Extending WarSampo Data Service and Semantic Portal for Digital Humanities Research on Military History

Annastiina Ahola[1,*], Eero Hyvönen[1,2,*], Petri Leskinen[1,2,*] and Heikki Rantala[1,2,*]

[1]*Aalto University, Department of Computer Science, Espoo, Finland*
[2]*University of Helsinki, Helsinki Centre for Digital Humanities (HELDIG), Helsinki, Finland*

Abstract

This paper introduces WarSampo 2.0, a new version of the Linked Open Data service and semantic portal *WarSampo – Finnish WW2 on the Semantic Web* targeted for Digital Humanities research on military history. The new WarSampo 2.0 interface allows for data analyses without programming skills based on an updated and enhanced version of underlying Knowledge Graph in a Linked Open Data service, which is demonstrated through three example cases of how the data and portal can be used for data analysis. To make the data more transparent to the end user, validation results against the metadata model schemas are presented as well as semantic tests made for checking truthfulness of the data with respect to the real world. An agentic AI approach based on the existing data and rules is employed to help propose additional rules to evaluate the truthfulness aspect.

Keywords

user interfaces, semantic portals, data analysis, data validation

1. Introduction

The original WarSampo data service and portal[1] (in Finnish: *Sotasampo*) (Hyvönen et al. 2016) is a widely used service with over a million users. Its Linked Open Data (LOD) in a SPARQL endpoint[2] contains data on World War II–era Finnish military history, covering the Winter War (1939–1940), Continuation War (1941–1944), and Lapland War (1944–1945). Its core data, originating from the National Archives of Finland[3], includes death records of virtually all 95 000 Finnish soldiers killed in action during these wars. This database was enriched with an ontology about the proceedings of the WW2 and was enriched with data aggregated from 13 other data sources, such as 160 000 authentic photographs provided by the Finnish Defense Forces, 3360 memoir articles of veterans published in a post-war magazine, historical placename data and old maps from the National Survey of Finland, biographies from the Finnish Literature Society Finland, an additional database of Finnish prisoners of war in the Soviet Union (Koho, Ikkala, and Hyvönen 2022) as well as additional data about some other 5000 significant Finns related to WW2. The general aim was to reassemble the lives of the fallen soldiers by data linking for relatives and researchers to study, and to promote peace by highlighting the human price of the war.

Published in 2016, the technology behind the portal has become outdated over the years. The portal software used for Sampo portals[4] (Hyvönen 2023) ten years ago, SPARQL Faceter (Koho, Heino, and

DHNB2027: Human(ities) in the Loop, March 01–05, 2027, Umeå, Sweden
*Corresponding author.
†These authors contributed equally.
annastiina.ahola@aalto.fi (A. Ahola); eero.hyvonen@aalto.fi (E. Hyvönen); petri.leskinen@aalto.fi (P. Leskinen); heikki.rantala@aalto.fi (H. Rantala)
https://seco.cs.aalto.fi/u/eahyvone/ (E. Hyvönen)
0009-0008-6369-4712 (A. Ahola); 0000-0003-1695-5840 (E. Hyvönen); 0000-0003-2327-6942 (P. Leskinen); 0000-0002-4716-6564 (H. Rantala)
© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
Digital Humanities in the Nordic and Baltic Countries Publications – ISSN: 2704-1441

[1]Original WarSampo portal: https://www.sotasampo.fi/en/
[2]WarSampo LOD service: https://www.ldf.fi/dataset/warsa
[3]The National Archives of Finland: https://kansallisarkisto.fi/en/
[4]Sampo data services and portal series: https://seco.cs.aalto.fi/applications/sampo/

Hyvönen 2016), did not support data-analytic tools now readily available in the Sampo-UI framework[5] (Ikkala et al. 2022; Rantala et al. 2023). The full potential of using the portal for data-analyses in Digital Humanities is therefore not possible. Furthermore, the underlying Google's AngularJS framework in use in 2015 is no longer supported. These needs for updates—and the jubilee year 2027 of 110 years of independence of Finland—inspired to start the WarSampo 2.0 project, a collaborative project between Aalto University, University of Helsinki, National Archives of Finland, and various Finnish veteran heritage preservation organizations led by Tammenlehvän Perinneliitto.

The technical aim of this work is to update the user interface (UI) and underlying software of the portal using Sampo-UI. At the same time, hundreds of suggestions for updates and corrections in the underlying death records will be checked and edited when appropriate by researchers at the National Archives of Finland. These suggestions have been collected from the public through the feedback channel of WarSampo. Furthermore, the original the war cemetery data and application perspective (Ikkala et al. 2017) of WarSampo covered only contemporary Finland and will be extended for cemeteries of the Karelia area that was annexed to the Soviet Union. Finally, a case study on data quality issues of linked data and will be made based on the WarSampo KG (Koho et al. 2021).

The project will publish the WarSampo 2.0 portal before December 6th, 2027 to commemorate Finland's 110th independence day.

2. Related work

The landscape of applying semantic web technologies to publishing and studying the World Wars and military history has been growing with numerous international projects. Data on the First World War has been published as part of various projects such as 1914-1918-Online[6], Europeana Collections 1914-1918[7], Out of the Trenches (Pan-Canadian Documentary Heritage Network (PCDHN) 2018), CENDARI[8], Muninn project[9] and WW1LOD (Lindquist et al. 2013; Mäkelä et al. 2017). In a Finnish context, WarVictimSampo (Rantala et al. 2022) focuses on the Finnish Civil War in 1918, but includes war victim data also from the Finnish Kindred Wars in 1914–1922.

Data related to the Second World War has been part of various projects and archives as well. Collins, Mulholland, and Zdrahal (2005) created a system to browse Bletchley Park's WW2-era historical accounts and interviews. Boer et al. (2013) digitized and enriched a Dutch WW2 historiography. The EHRI project (García-González and Bryant 2023) deals with trans-national Holocaust-related archival material and LOD Navigator (Sprugnoli, Moretti, and Tonelli 2019) is a tool created to explore and visualize data related to Italian victims of the Holocaust. In the Netherlands, archives such as Oorlogsbronnen[10] and Oorlog voor de Rechter[11], both part of the WO2Net[12], present Dutch archival data from WW2. In Germany, data from the archive Wiedergutmachung nationalsozialistischen Unrechts[13] dealing with claims by victims of National Socialist persecution is being transformed into LOD (Vafaie 2024; Yargan et al. 2026). In addition to WarSampo in Finland, WarMemoirSampo (Koho et al. 2022) also deals with WW2 and contains interview videos of Finnish WW2 veterans.

The application of LOD is not limited to the World Wars and 20th century. For example, outside of the World Wars, linked data has also been used with American Civil War data in projects such as Builders & Defenders (Sutton and Power 2024) and Civil War Data 150[14]. On the whole, these kinds

[5] Sampo-UI Framefork: https://seco.cs.aalto.fi/tools/sampo-ui/
[6] 1914-1918-Online. International Encyclopedia of the First World War (WW1): https://encyclopedia.1914-1918-online.net/
[7] Europeana Collections 1914-1918 | Remembering the First World War – a digital collection of outstanding sources from European national libraries: http://www.europeana-collections-1914-1918.eu/
[8] CENDARI: http://www.cendari.eu/about-cendari
[9] The Muninn Project: https://blog.muninn-project.org/
[10] Oorlogsbronnen | Personen, bronnen en verhalen uit de Tweede Wereldoorlog: https://www.oorlogsbronnen.nl/
[11] Oorlog voor de Rechter: https://oorlogvoorderechter.nl/
[12] WO2Net: https://www.wo2net.nl/
[13] Wiedergutmachung - Archivportal-D: https://www.archivportal-d.de/themenportale/wiedergutmachung
[14] Civil War Data 150 | Linking Civil War Data Across State and Federal Archives and Libraries: https://www.civilwardata150.net/

of projects demonstrate the importance of making military history and the human cost of war more accessible both to researchers and the general public.

3. Data

The original data[15,16] behind the original WarSampo portal consists of more than 14 000 000 triples hosted on an Apache Jena Fuseki[17] triplestore. It contains instances of various classes such as events, persons, military units, places, articles from the *Kansa taisteli* magazine, death records of casualties, photographs, cemeteries, and prisoner records of prisoners of war related to the Winter War (1939–1940), the Continuation War (1941–1944), and the Lapland War (1944–1945). Most of the persons (102 613 in total) included in the data are either casualties (94 677 instances of death records) or prisoners (4200 instances of prisoner records) of war (or both), but around 5000 additional notable individuals were also added to the data from other sources (Leskinen et al. 2017). More than 166 000 photographs included in the data document both the events of the war and the nearly 700 cemeteries in which casualties have been buried.

