P2 has type` property, which is supposed to point to a concept picked from e.g. a hierarchical thesaurus. As we are producing Linked Data, we use the RDFS standard for modeling the types of the instances (e.g. place types, types of military units), which enables more powerful semantics. The full WarSampo schema is published on GitHub¹⁰.

The dataset of the Finnish war casualties consists of about 95,000 death records, and originates from the Finnish National Archives. It has served as the primary source of person instances in WarSampo. Because it was created before the WarSampo Data Service, it uses a different namespace for some properties. As can be seen from Figure 1, the death records are modeled as CRM documents, which have custom properties for different pieces of information about the casualty. The death records are linked to corresponding WarSampo person instances, military units, military ranks, and wartime municipalities [11].

The death records were originally annotated with information about cemeteries, but this information was limited only to the cemetery name, and the municipality in which the cemetery is located. Because the “War Cemeteries in Finland” project and the original death record database used the same listing of war cemetery names as a starting point, it is a straightforward operation to append the new cemetery data and photographs to the existing cemetery instances by using only the name of the cemetery as a connecting link. The cemeteries provide a more stable basis for local history than the death records’ temporally changing municipalities, which have already changed greatly since the war.

4 New Views to WW2 Casualties

The new cemetery perspective and the new visualizations in the casualties perspective have been developed to gain new insights from the data based on the community-level aspect provided by the cemeteries. Since there is not enough data about the casualties to construct life stories of individual soldiers as biographies, the new views focus on supporting prosopographical study of the war casualties in which people are studied as groups. This approach is useful when there is not enough information about individual people to construct biographies, but the amount of individuals is large enough to study the data as groups of people using, for example, visualizations. Information visualization is a useful approach for analyzing complex spatiotemporal and multivariate data [9].

¹⁰ <https://github.com/SemanticComputing/Warsampo-schema>

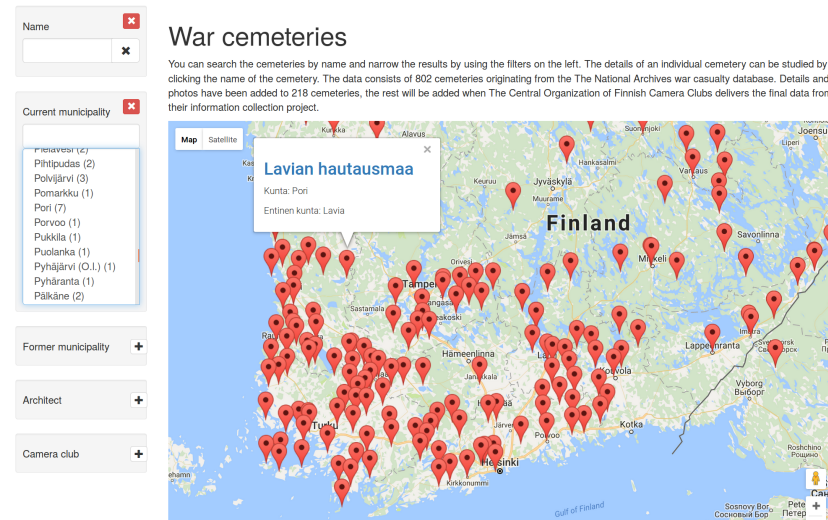


Fig. 2. Cemetery perspective

4.1 Cemetery Perspective and Information Pages

The user interface of the new Cemetery perspective¹¹ is presented in Figure 2. The user can browse all cemeteries, or search the cemeteries by name and narrow the results by using the filters on the left. The search results can be viewed as a listing, or on a map which provides a global view of the cemeteries. When the user clicks the name of a cemetery, a cemetery information page opens.

Figure 3 illustrates the top section of a cemetery information page. The top section contains all information provided by the “War Cemeteries in Finland” project. Below that the user can browse through the photographs of the cemetery with a full screen photo gallery. As is the case with all other WarSampo information pages, the right column contains links to related resources. In this context links to buried people and their units are provided. Below the photographs there is a map visualization based on the places of death of the buried people. This example shows that the death places are concentrated on the eastern front of Finland. By clicking the map pin the user can move onwards to the person information page.

The bottom section of a cemetery information page, partly illustrated in Figure 4, contains several visualizations for studying the buried people as a prosopographical group:

Distribution over units. A pie chart for showing the most common units of one’s hometown soldiers.

