
This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.

Koho, Mikko; Ikkala, Esko; Hyvönen, Eero

Reassembling the Lives of Finnish Prisoners of the Second World War on the Semantic Web

Published in:

Proceedings of the Third Conference on Biographical Data in a Digital World (BD 2019)

Published: 12/06/2022

Document Version

Publisher's PDF, also known as Version of record

Published under the following license:

CC BY

Please cite the original version:

Koho, M., Ikkala, E., & Hyvönen, E. (2022). Reassembling the Lives of Finnish Prisoners of the Second World War on the Semantic Web. In *Proceedings of the Third Conference on Biographical Data in a Digital World (BD 2019)* (pp. 31-39). (CEUR Workshop Proceedings; Vol. 3152). CEUR. http://ceur-ws.org/Vol-3152/BD2019_paper_5.pdf

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

Reassembling the Lives of Finnish Prisoners of the Second World War on the Semantic Web

Mikko Koho¹, Esko Ikkala¹, Eero Hyvönen^{1,2}

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

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

<http://seco.cs.aalto.fi/>, firstname.lastname@aalto.fi

Abstract

This paper presents the first results of a new, ninth application perspective for the semantic portal *WarSampo – Finnish WW2 on the Semantic Web*, based on a database of ca. 4 450 Finnish prisoners of war in the Soviet Union. Our key idea is to reassemble the life of each prisoner of war by using Linked Data, based on information about the person in different data sources. Using the enriched aggregated data, a biographical global “home page” for each prisoner of war can be created, that is more complete than information in individual data sources. The application perspective is targeted to the researchers of military history, to study and analyze the data in order to form new research questions or hypotheses, as well as to public in the large looking for information, e.g., about their relatives that were captured as prisoners of war. Employing the faceted search of the application perspective, prosopographical research on subgroups of prisoners is possible.

1 Introduction

Representing biographical texts as Linked Data leads to a paradigm change in publishing biographical collections (Hyvönen et al., 2019): the lives can then not only be read as texts by humans but also be processed and analyzed by computational means (Fokkens et al., 2017; Warren et al., 2016), opening new possibilities in Digital Humanities (Gardiner and Musto, 2015) research for biography and prosopography (Verboven et al., 2007) as well as for data reuse in applications. The same idea of Linked Data can be applied also when biographical data is available in semi-structured or structured form from different data sources: the data about a person can be aggregated, harmonized, and reassembled into a global knowledge graph that gives a more complete picture of the biographee than any individual source alone. Based on the knowledge graph, a biography of the biographee can be generated or alternatively a semi-structured “home page” presenting her/his life. The latter approach was introduced in the semantic portal *WarSampo – Finnish WW2 on the Semantic Web*¹ (Hyvönen et al., 2016), a web service in use in Finland that had 230 000 users in 2018, typically looking for information about their relatives killed in action during the Second World War (WW2).

This paper presents a new, ninth application perspective *Prisoners of War* to be included in WarSampo. This perspective was created for studying individual people, documented in a new prisoners of war (POW) database, as well as groups of them for prosopographical analysis. The new data was aligned with and integrated into the WarSampo person data, which is mostly based on the Finnish WW2

casualties of war² database of the National Archives of Finland. The new application perspective enables studying not only individuals but also prosopographical studies of the prisoners using either the whole dataset or subsets of it based on user interest and selections in a faceted search (Tunkelang, 2009) view.

The new prisoners of war dataset was originally published as a book (Alava et al., 2003). For integrating and publishing the data as a part of WarSampo, it has been further extended, cleaned, and validated by domain experts using, e.g., information from many war-time archives in Finland and Russia. This paper builds on previous work on WarSampo, which has discussed the Linked Data publication and data model (Koho et al., 2018a), and the data integration challenges (Koho et al., 2018b). Reconstructing the biographies of the casualties of war in WarSampo has been previously presented in (Koho et al., 2017). In contrast to the casualties of war dataset, the POW register can have multiple values for a single property, and contains sources of information for individual data values, creating a need for handling conflicting information about a person.

In the following, the underlying data model and data production process is first explained. After this, the main functionalities of the application from an end user perspective are explained, as well as the technical implementation. In conclusion, the contributions of the work are summarized and contrasted with related work.

2 Data Model and Data

The prisoners dataset consists mainly of a register of the Finnish prisoners of war in WW2, containing a spreadsheet of about 4 450 soldiers, auxiliary forces, and civilians captured by the army of the Soviet Union. Additional spreadsheets contain information about POW camps and hospitals, as well as the primary data sources. The data includes

¹This semantic portal was released in 2015 and is in use at <https://sotasampo.fi/en/>. More information about the project is available at home page <https://seco.cs.aalto.fi/projects/sotasampo/en/>.

²<http://kronos.narc.fi/menehtyneet/>

also separate documents about the prisoners of war to provide additional information, such as video interviews, images and archived documents.