The data model of WarSampo (Koho et al. 2021) is based on events and uses the CIDOC Conceptual Reference Model (CRM)[18] (Doerr 2003) with extensions to cover for cases where the pure CRM approach is not feasible due to limitations of the source data. The data on casualties and prisoners of war is stored in *death record* and *prisoner record* documents that document personal information about the person at the time of their death or capture, respectively. This includes biographical information (e.g., municipalities of birth, nationality, date of birth, and so on), as well as information on the person's military service information (e.g., military rank and unit). The WarSampo Metadata Schema[19] defines the classes and properties used in the dataset as well as the hierarchies between them. This original data serves as the basis for the new expanded WarSampo 2.0 KG including the new additions described in the next section.

4. WarSampo 2.0 portal and data service

The new UI for the WarSampo 2.0 portal is built using the Sampo-UI framework (Ikkala et al. 2022; Rantala et al. 2023). Sampo-UI is based on the Sampo model (Hyvönen 2023) and can be used to build semantic portals on top of virtually any SPARQL endpoint, regardless of the domain of the data. So far, it has been used to build more than 25 semantic portals both domestically and internationally[20], covering different domains such as numismatic coin finds (Rantala et al. 2024; Six et al. 2026), letters (Hyvönen, Leskinen, and Tuominen 2023; Hyvönen et al. 2026), and parliamentary data (Hyvönen et al. 2025). WarSampo 2.0 portal uses the latest major v4 version of Sampo-UI that has received updates to its libraries and a new and more easily maintainable architecture (Six et al. 2026).

While the original WarSampo portal uses an Apache Jena Fuseki triplestore to get its data, the data was ported over to a QLever triplestore (Bast and Buchhold 2017) for WarSampo 2.0 to enable faster queries in the UI. The original data is supplemented with (1) missing cemeteries from the annexed Karelia region now in Russia, and (2) new automatically created links related to casualties.

The WarSampo 2.0 portal re-implements three *application perspectives* from the original WarSampo portal—the other six perspectives will be integrated as they are in new portal version—thus far:

1. *Casualties.* This perspective contains information on all the casualties from the wars. This also includes some casualties who died after the wars, including a few prisoners of war in the Soviet Union who died later in captivity as well as war veterans whose reasons of death were not caused

[15] WarSampo dataset page on Linked Data Finland: https://www.ldf.fi/dataset/warsa
[16] WarSampo SPARQL endpoint: https://ldf.fi/warsa/sparql
[17] Apache Jena Fuseki: https://jena.apache.org/documentation/fuseki2/
[18] CIDOC CRM: https://cidoc-crm.org/
[19] WarSampo Metadata Schema: http://ldf.fi/schema/warsa/
[20] Sampo Model page with a list of different portals using Sampo-UI: https://seco.cs.aalto.fi/applications/sampo/

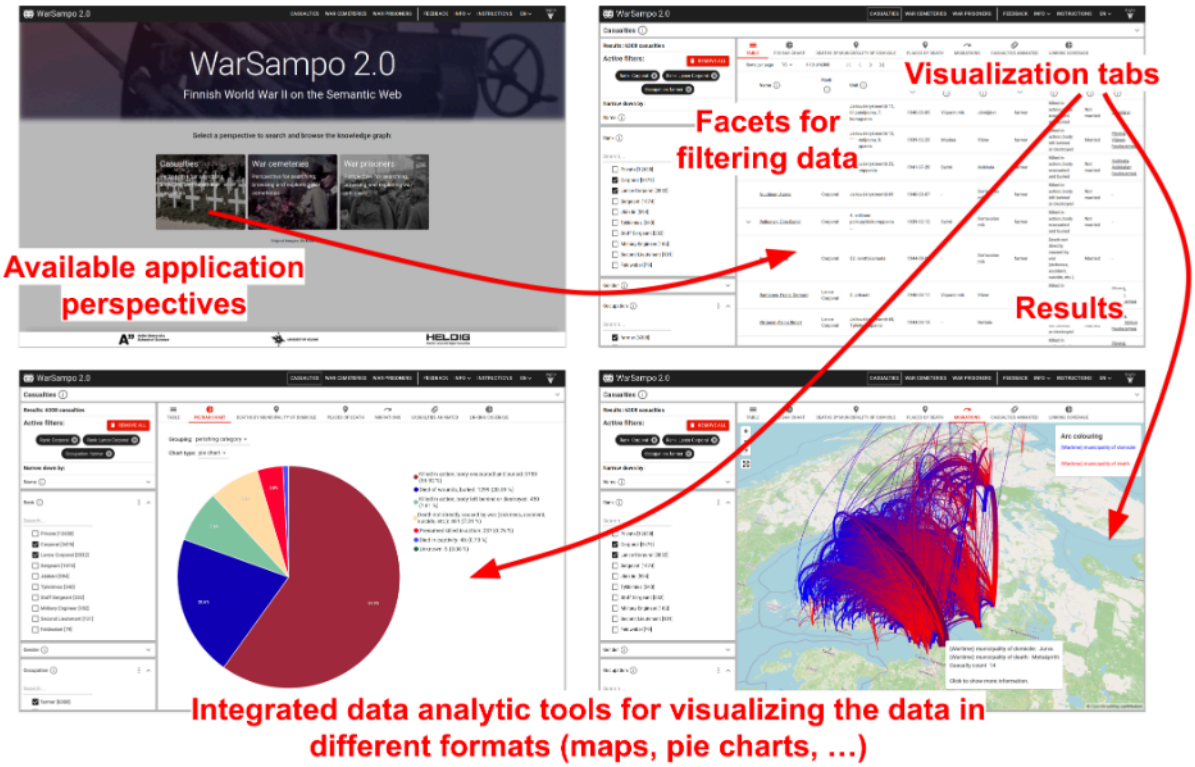


Figure 1: The user flow of using the WarSampo 2.0 portal to filter and visualize data

by direct enemy action (illness, accident, suicide, etc.). The information in this perspective is based on the death records of the original casualties database.

2. *War cemeteries.* This perspective contains information on all the different cemeteries in the data as well as who has been buried there. The information in this perspective is based on the original casualties database's cemetery records as well as photographs taken by hundreds of members of Finnish photography clubs who have documented the cemeteries in our earlier project (Ikkala et al. 2017). The new additions to the cemeteries data from the Karelia region are based on data from the Karjalanhaudat project[21].
3. *War prisoners.* This perspective contains information on the Finnish prisoners of war held captive in the Soviet Union. The information in this perspective is based on a dataset that was originally published as a book (Alava et al. 2002) and waS integrated as a new subgraoph into the WarSampo KG (Koho, Ikkala, and Hyvönen 2022).

Opening up an application perspective brings up a *faceted search view*, where the user can browse and filter the data (e.g., individual casualties in the *casualties* perspective) through *faceted search* (Hearst et al. 2002; Tunkelang 2009) as well as analyze the result sets using the different integrated data analytic tools. Clicking on the name of an individual resource brings up its *instance page*, where the user can view detailed information about that particular instance and use additional data-analytic tools.

Figure 1 shows an example user flow of the portal to filter and visualize. In the example, the user chooses to enter the *casualties* perspective (top left) and filters the data (top right) to only those casualties with the rank of *Corporal* or *Lance Corporal* (from the *Rank* facet) and the occupation of *farmer* (from the *Occupation* facet). The user can then visualize the filtered result set with the different visualization tabs available, such as the pie chart option to visualize perishing categories (bottom left) or the migrations visualization (bottom right), where the blue ends of the arcs indicate the municipalities of domicile, the red ends the municipalities of death, and the width the number of casualties with this combination of municipalities out of the filtered result set.