Distribution over ranks. The most common ranks as a pie chart.

¹¹ Preliminary version at <http://www.sotasampo.fi/en/cemeteries>

Hankaniemen hautausmaa

Foundation year: 1800 luvun alussa
 Number of graves: 280
 Current municipality: Kuhmo
 Street address: Kainuuntie 23 88900 Kuhmo
 URI: http://id.fswars/places/semetiesh0813_1

Panorama of the cemetery Single grave with a cross and a brass Memorial

Place of deaths

Links

Buried people (278) ▾

- [Eetu Jooeppi Ahtonen](#)
- [Matti Ahtonen](#)
- [Nikolai Ahtonen](#)
- [Viljo Olavi Ahtonen](#)
- [Reino Kullervo Appelgren](#)
- [Anarisa Hälonen](#)
- [Aame Johannes Hammila](#)
- [Arvo Johannes Haverinen](#)
- [Helkki Kaaleppi Haverinen](#)
- [Helkki Vilho Haverinen](#)

Units of buried people (118) ▾

- [Jalkaväkirykmentti 44, III pataljoona \(Talvisota\)](#)
- [15. rajakomppania \(Jatkosota\)](#)
- [Jalkaväkirykmentti 31, 9. komppania \(Jatkosota\)](#)
- [Jääkäripataljoona 5, Konekiväärikomppania \(Jatkosota\)](#)
- [Jalkaväkirykmentti 52, 1. komppania \(Jatkosota\)](#)
- [Eteläinen pataljoona 14, 2. komppania \(Talvisota\)](#)
- [Rajajätkäpataljoona 6, 2. komppania \(Jatkosota\)](#)
- [Jalkaväkirykmentti 27 \(Talvisota\)](#)

Fig. 3. The top section of a cemetery information page

Age distribution. The ages at death as a bar chart. This visualization can also be found in the Casualties perspective with customizable options.

Distribution over causes of death. Causes of death as a pie chart.

For prosopographical research the pie chart slices and bar chart bars were made interactive, so that by clicking them the user can examine the persons that belong to the group in question and visit their information pages. On the bottom there is a link to the casualties perspective, where the user can study the buried people further as a group with a faceted search interface and newly added visualizations. A design decision originating from the needs of the user groups was to place certain visualizations to the cemetery information page, and leave the more customizable visualizations (e.g. soldier life paths) to the more general casualties perspective.

4.2 Casualties Perspective Visualizations

The casualties perspective¹² of WarSampo provides a faceted search interface to the Finnish WW2 casualty data as a web page [11]. The faceted search web functionality is based on the SPARQL Faceter library [10]. The source code of the Casualties perspective in WarSampo is openly available in GitHub¹³.

The faceted search results have previously been displayed in a data table, but the perspective has been extended to also support visualizing the results based on the facet selections. This provides a way to easily analyze the data by

¹² <http://sotasampo.fi/en/casualties>

¹³ <https://github.com/SemanticComputing/WarSampo-death-records>

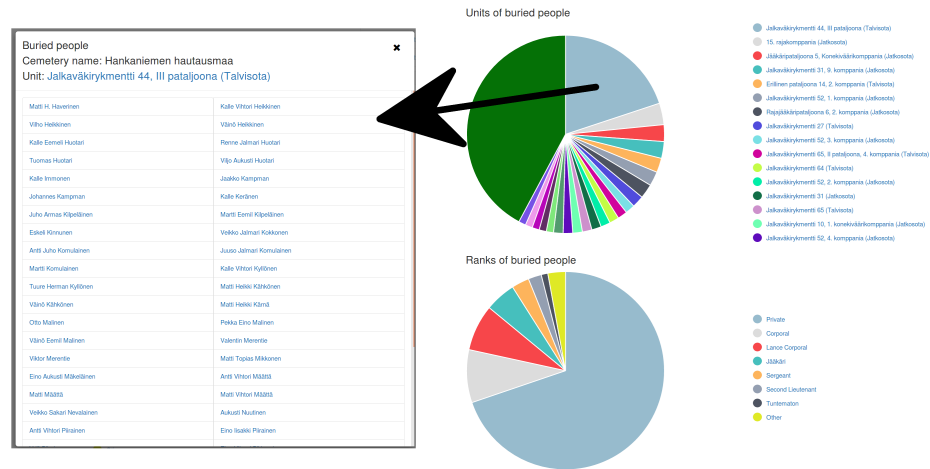


Fig. 4. Visualizations for studying the buried people as a prosopographical group on a cemetery information page

using the faceted search interface to filter the results based on what the user is interested in. For example you could select only soldiers that have been born in a certain place, or that have been buried in a certain cemetery.