The original information sources are mostly various registers in Finnish and Russian archives (Alava et al., 2003). Information in different sources can be contradictory, hence it is important to preserve the data source for each individual piece of information. A formatting was agreed upon to allow multiple values with source information already in the original spreadsheet that the domain experts worked on. The data formatting evolved as a collaboration between the domain experts maintaining the original dataset, and the WarSampo team of Linked Data experts. Also other agreements on the spreadsheet structure were needed: 1) separation and cleaning of values that will be linked to the WarSampo domain ontologies, 2) local identifiers for entities that are used in multiple spreadsheets, and 3) how to express partially or completely missing information.

The WarSampo infrastructure, data service, and semantic portal was chosen as the primary data publication platform by the stakeholders, which include the National Archives of Finland, and the Association for Cherishing the Memory of the Dead of the War.

21 Prisoners of War as Linked Data

The WarSampo *Linked Open Data infrastructure* is built to support integrating new datasets into WarSampo, by extending both the data model and the data content. The data is published openly online for everyone to use. The WarSampo web portal then provides different perspectives to the interlinked datasets, as customized web applications. New perspectives can be added to provide views to new datasets, or to show new features of the existing data.

In *Linked Data* (Heath and Bizer, 2011), information is presented as *RDF* graphs and all resources in the data have unique identifiers. This enables identifying and sharing common resources, e.g. people, places, and military ranks between the datasets, thus creating an interlinked *knowledge graph*.

A simple primary data model is used for the prisoner records, in which one prisoner record corresponds to one row in the source spreadsheet, with each column mapped to a distinct property. So all of the personal information about each captured individual is contained in the prisoner record, resembling the data model of the WarSampo death records (Koho et al., 2017). The properties and classes of prisoner records and death records have been harmonized using the dumb-down principle of Dublin Core³, i.e., by using shared super-properties and super-classes where applicable. By mapping columns directly to properties, the data can be shown to the end user in an intuitive way, resembling the original spreadsheet.

WarSampo uses the *CIDOC Conceptual Reference Model (CRM)*⁴ as the harmonizing data model. Prisoner records are modeled as instances of the CRM document class `E31.Document`.

³<http://dublincore.org/usage/documents/principles/>

⁴<http://cidoc-crm.org>

In addition, this data is then used to create CIDOC CRM descriptions of the actual people and events, when appropriate. WarSampo person instances (Leskinen et al., 2017) in the actor ontology are enriched using the prisoner records. New person instances are created for people that do not already exist in the ontology, which is the case for most of the war prisoners. The prisoner records then document the person instance through the CRM property `P70.documents`. The full WarSampo data model is published on GitHub⁵.

22 Data Conversion

It has been understood from our previous work, that the data transformations need to be repeatable, automated processes (Koho et al., 2018b), in the dynamic infrastructure where there is frequently a need to adapt to changes. An automatic data processing pipeline⁶ was developed to integrate the POW data into WarSampo linked data infrastructure. The pipeline handles data transformation, validation, linking, and harmonization.

The pipeline transforms the spreadsheets into RDF, mapping the spreadsheet columns to RDF properties, with possibly multiple values per property, and containing annotations for primary information sources. Automatic probabilistic entity linking processes then link the records to the WarSampo domain ontologies of military ranks, units, occupations, people, and places. Original literal values are also retained as separate properties.

The original POW register is maintained in spreadsheet format, which can be easily integrated into WarSampo with our automated transformation process when the spreadsheet is updated, provided that the structure stays the same. Also if the linked domain ontologies are updated, the whole integration process can be redone to account for the changes in the probabilistic entity linking.

The cell formatting is validated during the data transformation process. Also other simple data validation rules are applied to find anomalies during data conversions. The validation reports help the domain experts to improve the quality of the source data.

Some parts of the data had to be left out of the online data publication due to privacy issues. This is done automatically based on the date when a person has died. If there is no information about an individual's date of death, it is assumed that they may still be alive, and their personal information, including given names, is removed, effectively pseudonymizing them. For prisoners who are known to have died less than 50 years ago, health related information is removed, based on the columns of the original spreadsheet that might contain health related information.

23 Interlinking within WarSampo

Matching the people in the prisoner records to the ca. 100 000 people already existing in the WarSampo actor ontology is one of the most challenging aspects of the data transformation pipeline. The data model and contents are

⁵<https://github.com/SemanticComputing/Warsampo-schema>

⁶Source codes for data conversion and linking are available online: <https://github.com/SemanticComputing/WarPrisoners>.

different, and many pieces of personal information can be missing on both sides. In the first results of the person linking, we were able to link 1431 prisoner records to existing WarSampo person instances, corresponding to 32% of all prisoner records (Koho et al., 2018a). The person linking uses probabilistic *record linkage* (Gu et al., 2003; Gregg and Eder, 2019) (aka. *deduplication*) with a machine learning approach, in which each POW’s information is compared with the information in the WarSampo person instances to find matches that have high enough similarity. Initially the record linkage value comparisons were weighted based on domain knowledge, which was then iterated for better accuracy, and finally a manually curated list of matches was taken to serve as training data for the machine learning approach. The machine learning approach can adapt to data changes on both sides in the record linkage, without having to manually inspect the linking results and adjust the weights.