[21] Karjalanhaudat – Luovutetun Karjalan sankarihautausmaat: https://karjalansankarihautausmaat.fi/

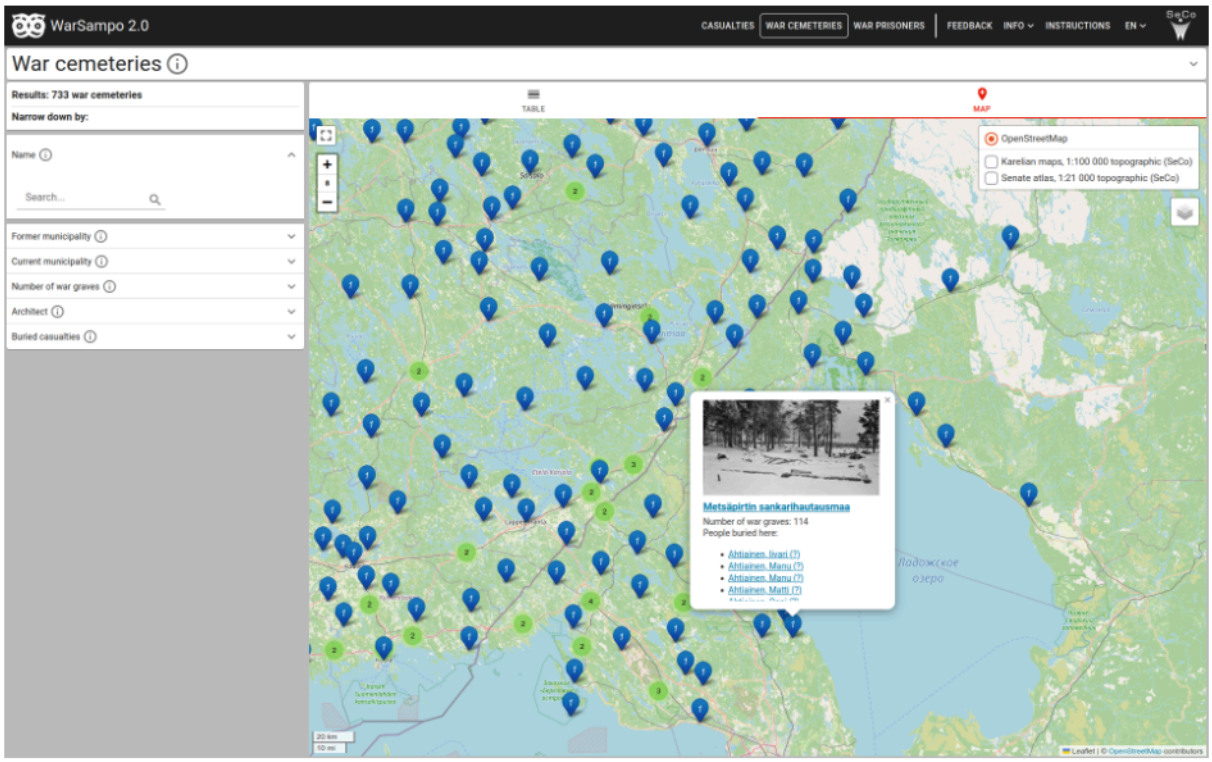


Figure 2: A map visualization showing the locations of war cemeteries with the new additions

4.1. New data additions

As part of WarSampo 2.0, there were two new additions made to the original data: (1) cemeteries located in the ceded Karelian areas not included in the original WarSampo data were added and linked to existing casualties, and (2) casualties were also automatically linked to other external digital sources.

4.1.1. Cemeteries in the ceded Karelian areas

The data on new cemeteries encompasses 61 war and field cemeteries from the ceded Karelian areas that were not included in the original casualties data. It has been compiled starting 2020 as a part of Karjalanhaudat project by Sotavainajien muiston vaalimisyhdistys[22] and Karjalan Liitto[23]. The data includes basic information about the cemeteries, such as descriptions, locations, and information on memorials, as well as both old and new images if available. After conversion, the new cemeteries were added as a part of the *cemeteries* perspective where they can be viewed alongside the originally included cemeteries as shown in Figure 2.

Data about who are buried in these different cemeteries can be extracted from the death records and from available documents included in the new cemetery source data. The Python package Splink[24] (Linacre et al. 2022) was used to automatically link casualties mentioned in the lists of buried casualties for the newly added cemeteries to existing WarSampo person instances. The amount of data recorded for the casualties differed between the documents. At minimum, the lists included the last and given names of the identified casualties, while some also had additional information such as the dates (or years) of birth and death, military unit and rank, place of death, and so on. After normalizing the names and dates between WarSampo and the data incoming from the documents, the Splink model was trained to predict links between persons based on available data for the following fields:

1. last name

[22]Sotavainajat – Sotavainajien muiston vaalimisyhdistys vuodesta 1998: https://sotavainajat.net/
[23]Karjalaisuuden ja karjalaisen kulttuurin yhteisö - Karjalan Liitto: https://www.karjalanliitto.fi/
[24]Splink: https://moj-analytical-services.github.io/splink/



Figure 3: An automatically linked cemetery of burial on a person's instance page

2. given name(s)
3. date of birth
4. date of death
5. year of birth (extracted from the exact date of birth for WarSampo data)
6. year of death (extracted from the exact date of death for WarSampo data)
7. municipality of the cemetery (burial lists) ↔ municipality of burial (WarSampo data)

The WarSampo person instances in the process were limited to those persons with no pre-existing cemetery of burial in the data (10 367 instances). For the incoming data, a total of 5850 persons and their information was extracted from the burial lists of the newly added cemeteries, covering 41 cemeteries. The Splink model trained on this data was able to create 5094 links between these persons and WarSampo person instances with high confidence. Due to there being no one-to-one requirement during linking and there being people with similar names and other biographical information within singular burial listings, the total amount of new person instance–cemetery links ended up at 5049.

Although the cemeteries from the ceded Karelian areas were not modeled in the data for the original WarSampo portal, the original casualties data had numerical values referring to specific cemeteries within Karelian municipalities but without the labels for indicating which number corresponded to which cemetery. With the burial lists, some of these numbers could be connected back to the names of the cemeteries based on shared casualties between sources and further links can be added based on these connections. This increased the total number of new person instance–cemetery links to 5660. These links were then used to add the corresponding cemetery as the cemetery of burial to the person instances in the WarSampo KG, making them also visible in the UI. As these links were created automatically and there are contradictions (different records indicating multiple different cemeteries of burial), there is a notice in the UI to inform the user of their unverified status as seen in Figure 3 and the list of buried casualties visible in Figure 2.

4.1.2. New links to external sources

The new automatic links from casualties to external sources were added to digitized military service cards[25] as well as to person pages on the Sotapolku.fi service[26], a citizen science portal for collecting information about WW2 Finnish soldiers and mapping their their paths during the war. The matching between people in WarSampo and Sotapolku was performed based on the titles of the archival units available from the digitized service cards of the National Archives which use the format *"[surname] [first name(s)], [date of birth as DD.MM.YYYY]"*. There were two cases in which two different people had the exact same full name and date of birth, in which case the correct links were manually checked based on other information available on the casualties in WarSampo data and the service cards, respectively. In the end, 93 290 out of the 94 677 casualties (98.54%) could be automatically linked to a digitized service card, leaving only 1387 without these links. For the Sotapolku.fi person page links, information

[25]Digitized military service cards on the National Archives' Astia access service: https://astia.narc.fi/uusiastia/kortti_aineisto.html?id=2661703739
[26]Sotapolku portal: https://sotapolku.fi/