After filtering the results with the facets, the user can choose from several different ways of displaying the results from the *Results Display* drop-down menu at the top of the page. The supported methods of displaying the faceted search results are:

Table. The original method of displaying the casualties as a table.

Age distribution. Age distribution of casualties as a column chart.

Soldier paths. Soldier life paths visualized as a sankey diagram. The default steps in the diagram are the municipalities of birth, residence and death, and the military cemetery. Steps can be customized by the user by selecting properties to be used for each step.

Statistics. Distribution over an arbitrary property as a bar chart. The user can select the property to use for the diagram.

Figure 5 shows the age distribution of casualties in Hietaniemi cemetery in Helsinki, which is the military cemetery with the highest number of casualties.

Figure 6 presents a screenshot of the soldier life paths diagram, showing the life paths of 40 soldiers. The diagram shows where the soldiers were born, where they lived, where they died, and where they are buried. The birth, residence, and death are given on a municipality level, whereas the cemetery is given as the exact cemetery. In this case the facets have been used to filter the casualties to only show persons buried in the cemetery of Inari in Ivalo. A large portion of places of death are unknown in the data, as is the case here as well. All in all 56% of the casualties in the dataset have an unknown place of death.

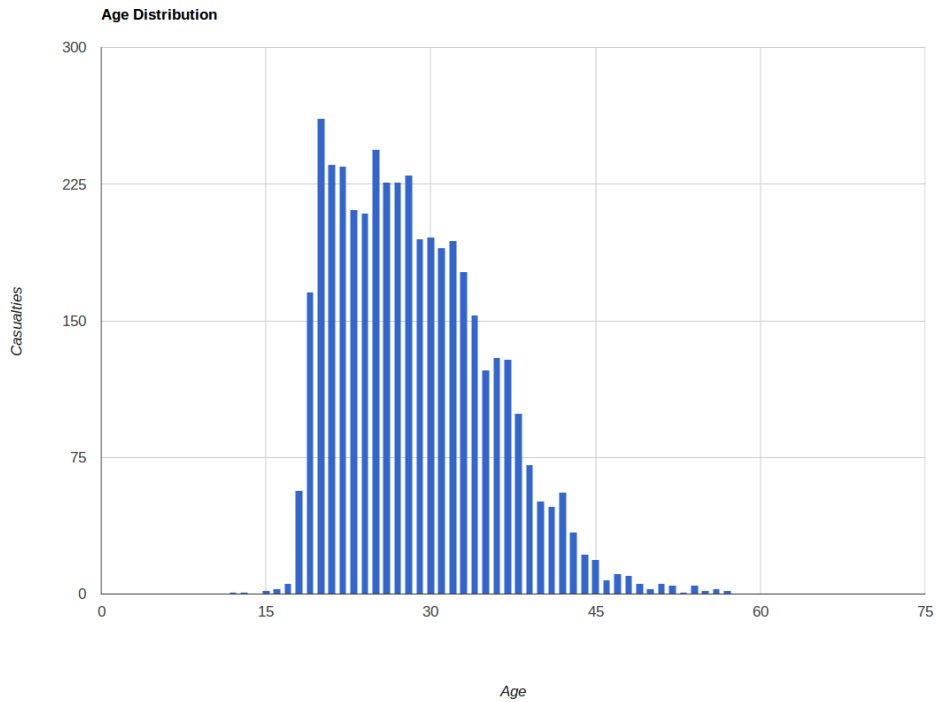


Fig. 5. Age distribution of casualties in the most "crowded" war cemetery, Hietaniemi in Helsinki.

The faceted search makes it possible to further narrow the search down e.g. to only include casualties who served in a certain military unit. Figure 7 show the whole user interface of the Statistics visualization view. Based on the facet selections, the diagram currently shows occupations of soldiers with the military rank of private, who were married, were born in Helsinki, lived in Helsinki, and are buried in Hietaniemi war cemetery in Helsinki. The most common occupations in the group are workman, chauffeur and labourer.