New person instances are created from the unlinked prisoner records and added into the actor ontology. With the probabilistic record linkage, it is possible that a record is not mapped simply because there is not enough information about either the POW record, or the person instance, to create a mapping between them. Modifying the information in either the POW data or in the actor ontology means that the whole record linkage process should be redone.

Other information is also linked to WarSampo domain ontologies. Of military ranks, 99% were linked to the WarSampo military ranks domain ontology. Of military units, 91% were linked to pre-existing military units in the actor ontology.

Domain ontologies differ from each other by nature. For example, covering and disambiguating all military ranks is clearly a simpler task than performing the same task with all wartime places. In general, it is not realistic to assume that the domain ontologies completely cover their domain. Other information still to be linked to WarSampo domain ontologies are war-time municipalities. More accurate place information could also be linked, but due to the ambiguous nature of the names, this would lead to a high level of error, based on initial experiments.

The created Linked Data stores source information when present in the original data. There are many ways of presenting this kind of provenance information in RDF (Hartig, 2009; Zhao et al., 2010). The approach used with the prisoners of war dataset is storing source information using RDF reification with the DCMI Metadata Terms⁷ property *source*.

24 Biographical Data

Each person’s basic personal information in the dataset contains columns like first and last names, dates of birth, return from captivity, and death, municipality of birth, domicile and death, and occupation, marital status, and number of children. These enable building some understanding about the life of the person before the war, and in case of survivors, also after the war.

⁷<http://dublincore.org/documents/dcmi-terms/>

Structured information is also gathered of the events of going missing and being captured, like the place and time. Biographically interesting information is also given as prose about being captured, the cause of death and burial place, and other information. These all are structured to contain the information source, and can often contain different pieces of information from different sources. Information on confiscated possessions and their estimated value sheds light to what kind of valuable personal possessions a person had. Information is also given about the occurrence of a person in Soviet war propaganda magazines or fliers, either in pictures or text.

3 Prisoners of War in the WarSampo Portal

A new application perspective was created into the WarSampo portal for studying, exploring and analyzing the prisoners of war dataset as a whole. Also the existing WarSampo Persons perspective, which generates a “home page” for each person in the WarSampo knowledge graph, was extended to show possibly contradictory data originating from multiple sources (e.g. death records, prisoner records, Wikipedia). The Prisoner perspective application is open-source, and available online⁸.

31 Biographical View in the Persons perspective

The WarSampo Persons perspective offers a general search of people in the WarSampo knowledge graph. Each person is provided with a biographical view, a home page, that re-assembles the biographical knowledge of the person from the WarSampo datasets, into a structured format.

Figure 1 shows an example of a soldier’s home page, where the information is combined from a prisoner record and a death record. The left side of the page contains a person selector and a text box for filtering the people by name. The details of a selected person are displayed on the right. Information usually exists from birth to death, with a clear and understandable focus on the war-time events. A property (e.g. occupation) may contain multiple values. In order to make the biographical view as transparent as possible, all values have been supplemented with a reference to the information source. In the figure, source number 2 refers to the POW register. There is a total of 12 sources of information for the particular person, which includes also a death record, and 10 different sources from the POW register.

The values that have been linked to WarSampo domain ontologies are shown as links to corresponding home pages. The idea here is that the WarSampo semantic portal acts as a customized graphical RDF browser, which makes it possible for the user to find surprising connections between the individual resources of the WarSampo knowledge graph.

32 Prosopographical Prisoners Perspective

The Prisoners perspective is based on the previously released Casualties perspective (Koho et al., 2017). The main design principle of these perspectives is to target one core class of WarSampo knowledge graph (e.g., prisoner record) and provide the user with a faceted search (Tunkelang,

⁸<https://github.com/SemanticComputing/prisoners-demo>

Persons

Search for known persons from the past Finnish wars by writing their name in the text input below and/or selecting a person from the list below. Information regarding the person and recommended links will appear on the right. **If you cannot find the person you are looking for, and know in which military unit they have served, you can take a look at the unit's timeline.**

Härmä, Osmo Juhani (Military)
Härmä, Tuomo Oskari (Private)

Osmo Juhani Härmä

[Information](#) [Timeline](#) [Photographs](#)

URI: http://ldf.fi/warsa/actors/person_p753249

Personal Details

Family name	Härmä ^[1, 2]
Given names	Osmo Juhani ^[1, 2]
Born	14.01.1924 ^[1, 2]
Municipality of birth	Kemi ^[1]
Municipality of domicile	Helsinki ^[1, 2]
Nationality	Finnish ^[1]
Mother tongue	Finnish ^[1, 2]
Marital status	Not married ^[1, 2]
Number of children	0 ^[1, 2]
Occupation	Driver ^[1, 2] Filer ^[2]
Rank	Military Engineer ^[2] Private ^[1]
Military Unit	Pioneeripataljoona 34, 1. komppania ^[1, 2]

Disappearance Details

Date of disappearance	12.05.1943 ^[1]
Place of disappearance	Uhtua ^[1]