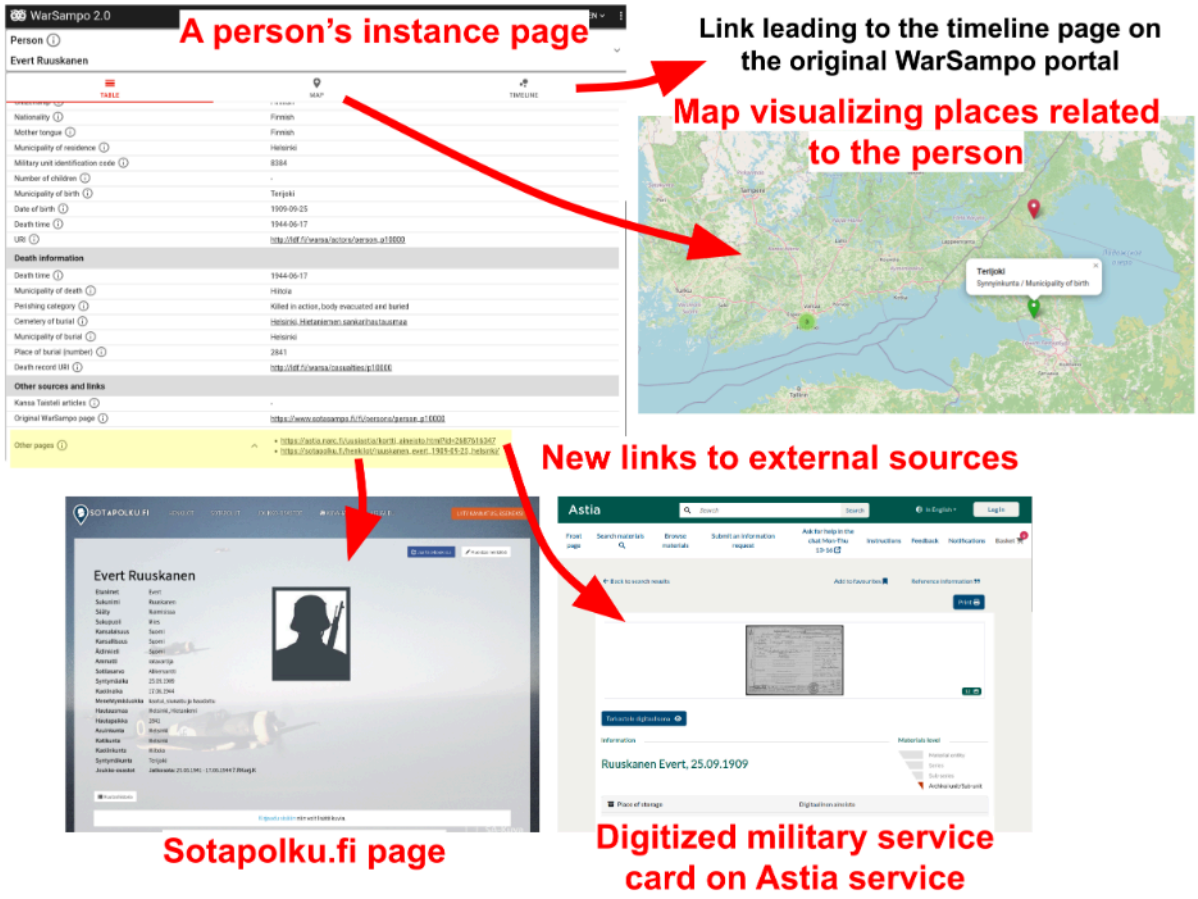


Figure 4: A person's instance page with its visualizations and external links

on the municipality of domicile, rank as well as unit was used to attempt to match as many casualties as possible with their possible Sotapolku.fi pages by comparing it to the information shown for each person in the search result tables on Sotapolku.fi. This process led to 93 372 out of 94 677 casualties (98.62%) being linked, with 1305 casualties having no link to a Sotapolku.fi page. For both of these cases, exact string matching was used for the names, and extending the linking process in the future to also attempt an approximate string matching could help link the remaining casualties to both external sites. The links to these external sites are shown to the user on a person's instance page (see Figure 4).

5. Using the data for data analysis

WarSampo 2.0 portal allows anyone to visualize the data within it using its integrated data analytic tools. This allows a non-technical user to view the data in different formats directly in the portal without the need for programming or SPARQL querying skills. This section presents three brief example cases to demonstrate how the portal can be used for data analysis.

5.1. The most common ranks of casualties

The data on the casualties of war includes information on their rank for the vast majority of casualties, with only ca. 1100 death records missing this information. This makes it possible to analyze, e.g., what the most common ranks of the casualties were in different times. The user can visualize this information in the *casualties* perspective by simply navigating to the *pie/bar chart* tab and changing the grouping to *rank* instead of the default *perishing categories*. The format can be changed from the default pie chart to a bar chart by selecting that as the chart type in the UI. Figure 5 illustrates the most common ranks as a

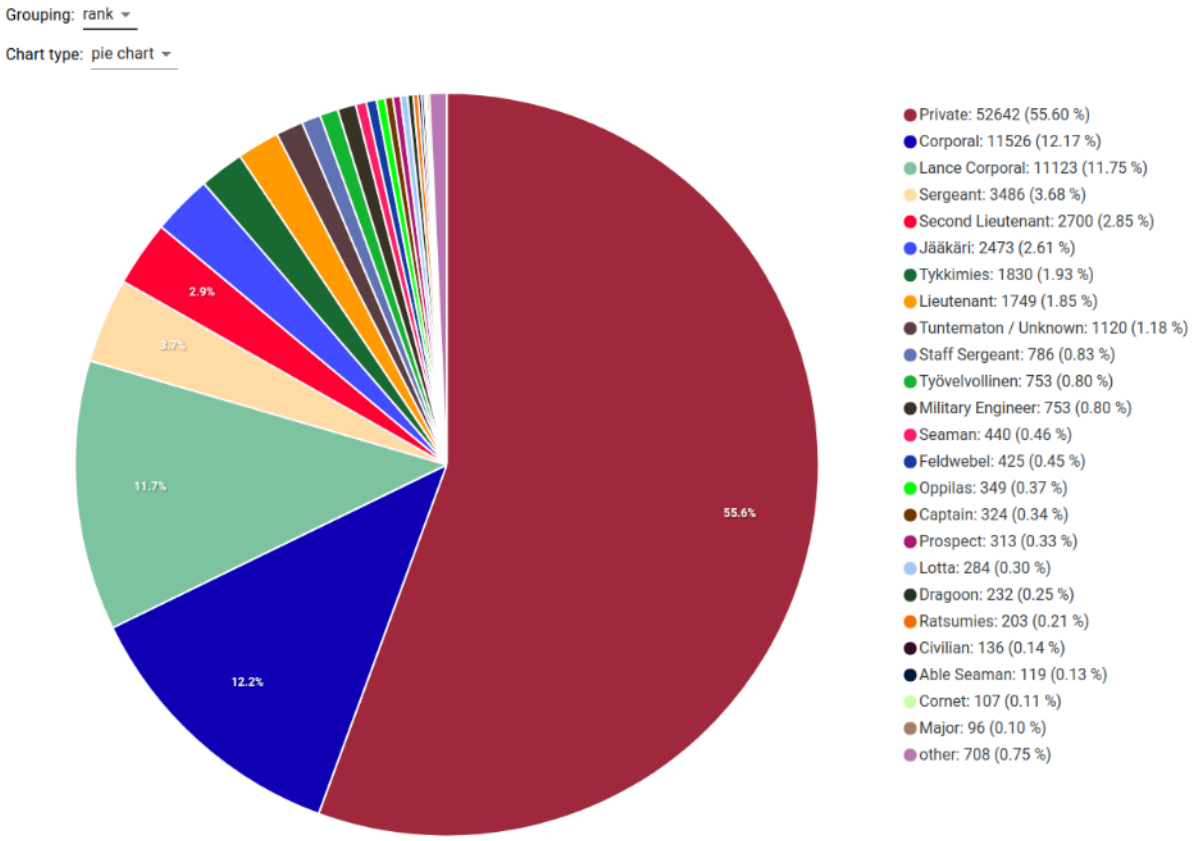


Figure 5: Casualties visualized as a pie chart grouped by their rank

pie chart taken from the UI.

5.2. Municipalities of domicile and their correlation to other factors

The data for casualties also includes information on the municipalities of birth, domicile, death, and so on. The coverage varies with, for example, the municipality of domicile missing for around 2000 casualties while the municipality of death is missing for nearly 53 000 casualties. Still, these pieces of information can be used to analyze those casualties that have the information.

One visualization related to these is the choropleth map available on the *casualties* perspective in the tab *deaths by municipality of domicile*. It visualizes the absolute number of casualties grouped by their municipality of domicile on the map. The darker the color, the higher the number of casualties with this municipality of domicile recorded. Hovering over a municipality will show its name, as well as the number of casualties. Making any selections in the facet will update the map to reflect the new data.

On the left, Figure 6 shows the choropleth map visualizing the number of casualties per each municipality of domicile in the perspective with no facet selections made. On the right side is an updated map with the municipality of *Salla* (a municipality located in Lapland, northern Finland) having been selected as the domicile of death in the facets. The differing colors show at a glance that the casualties who perished in Salla tend to also have domiciled either in Salla or other municipalities in northern Finland, instead of the domiciles being spread more evenly across Finland. Instead of Salla, another municipality or even other pieces of information altogether (e.g., casualties with specific ranks and/or occupation) could be used to view the differences and trends between different subsets of casualties.