5 Related Work, Discussion and Future Work

Applying Linked Data principles to cultural heritage data [6] and historical research [13] has been a promising approach to solve the problems of isolated and semantically heterogeneous data sources. For modeling general war related English data Historic England hosts various thesauri¹⁴. Also a number of previous research exists in Linked Data visualization [1,3].

¹⁴ <http://thesaurus.historicengland.org.uk/>

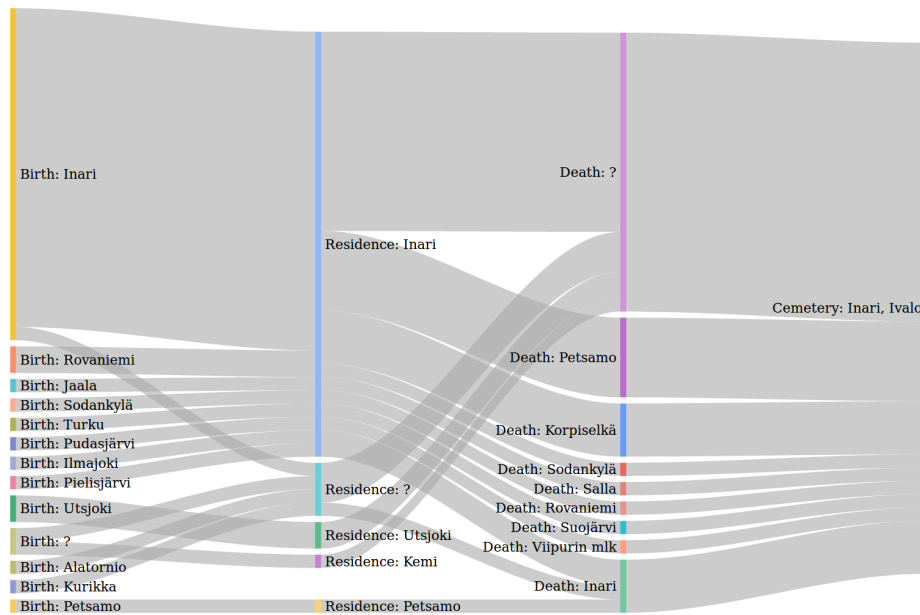


Fig. 6. Life paths of 40 soldiers buried in the cemetery of Inari in Ivalo.

The War Graves Photographic Project¹⁵, founded in 2008, aims to create an archive of names and photographs of all military graves and memorials from 1914 to the present day from any nationality, although the focus is on Commonwealth soldiers. Data collection is based on volunteer work, but unfortunately the data is not available through APIs or dumps, and the photographs are subject to a charge.

This paper presented how Finnish WW2 casualty data was linked to war cemetery data that is being collected for studying the casualties from a prosopographical perspective. The WarSampo portal was extended with several new interactive views to the casualty data based on cemeteries and other community level groups. A quick comparison shows that the War Graves Photographic Project has information on 668 people buried in Hietaniemi, the biggest war cemetery in Finland, whereas WarSampo now has information on 4268 people buried there.

The Cemetery perspective and the data visualizations on cemetery information pages are useful for all user groups mentioned in section 2, but because they provide easy access and a starting point from a local point of view, private citizens might benefit most from it. These visualizations offer instant information on divisions, for example, in ranks, age groups and units. By using a local starting point for the whole casualties dataset it is also much easier to detect

¹⁵ <https://www.twgpp.org/>

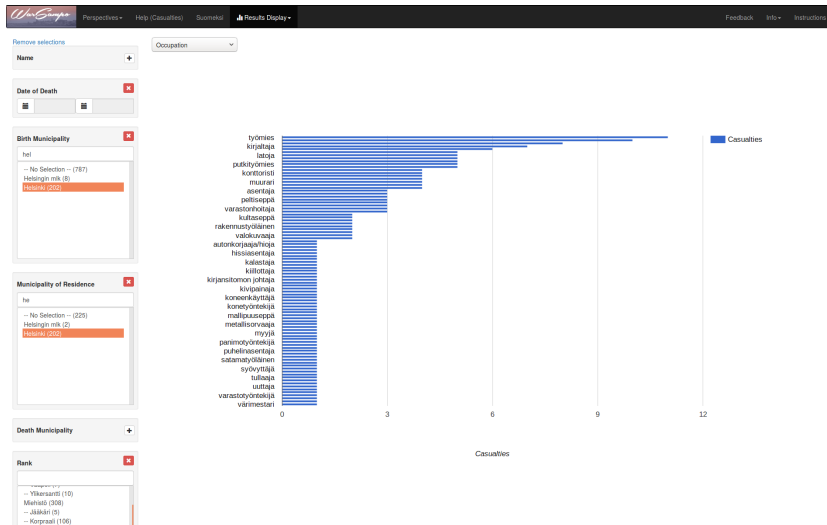


Fig. 7. The faceted search interface and the statistics view, visualizing occupations of soldiers with private military rank, who were married, born in Helsinki, living in Helsinki, and buried in Hietaniemi war cemetery in Helsinki.