Imprisonment details

Date of capture	12.05.1943 ^[2]
Municipality of capture	Uhtua ^[2]
Location of battle in which captured	Valkeajärven maasto ^[2]
Description of capture	12.5.1943 mennyt vihollisen puolelle Uhtuan suunnalla, Valkeajärven maastossa ^[2, 3] Loikannut Uhtuan suunnalla NL:n puolelle ^[2, 4] Antautui vangiksi (loikkari), koska 12.5. hänen piti joutua oikeuden eteen syytettynä tottelemattomuudesta ylempiarvoista komentajaa kohtaan sekä tämän pahoinpitelystä ^[2, 5] Karannut 11.5.1943 ^[2, 6]
Captivity locations	4 ^[2] 158 (02.07.1943 – 09.10.1943) ^[2] 241/6 (24.10.1943 – 09.12.1943) ^[2] 241/3 (09.12.1943 – 17.01.1944) ^[2] Sotavankisairaala 2074 (03.02.1944 – xx.xx.1944) ^[2]
Additional information	Syntynyt noin 1924, kotoisin Kemistä, asunut viimeksi Helsingissä, loikannut Uhtuan suunnalla NL:n puolelle, olivat yhdessä Tsherepovetsin leirillä v. 1943 ja Uralissa leirillä 241/6, josta joulukuussa 1943 vietiin johonkin sairaalaan ^[2, 7] Ilmoittaja Vahalinnan Aarre. Härmä Osmo, syntynyt Kemissä 1924. Kuollut Pinjugissa 29.7.1944. Toinen ilmoittaja Malinen Aarne ^[2, 8] Kuolintodistuksen mukaan kuollut 29.7.1944. Sairauskertomuksen mukaan kuollut 29.7.1944 klo 7:00. Ruumiinavauspäiväksi ilmoitetaan 31.7.1944, hautauspäiväksi 30.7.1944. ^[2, 5]

Death Details

Date of death	29.07.1944 ^[10, 11, 2, 5, 9] 29.08.1944 ^[12, 2, 5] 22.07.1943 ^[1]
---------------	--

Figure 1: The Persons perspective showing part of a person's home page.

2009; Oren et al., 2006) interface, which initially renders a result set that contains all instances of the target class as a paginated table. This way we ease off the “blank search field problem”, where a new user does not know what kind of query terms should be used for meaningful results. The initial result set can be narrowed down by using various facets (e.g., military unit or prison camp).

Figure 2 shows a part of the Prisoners perspective user interface. Facets are presented on the left of the user interface. The number of hits (instances of the target class) produced by each facet value is calculated dynamically and is shown in parenthesis. Facet values leading to an empty result set are hidden. To reduce unnecessary data fetching, most of the facets are disabled by default. They can be activated by clicking the plus sign on the facet header. The facets are *name, date of being captured as a POW, date of death, military unit, military rank, POW camps where the person has been, occupation, marital status, number of children, birth municipality, place of being captured, and place of death*. The results are displayed on the right side of the user interface. The result set, based on the facet selections, can be shown as a table, or shown with three different visualizations:

1. a distribution chart over a selected property, with property choices: military rank, military unit, occupation, number of children, birth municipality, municipality of residence, place of being captured, and place of death,
2. an age distribution chart at the time of capturing,
3. a sankey diagram of soldier life paths based on known geographical locations at different times, starting from the municipality of birth, and ending to the municipality of death.

The results display mode can be selected using the button in the top bar. In Figure 2, the results are displayed as a table, with each row corresponding to a single prisoner record, with several key properties mapped to separate columns.

Figure 3 shows the age distribution of all soldiers whose rank is private at the time when they have been captured as a prisoner of war. Figure 4 shows the military rank distribution of the soldiers that were born in Helsinki.

The common usage scenario of the average user is to search for information about their relatives who have participated in the war. This can be achieved most easily with the table view of results and using the different facets, and mostly the name facet, where a person can search with just a part of the name to get all the results containing that. Another way to find relatives, who historically are often situated in the same region, is to filter the results with the birth municipality facet.

Another usage scenario is studying and analyzing the data by a historian or an interested citizen. The facets already provide distributions of the facet values, with the number of hits after each value. When a selection is made in one of the facets, all of the facets are updated to show the distribution of values with that selection. Further analysis can be done with the various visualizations of the facet results. New visualizations, e.g. locations of the POW camps on a

map, can be added rather easily to the application, and the existing ones extended as needed.

4 Implementation

The Prisoners perspective is an AngularJS⁹ web application, which consists of several modules. The facet functionality is implemented using SPARQL Faceter¹⁰ (Koho et al., 2016), a module that provides

- a set of directives that work as configurable facets,
- a service that synchronizes the facet selections,
- a service for updating the URL parameters based on facet selections, and retrieving the facet values from URL parameters,
- a service for retrieving SPARQL results based on the facet selections, using a configurable query template.

For querying the SPARQL endpoint, mapping the SPARQL results into JavaScript objects and paging the results, we have developed another general module¹¹ that is being used across the WarSampo semantic portal.