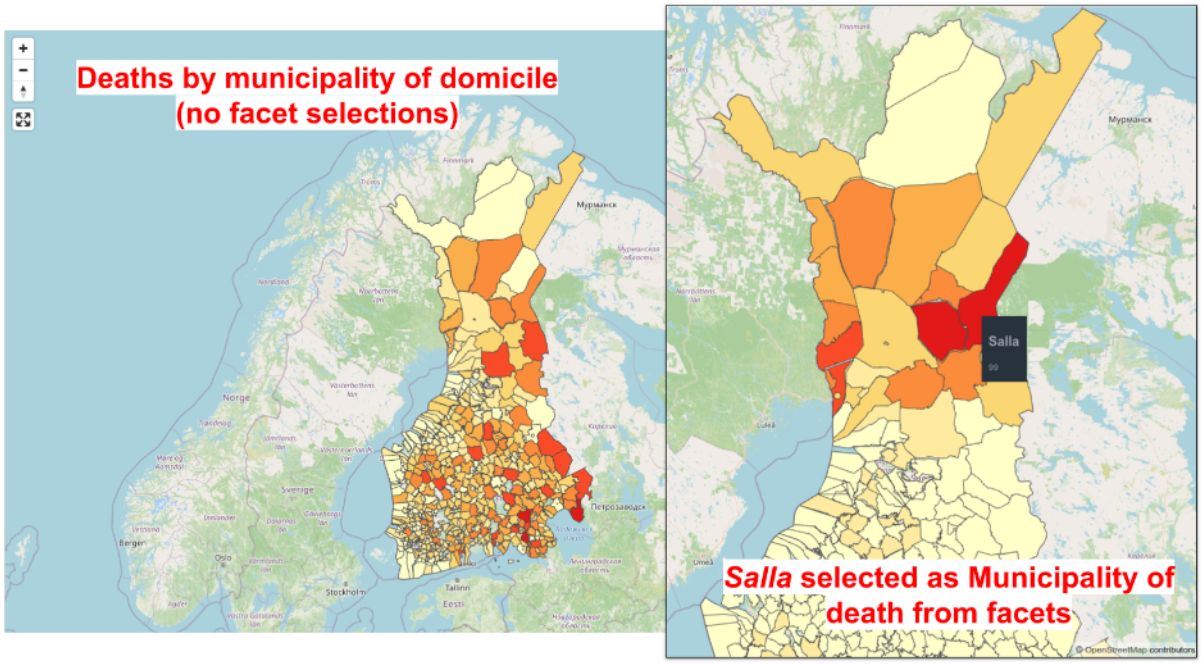


Figure 6: Absolute number of casualties by the municipality of domicile visualized as a choropleth map (underlying map tiles © OpenStreetMap contributors)

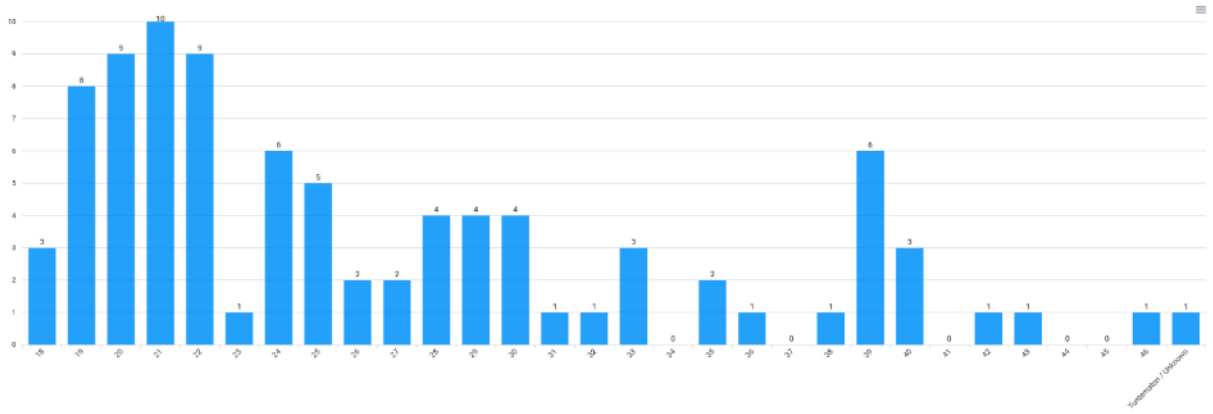


Figure 7: Ages of casualties buried in *Koiviston sankarihautausmaa* (Koivisto War Cemetery) visualized as a bar chart

5.3. Casualties and their ages at the time of death

The user could also be interested in the cemeteries and the casualties buried in a certain cemetery, for example, the user's local cemetery. In the *cemeteries* perspective, the instance pages of individual cemeteries offer different visualizations about the casualties buried there. Since the death records record both dates of birth and death when available, it is possible to calculate the ages of death for many of the casualties and offer visualizations of this information in the UI.

Figure 7 illustrates the age distribution of casualties at the time of their death that have been buried in *Koiviston sankarihautausmaa* (Koivisto War Cemetery), a cemetery located in the town of Primorsk (in Finnish *Koivisto*) in the ceded Karelian area, taken from the *ages of buried casualties* tab on the cemetery's instance page. The three youngest casualties buried there were 19 years old at the time of their death, while the older was 46 years old. One's age is marked as unknown, which is the grouping used for those casualties where either date of birth and/or death is completely or partially missing (e.g., only the year is known). This visualization directly utilizes the new data additions of cemeteries from the ceded Karelian area as well as the data from the automatic linking of casualties to these cemeteries

and thus would not have been possible with the original WarSampo data.

For more complex and customized queries, more technical users can always utilize the open WarSampo data service to query and extract the data they need and then use tools such as Python visualization libraries to create their own visualizations, as done by Koho, Rantala, and Hyvönen (2022). As shown, the integrated data analytic tools in the WarSampo 2.0 portal enable the user to easily and quickly explore and visualize different subsets of data with only a few clicks. The user does not necessarily have to know what they are looking for in advance, but rather can serendipitously come across interesting facts and trends while trying out different combinations of filters in different perspectives.

6. Data validation

To validate the data, two approaches are employed: (1) user feedback on data issues, and (2) Shapes Constraint Language (SHACL)[27] constraints–based validation.

6.1. User feedback

With the original data in the database being based on documents available at the time of compiling the database, there were bound to be inaccuracies in the data contents itself stemming both from inaccuracies in the original documents as well as mistakes made during the digitizing process (e.g., typos in names or dates). The original WarSampo portal includes a feedback page that the user can use to send information on data inaccuracies or missing data. Many of these corrections involve the data on casualties, where the user is often a relative of the casualty. This feedback data is passed onto the National Archives of Finland, who investigate the reports and make corrections if necessary. This process allows some level of data content validation to be performed that cannot be done using rules like in the next subsection.

6.2. SHACL-based validation

The Shapes Constraint Language SHACL[28] is the language recommendation of W3C for validating RDF graphs against a set of conditions. SHACL-based validation can be used to validate the WarSampo data stored in the endpoint against its defined schema, as well as custom rules written for the dataset to catch errors not explicitly defined in the schema itself (e.g., a start date of something should be before the end date in time). ShEx - Shape Expressions[29] could also be used to validate the shapes in the data against a schema, but SHACL is a better fit for this use case due to its support for SPARQL-based constraints (Gayo et al. 2017, Ch. 7) with SHACL-SPARQL. Additionally, SHACL validation can be run easily with Fuseki through Apache Jena SHACL[30], allowing for an easier validation process via a Fuseki triplestore container.

6.2.1. Schema-based SHACL rules

The foundation for the SHACL rules are derived from the WarSampo schema. A Python script using `RDFLib`[31] generates SHACL rules based on the property types (e.g., OWL's[32] `ObjectProperty` should be getting IRI values or `DatatypeProperty` should be getting literals) as well as domains and ranges if specified. A SPARQL constraint is used for checking class constraints to account for cases where a subclass of a specified range is provided instead of the specified class itself. In cases where the property type and range conflict (e.g., `DatatypeProperty` but the range is a class), the range is treated as the truth and a warning of this conflict is printed on the console running the Python script.

[27] Shapes Constraint Language (SHACL): https://www.w3.org/TR/shacl/
[28] SHACL: https://www.w3.org/TR/shacl/
[29] ShEx - Shape Expressions: https://shex.io/
[30] Apache Jena SHACL: https://jena.apache.org/documentation/shacl/
[31] RDFLib: https://rdflib.readthedocs.io/
[32] OWL Web Ontology Language: https://www.w3.org/TR/owl-features/

Table 1
Validation report results for rules generated from the WarSampo schema

(a) Initial validation results for schema-based rules

Constraint component type	Count
Class (including subclass) constraint	153 633
Datatype constraint	191
Less than constraint	0
Pattern constraint	0
SPARQL constraint	0
Unique language constraint	0
Severity level	**Count**
Violation	153 824
Warning	0
Information	0

(b) Validation results after adjusting schema

Constraint component type	Count
Class (including subclass) constraint	5199
Datatype constraint	191
Less than constraint	9
Pattern constraint	0
SPARQL constraint	0
Unique language constraint	0
Severity level	**Count**
Violation	5390
Warning	0
Information	0

Table 2
Validation report results after adding additional rules

(a) Validation results after adding some event-related rules

Constraint component type	Count
Class (including subclass) constraint	5199
Datatype constraint	191
Less than constraint	9
Pattern constraint	3
SPARQL constraint	28
Unique language constraint	1645
Severity level	**Count**
Violation	7073
Warning	0
Information	2

(b) Validation results after generalizing two of the new event-related rules

Constraint component type	Count
Class (including subclass) constraint	5199
Datatype constraint	191
Less than constraint	9
Pattern constraint	3
SPARQL constraint	30
Unique language constraint	1647
Severity level	**Count**
Violation	7077
Warning	0
Information	2

Validating the dataset against these schema-based constraints returns a huge validation report with 153 824 violations (see Table 1a), the vast majority of them (153 633) being class constraint violations and 191 of them being datatype constraint violations. The number of violations being extremely high for specific properties (e.g., a singular property having 125 288 violations) indicated clear issues with the schema being partially outdated or missing information. The schema was adjusted by:

1. adding missing class definitions (e.g., the schema referenced CIDOC CRM `E1 CRM Entity` class in its ranges but did not include its definition as a class) and their subclass hierarchies
2. fixing typos (e.g., `E2 Temporal Enitity` to `E2 Temporal Entity`)
3. replacing incorrect classes as range and domain values with the ones used in the actual data
4. removing the range specification for properties whose values are just external IRI references and not included in the actual WarSampo dataset (and whose types cannot thus be validated from within the dataset)

This reduced the number of class constraint violations down to 5199, leading to a total of 5390 violations (see Table 1b).