local irregularities in data that might be worth exploring in more detail. The new visualizations of the Casualties perspective offer more customizable views to the casualty data as a whole, which are especially useful for academic researchers.

While the current tools allow for many ways to analyze and explore the data, there is still a vast amount of different kinds of visualizations that could be tried, e.g. many that would be based on aggregating the data in some manner. For example displaying the average amount of children in relation to some other property, like place of residence, might reveal some new insights.

The casualty and cemetery data collected from various sources contains errors and inconsistencies. To improve the quality of the data, a Linked Data tool for suggesting corrections and collecting new data from the users is being planned.

Acknowledgements Our work is funded by the Memory Foundation for the Fallen¹⁶, Open Science and Research Initiative¹⁷ of the Finnish Ministry of Education and Culture, the Finnish Cultural Foundation, and the Academy of Finland.

¹⁶ <http://www.muistosaaatio.fi/>

¹⁷ <http://openscience.fi/>

References

1. Bikakis, N., Sellis, T.: Exploration and visualization in the web of big linked data: A survey of the state of the art. In: Proceedings of the Workshops of the EDBT/ICDT 2016 Joint Conference. CEUR Workshop Proceedings, Vol-1558 (2016)
2. Bizer, C., Heath, T., Berners-Lee, T.: Linked Data—The Story So Far. Semantic services, interoperability and web applications: emerging concepts pp. 205–227 (2009)
3. Dadzie, A.S., Rowe, M.: Approaches to visualising Linked Data: A survey. *Semantic Web* 2(2), 89–124 (2011)
4. Doerr, M.: The CIDOC CRM – an ontological approach to semantic interoperability of metadata. *AI Magazine* 24(3), 75–92 (2003)
5. Fleming, R.: Writing biography at the edge of history. *The American Historical Review* 114(3), 606–614 (2009), <http://dx.doi.org/10.1086/ahr.114.3.606>
6. Hyvönen, E.: Publishing and Using Cultural Heritage Linked Data on the Semantic Web. *Synthesis Lectures on the Semantic Web: Theory and Technology*, Morgan & Claypool, Palo Alto, CA, USA (2012)
7. 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), vol 1608
8. 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 (2016)
9. Kehrer, J., Hauser, H.: Visualization and visual analysis of multifaceted scientific data: A survey. *IEEE transactions on visualization and computer graphics* 19(3), 495–513 (2013)
10. Koho, M., Heino, E., Hyvönen, E.: SPARQL Faceter—Client-side Faceted Search Based on SPARQL. In: Troncy, R., Verborgh, R., Nixon, L., Kurz, T., Schlegel, K., Vander Sande, M. (eds.) *Joint Proc. of the 4th International Workshop on Linked Media and the 3rd Developers Hackshop*. No. 1615, CEUR Workshop Proceedings (2016), <http://ceur-ws.org/Vol-1615/semdevPaper5.pdf>
11. Koho, M., Hyvönen, E., Heino, E., Tuominen, J., Leskinen, P., Mäkelä, E.: Linked Death - Representing, Publishing, and Using Second World War Death Records as Linked Open Data. In: *The Semantic Web: ESWC 2016 Satellite Events*. Springer (June 2016)
12. Leskinen, J., Juutilainen, A. (eds.): *Jatkosodan pikkujättiläinen*. WSOY, Finland (2005)
13. Meroño-Peñuela, A., Ashkpour, A., Van Erp, M., Mandemakers, K., Breure, L., Scharnhorst, A., Schlobach, S., Van Harmelen, F.: Semantic technologies for historical research: A survey. *Semantic Web* 6(6), 539–564 (2015)
14. Veit, R.F., Nonestied, M.: *New Jersey cemeteries and tombstones: history in the landscape*. Rutgers University Press (2008)
15. 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), <http://dx.doi.org/1854/8212>