In addition to the default paginated table result view, powered by the ngTable¹² directive, we have implemented several reusable visualization directives for displaying the results on modern or historical maps or as statistical distributions. For the Prisoners perspective, a new sankey visualization directive was built using Google Charts.¹³

The Persons perspective is part of the WarSampo portal AngularJS core infrastructure¹⁴. It was extended to fetch data to the person’s homepage from the prisoner records, along with the source reifications. The page was redesigned and restructured to be able to integrate the data from the prisoner records, and to show the prisoner record data along with the information from a person instance and a death record, of which the latter may or may not be present. Showing and numbering the information sources was also a new addition.

5 Discussion

This paper presented first results of publishing the prisoners of war dataset as part of WarSampo. The POW data contains sensitive information about the individual citizens, some of whom are still alive. The publication of the data has been delayed due to the evaluation as to what information can be legally published about the individuals, and what needs to be hidden. The dataset and new portal is expected to be finally published in November 2019.

The combination of faceted search and various result visualization components forms the base of the user interface

⁹<https://angularjs.org/>

¹⁰<https://github.com/SemanticComputing/angular-semantic-faceted-search>

¹¹<https://github.com/SemanticComputing/angular-paging-sparql-service>

¹²<https://github.com/esvit/ng-table>

¹³<https://github.com/angular-google-chart/angular-google-chart>

¹⁴<https://github.com/SemanticComputing/warsampo-angular-app>

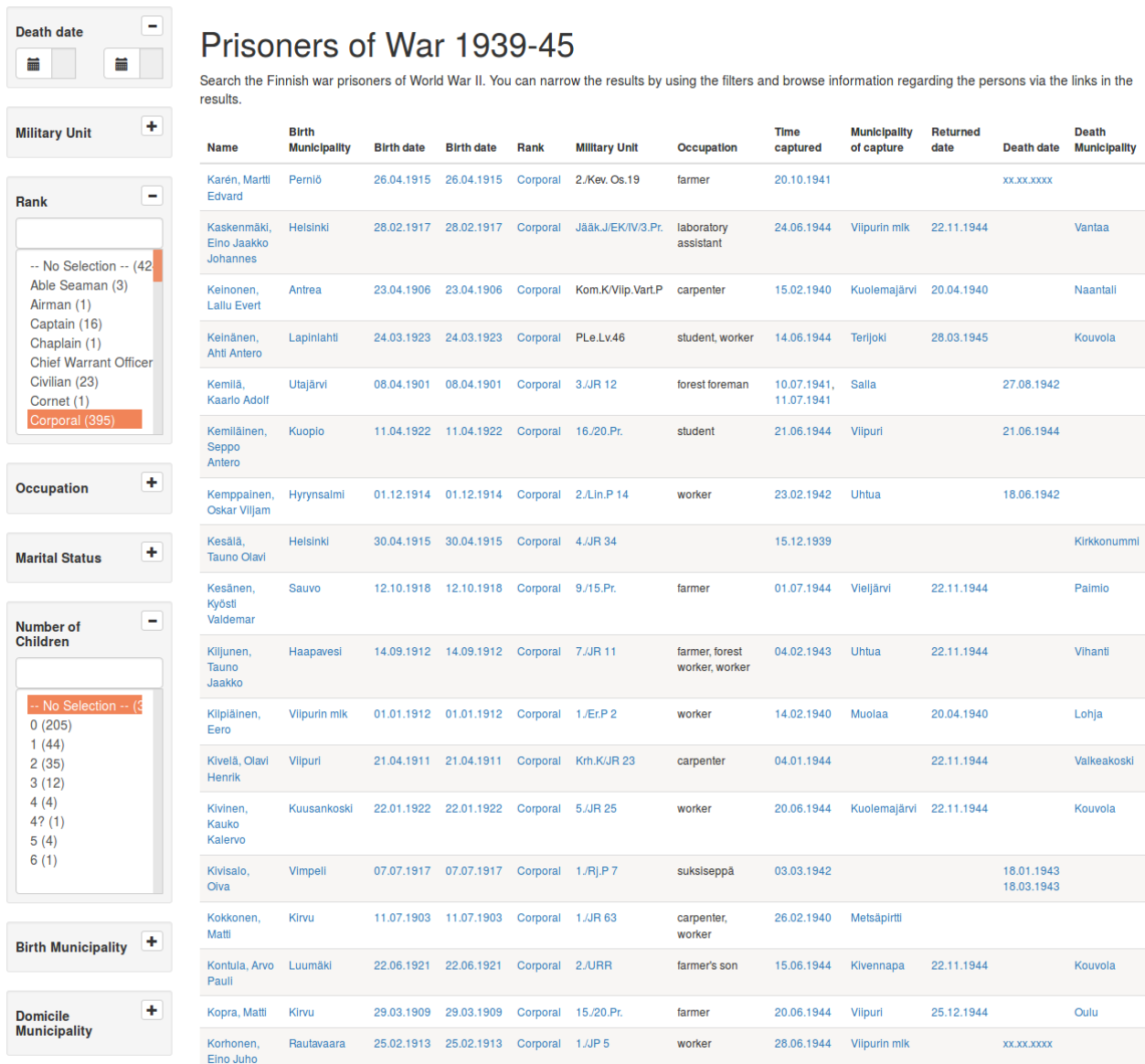


Figure 2: Prisoners perspective: facet selection results shown as a table view.

of the Prisoners perspective. This design has proved to be broadly applicable to many kinds of datasets. By browsing through the facets, the user can quickly see what kind of values have been used for different properties. This often reveals inconsistencies and spelling errors, if the property values have not been systemically entered or harmonized, or they are completely missing for a large number of resources. For estimating the completeness and the reliability of the dataset, looking at the actual property values is often more important than focusing on data modeling details.