6.2.2. Expanding rules to validate things outside of the schema

Kourijoki (2020, Appendix E) created additional SHACL-based rules to validate some event-related parts of the WarSampo data:

1. preferred labels based on the integrity condition S14 of SKOS[33] (one preferred label per language)
2. related links for military and political being to the YSA vocabulary[34] (with severity being set to information instead of violation)
3. aircraft label pattern for battles, (4) start times being before time ends for battles
4. start and end times being of type xsd[35]:time
5. there being no empty string literals as values for properties of battles
6. the actor carrying out a battle should also be a participant in the battle

Notably, some of these rules are validating the actual contents of the values given to properties rather than just the types of the values by, e.g., enforcing specific patterns that aircraft labels should follow and setting chronological constraints.

These rules were integrated on top of the schema-based rules created. This lead to the total number of violations increasing to 7073 as well as two information-level reports (see Table 2a). The rules for preferred labels and empty string literal values lent themselves well for generalized versions that could apply to the dataset at large. They were generalized to apply to all preferred labels and all properties, respectively. This led to only a total of four new violations, two from empty string values and two from duplicate preferred labels per language (see Table 2b).

6.2.3. Letting an agent propose new rules

The rules written by Kourijoki (2020, Appendix E) were aimed at validating a very specific part of the dataset and are specific to the types of events they are validating. Other parts of the dataset would also benefit from similarly tailored rules, especially rules tailored to check the factual correctness of the data in respect to the real world, which goes beyond what the schema can automatically give in terms of generated rules. This kind of factual correctness checking is part of the 8-star model for LOD publishing (Hyvönen and Tuominen 2024) extending the traditional 5-star model[36] of Tim Berners-Lee for linked data quality: 6th star is earned by publishing not only data but also its schema, 7th star is obtained by showing that the data actually conforms to the schema, and 8th star is given for truthfulness of the data with respect to the real wold.

The task of generating these rules was given to an AI agent, which had the access to both the data to be validated as well as previous iterations of the SHACL rules. The model used was the open-weights model `zai-org/GLM-5.3-Flash`[37] (Zeng et al. 2026). By having access to the data, the agent could discover related properties and instances whose values could be cross-referenced against each other to spot factual mistakes (e.g., a person cannot have been wounded after their death) as well as spotting properties with limited possible values (e.g., a total of 191 Mannerheim Crosses were awarded, so only cross numbers between 1...191 can be correct). The output from the model is a set of rules, so the human-in-the-loop can choose to only include some of the rules and reject others, or alternatively adjust the proposed rules based on domain knowledge the AI agent is missing.

Table 3 lists all the rules proposed by the agent — all sensible for the domain. These new rules led to the number of violations increasing to 7235 with also nearly 1000 new warnings and over 30 new information information-severity validation results (see Table 4a). As shown in Table 4b, the majority of the warnings (889 warnings) came from the rule regarding events overlapping with their related

[33] SKOS Simple Knowledge Organization System: https://www.w3.org/TR/skos-reference/
[34] YSA – Yleinen suomalainen asiasanasto: https://finto.fi/ysa/en/
[35] `xsd` prefix is used here for the namespace <http://www.w3.org/2001/XMLSchema#>
[36] 5-star model: https://www.w3.org/2011/gld/wiki/5_Star_Linked_Data
[37] zai-org/GLM-5.3-Flash on Hugging Face: https://huggingface.co/zai-org/GLM-5.3-Flash
[38] `wsch` prefix is used here for the namespace <http://ldf.fi/schema/warsa/>
[39] `ps` prefix is used here for the namespace <http://ldf.fi/schema/warsa/prisoners/>
[40] `crm` prefix is used here for the namespace <http://www.cidoc-crm.org/cidoc-crm/>
[41] `wevs` prefix is used here for the namespace <http://ldf.fi/schema/warsa/events/>
[42] `org` prefix is used here for the namespace <http://rdf.muninn-project.org/ontologies/organization#>
[43] `wacs` prefix is used here for the namespace <http://ldf.fi/schema/warsa/actors/>
[44] `wgs84` prefix is used here for the namespace <http://www.w3.org/2003/01/geo/wgs84_pos#>

Table 3
The rule proposals made by the AI agent

#	Rule	Target	Severity	Checks
Death records				
1.1	birth before death	wsch[38]:DeathRecord	Violation	wsch:date_of_death ≥ wsch:date_of_birth
1.2	wounding before death	wsch:DeathRecord	Violation	wsch:date_of_wounding ≤ wsch:date_of_death
1.3	MIA before death	wsch:DeathRecord	Warning	wsch:date_of_going_mia ≤ wsch:date_of_death
1.4	death period	wsch:DeathRecord	Info	wsch:date_of_death within 1939-11-30 … 2000-12-31
1.5	birth period	wsch:DeathRecord	Info	wsch:date_of_birth within 1870-01-01 … 1939-11-29
Prisoner records				
2.1	birth before death	wsch:PrisonerRecord	Violation	ps[39]:date_of_death ≥ wsch:date_of_birth
2.2	capture before return	wsch:PrisonerRecord	Violation	ps:date_of_return ≥ ps:date_of_capture
2.3	capture before death	wsch:PrisonerRecord	Violation	ps:date_of_death ≥ ps:date_of_capture
2.4	declaration of death after death / MIA	wsch:PrisonerRecord	Warning	ps:date_of_declaration_of_death ≥ death / MIA date
Captivity				
3.1	begin before end	ps:Captivity	Violation	ps:date_end ≥ ps:date_begin
3.2	within capture and return	subjects of ps:captivity	Warning	captivity interval inside capture…return window
Event timings				
4.1	(timespan) begin before end	crm[40]:E52_Time-Span	Violation	crm:P82a_begin_of_the_begin ≤ crm:P82b_end_of_the_end
4.2	person's birth before involved event end	person-event classes	Violation	event end ≥ birth of the involved person
4.3	person's death after involved event start	person-event classes	Violation	event begin ≤ death of the involved person
4.4	event's related period matches	subjects of wevs[41]:related_period	Warning	event overlaps its related war period
Units				
5.1	dissolution after formation	wsch:Dissolution	Warning	dissolution begins after a formation of the same unit
Ranks				
6.1	no self-seniority	wsch:Rank	Violation	no org[42]:rankSeniorTo / org:equalTo self-loop
6.2	no mutual seniority	wsch:Rank	Violation	seniority is antisymmetric
Cemeteries				
7.1	unveiling after foundation	wsch:Cemetery	Warning	memorial unveiled in/after the foundation year
Ranges				
8.1	(knight) number range	subjects of wacs[43]:knight	Violation	Mannerheim Cross number in 1…191
8.2	(coordinates) latitude / longitude range	subjects of wgs84[44]:lat and wgs84:long	Violation	latitude ∈ [-90, 90], longitude ∈ [-180, 180]

periods, while the most common new violation came from a person being involved in an event after their death (101 violations). These stem from, for example, conflicting information from casualty and war prisoner records: The death record might have an earlier death date for a casualty that matches the day they disappeared with the perishing category *Presumed killed in action*, while the prisoner records list a later, actual date of death. This leads to a situation where the captivity event starts after one

Table 4
Validation report results after integrating the AI agent's proposed rules

(a) Validation report results by constraint type and severity level

Constraint component type	Count
Class (including subclass) constraint	5199
Datatype constraint	191
Less than constraint	9
Pattern constraint	3
SPARQL constraint	1186
Unique language constraint	1647

Severity level	Count
Violation	7235
Warning	965
Information	35

(b) New results by rule

Rule		Severity	Count
Death records	1.3	Warning	14
	1.4	Info	32
	1.5	Info	1
Prisoner records	2.2	Violation	3
	2.3	Violation	7
Captivity	3.1	Violation	16
	3.2	Warning	52
Event timings	4.1	Violation	26
	4.2	Violation	5
	4.3	Violation	101
	4.4	Warning	889
Units	5.1	Warning	2
Cemeteries	7.1	Warning	8

of the listed dates of death. Multiple different dates of death for a single person also caused some of these violations, with one death event taking place after the first date of death. Even if not all of the new violations lead to corrections being made (e.g., a medal being posthumously awarded triggering a violation under Rule 4.3 or multiple captivity periods triggering Rule 2.2), they help pinpoint data instances that need verification from the total of more than 1 800 000 instances of different classes in the dataset.