Maintaining interlinked datasets and domain ontologies present new challenges (Auer et al., 2012; Maedche et al., 2003), as changes in one part need to be accounted for in other interlinked parts. The Linked Data environment is not yet mature enough to have easy-to-use tools for non-technical people to use for editing and maintaining interlinked data. Hence, the POW data is still maintained using the spreadsheet with agreed upon formatting and structur-

ing, which can then be re-integrated easily into WarSampo. The Linked Data approach requires tighter co-operation with the domain experts and data publishers, especially in the creation phase of historical information (Boonstra et al., 2004), than more traditional data publishing ways. However, it is possible using Linked Data to create an understanding about the whole of the war, by combining information from several datasets together, which would not be easy by studying the individual datasets directly.

The historical occupations in the WarSampo datasets have recently been harmonized into a manually curated SKOS-based¹⁵ ontology *AMMO* (Koho et al., 2019), to which the prisoner records are linked. The ontology combines synonymous occupational labels into harmonized occupation resources, and provides structures of social stratification and occupational groups. It will enable studying the prisoner records using new facets in the future, such as social

¹⁵<https://www.w3.org/TR/skos-primer/>

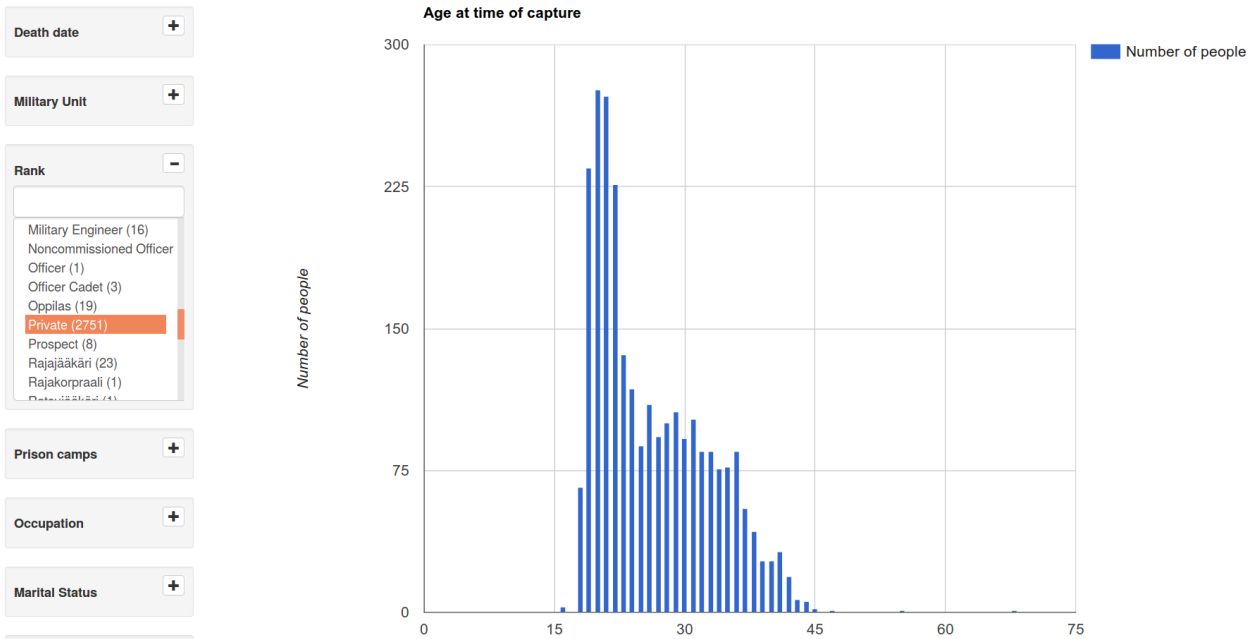


Figure 3: Prisoners perspective: age distribution of the soldiers with the military rank private.