The results as a whole demonstrate the importance of data validation for the trustworthiness and the actual truthfulness of the data contents. Even though the original WarSampo data is part of a published service and has an available schema, the data validation process exposed issues in both the schema definition and the actual existing data—many of which are easy to remedy as long as the data maintainer is aware of them. For example, 4519 of the class (or subclass) constraint violations stem from the unit categories in the data missing their intended class definition, which is actually caused by a mismatch in the namespace: In the data defining the unit categories, they are defined to be in the `<http://ldf.fi/warsa/unitCategories/>` namespace, while military unit instances try to link to unit categories within the `<http://ldf.fi/warsa/actors/>` namespace. This has led to the unit categories information not being correctly linked and thus missing the correct class definition as well. Similarly, the 191 datatype constraint violations can be fixed by casting the values as the intended integers instead of storing the values (consisting of integers) as string values.

The application of the agentic AI approach to generating SHACL rules also demonstrated the approach's viability for moving towards data fulfilling the conditions for the 8th star in the 8-star model for LOD publishing, which requires rules to be tailored to the datasets in such a way that generating them just based on a schema automatically is difficult. The agent was able to come up with more than 20 rules to cover multiple different parts of the dataset that led to more than 150 new violations for verification as well as a little under a thousand new warnings and more than 30 information-level validation results. As the agent provided the rules themselves instead of the validation results directly, the validation process can be rerun again if needed and the process and its results remains explainable. If needed, the human can easily modify or exclude any rules they deem unnecessary or otherwise problematic, being ultimately in the charge of the validation process.

7. Discussion & future work

This paper presents the work on modernizing parts of the original in-use WarSampo portal and developing rules to automatically validate the data. The new portal and data additions add more possibilities

to easily use the WarSampo KG for data analysis without the need for advanced programming skills or external software. Compared to the original WarSampo portal, the new WarSampo 2.0 portal not only has more visualization options, but also includes the cemeteries from the ceded Karelian areas that are missing from the original WarSampo data and thus could not be analyzed and visualized in the UI before. Additionally, the work helps ensure that the portal will remain usable also in the future to preserve the knowledge of Finnish WW2-era military history in a way that not only enables people to use it for research, but also makes the data accessible to the general public interested in it.

The results from both the SHACL-based validation and user feedback have proved fruitful in the task of diagnosing issues in the data and highlight the need for resources devoted to correcting them, both in terms of the actual content (e.g., contradicting dates) and more "technical" issues like fixing missing class definitions for instances. If additional funding is secured, this kind of work is planned to be carried out before the official publication of the WarSampo 2.0 portal. One part of this work would be to also set up a working infrastructure to handle future data maintenance primarily in native RDF with a tool such as the SAHA metadata editor (Mäkelä and Hyvönen 2014) or equivalent to avoid potential future problems arising from having to transform the WarSampo data between different formats to carry out edits.

The agentic approach employed in this paper to generate new SHACL rules tailored to the WarSampo dataset helped to discover more than 150 additional violations. The agent could be tasked with making new batches of rules to complement its existing ones and could even be guided to focus on a certain subset of data if needed. The agent as a rule creator saves the human time by not having to go through the data manually to find pairs of properties that could have contradicting values (e.g., date of death and birth) or coming up with missing ranges, and still provides an explainable, re-runnable process for catching violations.

In this case, the AI agent's task was explicitly confined to creating new validation rules, but it could also be tasked with trying to discover mistakes that pure SHACL rules cannot catch (e.g., typos or other inconsistencies in data values). The agent could then also make suggestions for fixes — possibly with reasoning as to why a specific suggestion was made by providing sources — to the human editor of the data, who could either accept or reject them.

Acknowledgments: Support of the Finnish Digital Humanities research infrastructure initiative FIN-CLARIAH/DARIAH-FI funded by the Research Council of Finland and NextConnectionEU is acknowledged. Computational resources of CSC – IT Center for Science have been used in the research. The computational resources provided by the Aalto Science-IT project are acknowledged. The first author acknowledges the support of Kaatuneiden Muistosäätiö for this project through a personal working grant.

References

Alava, Teuvo, Dmitri Frolov, Reijo Nikkilä, and Timo Karvonen. 2002. *Rukiver! Suomalaiset sotavangit Neuvostoliitossa.* Edited by Dmitri Frolov, Reijo Nikkilä, and Timo Karvonen. Helsinki: Edita.

Bast, Hannah, and Björn Buchhold. 2017. "QLever: A Query Engine for Efficient SPARQL+Text Search." In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management,* 647–656.

Boer, Victor de, Johan van Doornik, Lars Buitinck, Maarten Marx, Tim Veken, and Kees Ribbens. 2013. "Linking the kingdom: enriched access to a historiographical text." In *Proceedings of the Seventh International Conference on Knowledge Capture,* 17–24. K-CAP '13. Banff, Canada: Association for Computing Machinery. ISBN: 9781450321020. https://doi.org/10.1145/2479832.2479849.

Collins, Trevor, Paul Mulholland, and Zdenek Zdrahal. 2005. "Semantic Browsing of Digital Collections." In *The Semantic Web – ISWC 2005,* edited by Yolanda Gil, Enrico Motta, V. Richard Benjamins, and Mark A. Musen, 127–141. Berlin, Heidelberg: Springer Berlin Heidelberg. ISBN: 978-3-540-32082-1.

Doerr, Martin. 2003. "The CIDOC conceptual reference module: an ontological approach to semantic interoperability of metadata." *AI magazine* 24 (3): 75–75.

García-González, Herminio, and Mike Bryant. 2023. "The Holocaust Archival Material Knowledge Graph." In *The Semantic Web – ISWC 2023,* edited by Terry R. Payne, Valentina Presutti, Guilin Qi, María Poveda-Villalón, Giorgos Stoilos, Laura Hollink, Zoi Kaoudi, Gong Cheng, and Juanzi Li, 362–379. Cham: Springer Nature Switzerland. ISBN: 978-3-031-47243-5.

Gayo, Jose Emilio Labra, Eric Prud'hommeaux, Iovka Boneva, and Dimitris Kontokostas. 2017. *Validating RDF Data.* 7:1–328. Synthesis Lectures on the Semantic Web: Theory and Technology 1. Morgan & Claypool Publishers LLC, September. https://doi.org/10.2200/s00786ed1v01y201707wbe016.

Hearst, M., A. Elliott, J. English, R. Sinha, K. Swearingen, and K.-P. Lee. 2002. "Finding the flow in web site search." *CACM* 45 (9): 42–49.

Hyvönen, Eero. 2023. "Digital Humanities on the Semantic Web: Sampo Model and Portal Series." *Semantic Web – Interoperability, Usability, Applicability* 14 (4): 729–744. https://doi.org/10.3233/SW-223034.

Hyvönen, Eero, 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 *Proceedings of the 13th International Conference on The Semantic Web. Latest Advances and New Domains - Volume 9678,* 758–773. Berlin, Heidelberg: Springer-Verlag. ISBN: 9783319341286. https://doi.org/10.1007/978-3-319-34129-3_46.

Hyvönen, Eero, Petri Leskinen, Henna Poikkimäki, Heikki Rantala, Rafael Leal, Jouni Tuominen, Senka Drobac, Ossi Koho, Ilona Pikkanen, and Hanna-Leena Paloposki. 2026. "Searching, Exploring, and Analyzing Historical Letters and the Underlying Networks: LetterSampo Finland – Finnish 19th-Century Letters on the Semantic Web." *Digital Humanities in the Nordic and Baltic Countries Publications* 7 (4). https://doi.org/10.5617/dhnbpub.12932.

Hyvönen, Eero, Petri Leskinen, and Jouni Tuominen. 2023. "LetterSampo – Historical Letters on the Semantic Web: A Framework and Its Application to Publishing and Using Epistolary Data of the Republic of Letters." *Journal on Computing and Cultural Heritage* 16 (1).

Hyvönen, Eero, Laura Sinikallio, Petri Leskinen, Senka Drobac, Rafael Leal, Matti La Mela, Jouni Tuominen, Henna Poikkimäki, and Heikki Rantala. 2025. "Publishing and Using Parliamentary Linked Data on the Semantic Web: ParliamentSampo System for Parliament of Finland." *Semantic Web* 16 (1). https://doi.org/10.3233/SW-243683.