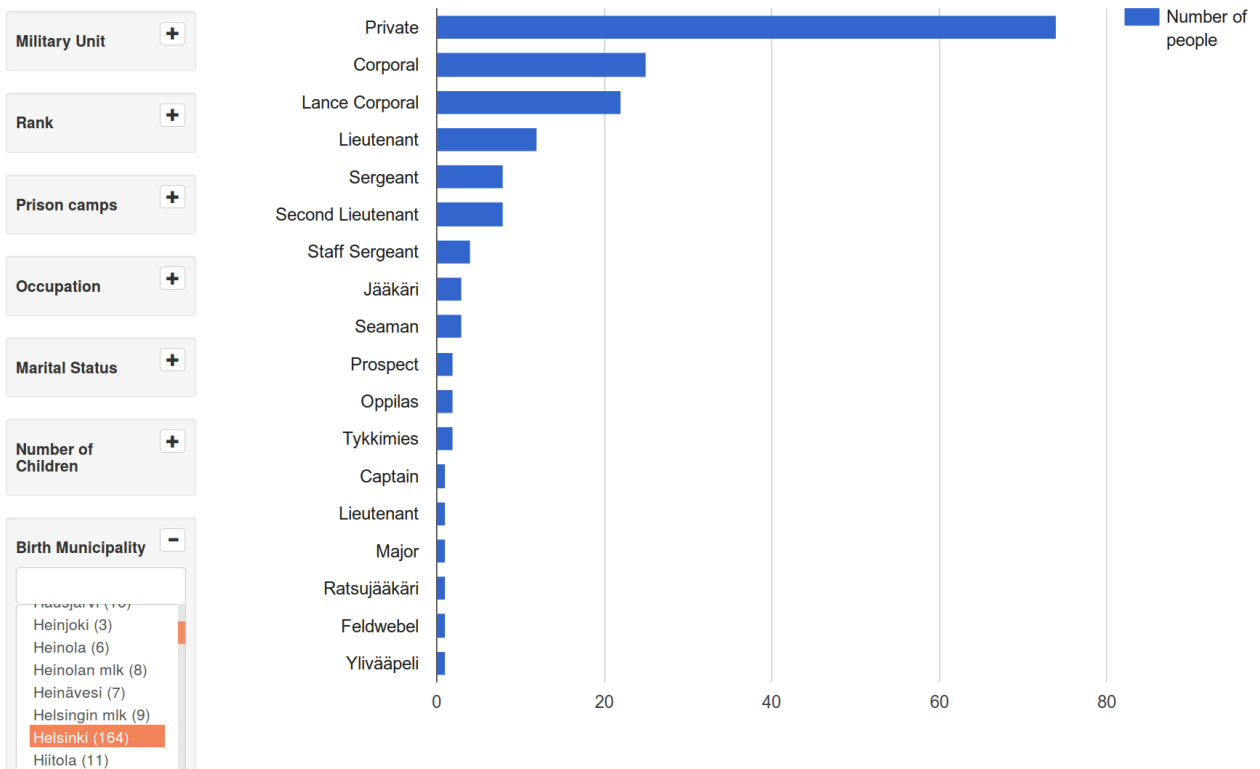


Figure 4: Prisoners perspective: statistics view

class and field of work, and facilitate the use of the dataset to answer new kinds of research questions of collaborating historians.

Integration of videos and other documents relating to the prisoners of war, will be implemented later, and will consist of expressing the document metadata in terms of CIDOC CRM, and linking the prisoners to the related document re-

sources, which in turn contain URL links to the document files.

Integrating data into a Linked Data infrastructure is more laborious than simpler ways of publishing the data as an independent data object, which does not communicate with other datasets. However, the result of the integration is an interlinked knowledge base, where the interlinked graphs

enrich each other, creating a whole that is greater than the sum of its parts (Hyvönen, 2012).

Acknowledgements

Reijo Nikkilä, Tiia Moilanen, and Pertti Suominen of The National Prisoners of War Project worked on the data as domain experts. Katri Miettinen indexed related documents for linking with persons.

Our work was funded by the Association for Cherishing the Memory of the Dead of the War¹⁶, Teri-Säätiö, Open Science and Research Initiative¹⁷ of the Finnish Ministry of Education and Culture, the Finnish Cultural Foundation, and the Academy of Finland.

The authors wish to acknowledge CSC – IT Center for Science, Finland, for computational resources.

6 References

- Teuvo Alava, Dmitri Frolov, and Reijo Nikkilä. 2003. Rukiver. *Suomalaiset sotavangit Neuvostoliitossa. Helsinki: Edita.*
- Sören Auer, Theodore Dalamagas, Helen Parkinson, and Bancilhon and. 2012. Diachronic linked data: towards long-term preservation of structured interrelated information. In *Proceedings of the First International Workshop on Open Data*, pages 31–39. ACM.
- Onno Boonstra, Leen Breure, and Peter Doorn. 2004. Past, present and future of historical information science. *Historical Social Research*, 29(2):4–132.
- Antske Fokkens, Serge ter Braake, Niels Ockeloen, Piek Vossen, Susan Legêne, Guus Schreiber, and Victor de Boer. 2017. Biographynet: Extracting relations between people and events. In *Europa baut auf Biographien*, pages 193–224. New Academic Press, Wien.
- Eileen Gardiner and Ronald G. Musto. 2015. *The Digital Humanities: A Primer for Students and Scholars*. Cambridge University Press, New York, NY, USA.
- Forest Gregg and Derek Eder. 2019. Dedupe. <https://github.com/dedupeio/dedupe>.
- Lifang Gu, Rohan Baxter, Deanne Vickers, and Chris Rainsford. 2003. Record linkage: Current practice and future directions. *CSIRO Mathematical and Information Sciences Technical Report*, 3:83.
- Olaf Hartig. 2009. Provenance information in the web of data. In *Proceedings of the WWW2009 Workshop on Linked Data on the Web*, volume 538 of *CEUR Workshop Proceedings*.
- Tom Heath and Christian Bizer. 2011. *Linked Data: Evolving the web into a global data space*. Synthesis Lectures on The Semantic Web: Theory and Technology. Morgan & Claypool Publishers, Palo Alto, USA.
- Eero Hyvönen, Erkki Heino, Petri Leskinen, Esko Ikkala, Mikko Koho, Minna Tamper, Jouni Tuominen, and Eetu Mäkelä. 2016. 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)*, pages 758–773. Springer-Verlag.
- Eero Hyvönen, Petri Leskinen, Minna Tamper, Heikki Rantala, Esko Ikkala, Jouni Tuominen, and Kirsi Keravuori. 2019. Biographysampo – publishing and enriching biographies on the semantic web for digital humanities research. In *Proceedings of the 16th Extended Semantic Web Conference (ESWC 2019)*, pages 574–589. Springer-Verlag.
- Eero Hyvönen. 2012. *Publishing and Using Cultural Heritage Linked Data on the Semantic Web*. Synthesis Lectures on the Semantic Web: Theory and Technology. Morgan & Claypool Publishers, Palo Alto, USA.
- Mikko Koho, Erkki Heino, and Eero Hyvönen. 2016. SPARQL Faceter—Client-side Faceted Search Based on SPARQL. In *Joint Proc. of the 4th International Workshop on Linked Media and the 3rd Developers Hackshop*, number 1615. CEUR Workshop Proceedings.
- Mikko Koho, Eero Hyvönen, Erkki Heino, Jouni Tuominen, Petri Leskinen, and Eetu Mäkelä. 2017. Linked death—representing, publishing, and using Second World War death records as linked open data. In Eva Blomqvist, Katja Hose, Heiko Paulheim, Agnieszka Ławrynowicz, Fabio Ciravegna, and Olaf Hartig, editors, *The Semantic Web: ESWC 2017 Satellite Events*, pages 369–383. Springer-Verlag.
- Mikko Koho, Erkki Heino, Esko Ikkala, Eero Hyvönen, Reijo Nikkilä, Tiia Moilanen, Katri Miettinen, and Pertti Suominen. 2018a. Integrating prisoners of war dataset into the WarSampo linked data infrastructure. In *Proceedings of the Digital Humanities in the Nordic Countries 3rd Conference (DHN 2018)*. CEUR Workshop Proceedings, March. Vol 2084.
- Mikko Koho, Esko Ikkala, Erkki Heino, and Eero Hyvönen. 2018b. Maintaining a linked data cloud and data service for Second World War history. In *Digital Heritage. Progress in Cultural Heritage: Documentation, Preservation, and Protection. 7th International Conference, EuroMed 2018, Nicosia, Cyprus*, volume 11196. Springer-Verlag, October–November.
- Mikko Koho, Lia Gasbarra, Jouni Tuominen, Heikki Rantala, Ilkka Jokipii, and Eero Hyvönen. 2019. AMMO Ontology of Finnish Historical Occupations. In *Proceedings of the The First International Workshop on Open Data and Ontologies for Cultural Heritage (ODOCH'19)*, volume 2375. CEUR Workshop Proceedings, June.
- Petri Leskinen, Mikko Koho, Erkki Heino, Minna Tamper, Esko Ikkala, Jouni Tuominen, Eetu Mäkelä, and Eero Hyvönen. 2017. Modeling and using an actor ontology of Second World War military units and personnel. In *Proceedings of the 16th International Semantic Web Conference (ISWC 2017)*. Springer-Verlag, October.
- Alexander Maedche, Boris Motik, Ljiljana Stojanovic, Rudi Studer, and Raphael Volz. 2003. An infrastructure for searching, reusing and evolving distributed ontologies. In *Proc. of the twelfth international conference on World Wide Web*, pages 439–448. ACM Press.
- Eyal Oren, Renaud Delbru, and Stefan Decker. 2006. Extending faceted navigation for RDF data. In *International semantic web conference*, pages 559–572. Springer-Verlag.

¹⁶http://www.sotavainajat.net/in_english

¹⁷<http://openscience.fi/>

- Daniel Tunkelang. 2009. *Faceted search*. Synthesis lectures on information concepts, retrieval, and services. Morgan & Claypool Publishers.
- Koenraad Verboven, Myriam Carlier, and Jan Dumolyn. 2007. A short manual to the art of prosopography. In *Prosopography approaches and applications. A handbook*, pages 35–70. Unit for Prosopographical Research (Linacre College).
- Christopher Warren, Daniel Shore, Jessica Otis, Lawrence Wang, Mike Finegold, and Cosma Shalizi. 2016. Six degrees of Francis Bacon: A statistical method for reconstructing large historical social networks. *Digital Humanities Quarterly*, 10(3).
- Jun Zhao, Christian Bizer, A Gil, Paolo Missier, and Satya Sahoo. 2010. Provenance requirements for the next version of RDF. In *Proceedings of the W3C Workshop – RDF Next Steps*. W3C. <https://www.w3.org/2009/12/rdf-ws/papers/ws08>.