Hyvönen, Eero, and Jouni Tuominen. 2024. "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. https://ceur-ws.org/Vol-3759/paper4.pdf.

Ikkala, Esko, Eero Hyvönen, Heikki Rantala, and Mikko Koho. 2022. "Sampo-UI: A Full Stack JavaScript Framework for Developing Semantic Portal User Interfaces." *Semantic Web – Interoperability, Usability, Applicability* 13, no. 1 (January): 69–84. https://doi.org/10.3233/SW-210428.

Ikkala, Esko, Mikko Koho, Erkki Heino, Petri Leskinen, Eero Hyvönen, and Tomi Ahoranta. 2017. "Prosopographical Views to Finnish WW2 Casualties Through Cemeteries and Linked Open Data." In *Proceedings of the Second Workshop on Humanities in the Semantic Web (WHiSe II),* edited by Alessandro Adamou, Enrico Daga, and Leif Isaksen, 45–56. CEUR Workshop Proceedings, Vol. 2014. http://ceur-ws.org/Vol-2014/.

Koho, Mikko, Erkki Heino, and Eero Hyvönen. 2016. "SPARQL Faceter - Client-side Faceted Search Based on SPARQL." In *Joint Proceedings of the 4th International Workshop on Linked Media and the 3rd Developers Hackshop.* Vol 1615. Heraklion, Crete, Greece: CEUR Workshop Proceedings, May. http://www.ceur-ws.org/Vol-1615.

Koho, Mikko, Esko Ikkala, and Eero Hyvönen. 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),* 31–39. CEUR Workshop Proceedings, June. http://ceur-ws.org/Vol-3152/BD2019_paper_5.pdf.

Koho, Mikko, Esko Ikkala, Petri Leskinen, Minna Tamper, Jouni Tuominen, and Eero Hyvönen. 2021. "WarSampo Knowledge Graph: Finland in the Second World War as Linked Open Data." *Semantic Web – Interoperability, Usability, Applicability* 12, no. 2 (January): 265–278. https://doi.org/10.3233/SW-200392.

Koho, Mikko, Rafael Leal, Esko Ikkala, Minna Tamper, Heikki Rantala, and Eero Hyvönen. 2022. "Building Lightweight Ontologies for Faceted Search with Named Entity Recognition: Case WarMemoir-Sampo." In *Proceedings of the 1st International Workshop on Knowledge Graph Generation From Text and the 1st International Workshop on Modular Knowledge co-located with 19th Extended Semantic Conference (ESWC 2022),* edited by Sanju Tiwari, Nandana Mihindukulasooriya, Francesco Osborne, Dimitris Kontokostas, Jennifer D'Souza, and Mayank Kejriwal, 3184:19–35. International Knowledge Graph Generation From Text (TEXT2KG). CEUR Workshop Proceedings, May. http://ceur-ws.org/Vol-3184/TEXT2KG_Paper_2.pdf.

Koho, Mikko, Heikki Rantala, and Eero Hyvönen. 2022. "Digital Humanities and Military History: Analyzing Casualties of the WarSampo Knowledge Graph." In *DHNB 2022 The 6th Digital Humanities in Nordic and Baltic Countries Conference,* edited by Karl Berglund, Matti La Mela, and Inge Zwart, vol. 3232. Uppsala, Sweden: CEUR Workshop Proceedings, March. http://ceur-ws.org/Vol-3232/paper29.pdf.

Kourijoki, Alex. 2020. "Linkitetyn datan validointi ja korjaus" [in Finnish]. Master's thesis, Aalto University, School of Science, Master's Programme in Computer, Communication and Information Sciences, June. http://urn.fi/URN:NBN:fi:aalto-202006213894.

Leskinen, Petri, 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 *The Semantic Web – ISWC 2017,* edited by Claudia d'Amato, Miriam Fernandez, Valentina Tamma, Freddy Lecue, Philippe Cudré-Mauroux, Juan Sequeda, Christoph Lange, and Jeff Heflin, 280–296. Cham: Springer International Publishing. ISBN: 978-3-319-68204-4.

Linacre, Robin, Sam Lindsay, Theodore Manassis, Zoe Slade, Tom Hepworth, Ross Kennedy, and Andrew Bond. 2022. "Splink: Free software for probabilistic record linkage at scale." *International Journal of Population Data Science* 7, no. 3 (August). https://doi.org/10.23889/ijpds.v7i3.1794. https://ijpds.org/article/view/1794.

Lindquist, Thea, Michael Dulock, Juha Törnroos, Eero Hyvönen, and Eetu Mäkelä. 2013. "Using linked open data to enhance subject access in online primary sources." *Cataloging & Classification Quarterly* 51 (8): 913–928.

Mäkelä, Eetu, and Eero Hyvönen. 2014. "SPARQL SAHA, a Configurable Linked Data Editor and Browser as a Service." In *The Semantic Web: ESWC 2014 Satellite Events. ESWC 2014,* edited by V. Presutti, E. Blomqvist, R. Troncy, H. Sack, I. Papadakis, and A. Tordai, 434–438. Springer-Verlag, May. https://doi.org/10.1007/978-3-319-11955-7_62.

Mäkelä, Eetu, Juha Törnroos, Thea Lindquist, and Eero Hyvönen. 2017. "WW1LOD: An application of CIDOC-CRM to World War 1 linked data." *International Journal on Digital Libraries* 18 (4): 333–343.

Pan-Canadian Documentary Heritage Network (PCDHN). 2018. *Out of the Trenches : A Linked Open Data Project.* V. V2. https://doi.org/10.7939/DVN/URXSGC.

Rantala, Heikki, Annastiina Ahola, Esko Ikkala, and Eero Hyvönen. 2023. "How to create easily a data analytic semantic portal on top of a SPARQL endpoint: introducing the configurable Sampo-UI framework." In *Proceedings of 8th International Workshop on the Visualization and Interaction for Ontologies and Linked Data co-located with the 22nd International Semantic Web Conference (ISWC 2023) in Athens, Greece*. CEUR Workshop Proceedings, Vol. 3508. https://ceur-ws.org/Vol-3508/paper3.pdf.

Rantala, Heikki, Ilkka Jokipii, Esko Ikkala, and Eero Hyvönen. 2022. "WarVictimSampo 1914–1922: a National War Memorial on the Semantic Web for Digital Humanities Research and Applications." *ACM Journal on Computing and Cultural Heritage* 15, no. 1 (February). https://doi.org/10.1145/3477606.

Rantala, Heikki, Eljas Oksanen, Frida Ehrnsten, and Eero Hyvönen. 2024. "Publishing numismatic public finds on the semantic web for digital humanities research–coinsampo linked open data service and semantic portal." In *CEUR Workshop Proceedings*, vol. 3724. CEUR.

Six, Joren, Jahid Chetti, Annastiiina Ahola, Frederic Lamsens, Heikki Rantala, and Eero Hyvönen. 2026. "Sampo-UI Meets Ontop: Publishing Relational Digital Humanities Data on the Semantic Web." In *Posters, Demos, Blue Sky, and Tutorials at SEMANTiCS 2026*. CEUR Workshop Proceedings, Vol. 4293. https://ceur-ws.org/Vol-4293/paper2.pdf.

Sprugnoli, Rachele, Giovanni Moretti, and Sara Tonelli. 2019. "LOD Navigator: Tracing Movements of Italian Shoah Victims." *Umanistica Digitale* 3 (4). https://doi.org/10.6092/issn.2532-8816/9050.

Sutton, Angela, and Jessica Power. 2024. "Community-Driven Linked Data Approaches in Builders and Defenders: Nashville's Historical Black Civil War Database." *DHQ: Digital Humanities Quarterly* 18 (4).

Tunkelang, Daniel. 2009. *Faceted search*. Synthesis Lectures on Information Concepts, Retrieval, and Services. Springer Cham. https://doi.org/10.1007/978-3-031-02262-3.

Vafaie, Mahsa. 2024. "German National Socialist Injustice on the Semantic Web: from Archival Records to a Knowledge Graph." In *DC@ ISWC*.

Yargan, Dilek, Jörg Waitelonis, Mahsa Vafaie, and Harald Sack. 2026. *BZKO: An Ontology for the Card Index of German Post-War Compensation Records*. arXiv: 2608.06918 [cs.IR]. https://arxiv.org/abs/2608.06918.

Zeng, Aohan, Xin Lv, Zhenyu Hou, Zhengxiao Du, Qinkai Zheng, Bin Chen, Da Yin, et al. 2026. *GLM-5: from Vibe Coding to Agentic Engineering*. arXiv: 2602.15763 [cs.LG]. https://arxiv.org/abs/2602.15763.