WarVictimSampo 1914–1922: a National War Memorial on the Semantic Web for Digital Humanities Research and Applications

HEIKKI RANTALA, Aalto University, Semantic Computing Research Group (SeCo), Finland
ILKKA JOKIPII, University of Helsinki and the National Archives of Finland, Finland
ESKO IKKALA, Aalto University (SeCo), Finland
EERO HYVÖNEN, University of Helsinki (HELDIG) and Aalto University (SeCo), Finland

This paper presents the semantic portal and Linked Open Data (LOD) service WarVictimSampo 1914–1922 about the war victims, battles, and prisoner camps in the Finnish Civil and other wars in 1914–1922. The system is based on a database of the National Archives of Finland and additional related data created, compiled, and linked during the project. The system contains detailed information about some 40 000 deaths extracted from several data sources, and data about over 1000 battles of the Civil War. A key novelty of WarVictimSampo 1914–1922 is the integration of ready-to-use Digital Humanities visualizations and data analysis tooling with semantic faceted search and data exploration, which allows, e.g., studying data about wider prosopographical groups in addition to individual war victims. The paper focuses on demonstrating how the tools of the portal, as well as the underlying SPARQL endpoint openly available on the Web, can be used to explore and analyze war history in flexible and visual ways. WarVictimSampo 1914–1922 is a new member in the series of "Sampo" model based semantic portals. The portal is in use and has had 23 000 users including both war historians and the general public seeking information about their deceased relatives.

CCS Concepts: • **Information systems → Resource Description Framework (RDF)**; **Digital libraries and archives**; • **Applied computing → Arts and humanities**.

Additional Key Words and Phrases: semantic web, linked data, digital humanities, military history

ACM Reference Format:
Heikki Rantala, Ilkka Jokipii, Esko Ikkala, and Eero Hyvönen. 2021. WarVictimSampo 1914–1922: a National War Memorial on the Semantic Web for Digital Humanities Research and Applications. *ACM J. Comput. Cult. Herit.* 1, 1, Article 1 (January 2021), 19 pages. https://doi.org/10.1145/3477606

1 INTRODUCTION

This paper presents the semantic portal and Linked Open Data (LOD) service WarVictimSampo 1914–1922[1] about the war victims, battles, and prisoner camps of the Finnish Civil in 1918, including war victim data also from other Finnish wars in 1914–1922. The tools offered by the service help researchers and general public to better access and analyse the historical data. The core of the service is the LOD knowledge graph extracted from the Finnish War Victims 1914–1922 database that includes some 40 000 victims. The data was enriched with related linked data sources. Most,

[1] WarVictimSampo 1914–1922 Semantic Portal: https://sotasurmat.narc.fi/en

Authors' addresses: Heikki Rantala, Aalto University, Semantic Computing Research Group (SeCo), Finland, heikki.rantala@aalto.fi; Ilkka Jokipii, University of Helsinki and the National Archives of Finland, Helsinki, Finland, ilkka.jokipii@arkisto.fi; Esko Ikkala, Aalto University (SeCo), Finland, esko.ikkala@aalto.fi; Eero Hyvönen, University of Helsinki (HELDIG) and Aalto University (SeCo), Finland, eero.hyvonen@aalto.fi.

over 90 percent, of the deaths recorded here are related to the Finnish Civil War in 1918, and the rest are related to the other wars of the period. Fig. 1 depicts the distribution of death dates in the data during 1918. The Finnish Civil War is considered to have lasted from 27 January 1918 to 15 May 1918. As can be seen from Figure 1, many deaths fall outside of this period, and many people died in prison camps after the war by being executed or by hunger and diseases. The data includes people who have died in Finland and Finnish people who died abroad in war like conditions during the period between 1914 and 1922. The data is maintained by the National Archives of Finland.

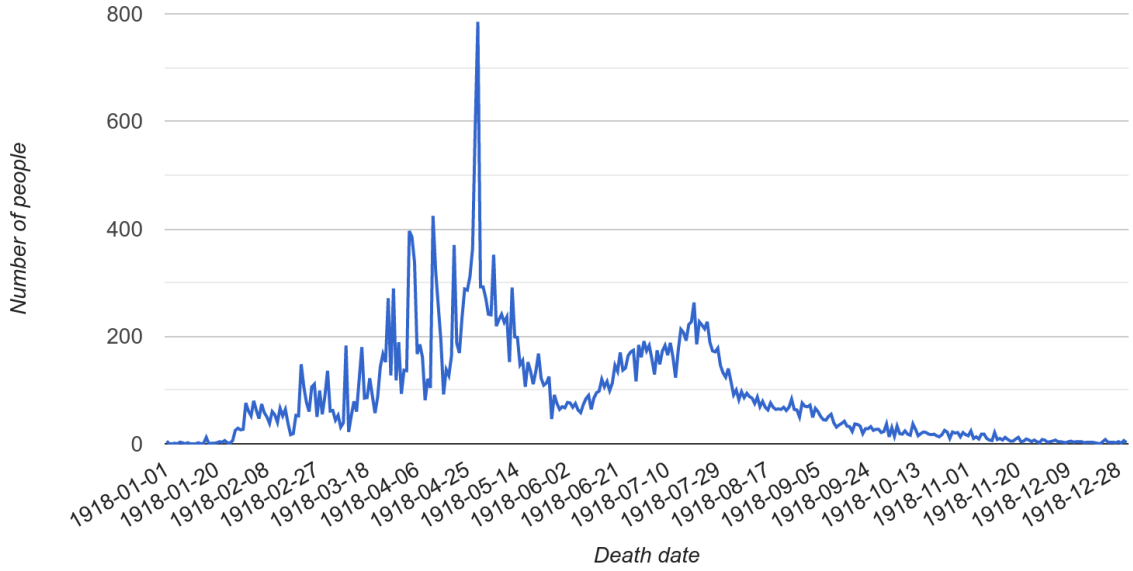

Fig. 1. Distribution of death dates in 1918 in the knowledge graph of WarVictimSampo 1914–1922

1.1 Death Records as National Memory

There are multitude of historical war-related databases around the Web. They differ in their scope and size. Naturally the databases serve the purposes of research, but they are also important places for commemoration of the victims. Probably the largest databases are The Central Database of Shoah Victims' Names[2] (approx. 4.8 million) and Memorial Foundation's database[3] Victims of political terror in the Soviet Union (more than 3 million). Both of the databases are still ongoing projects even though the databases were launched in 2004 and 2007 respectively. There are also many databases of the victims of the second world war. Also finally in the year 2018 first central database[4] of the victims of Spanish Civil War and Franco's dictatorship was published.

Finland has been in the forefront in creating war victim databases [15]. First database in Finland was the War Victims Database of the Second World War. The project started in the 1985 as a series of books but quickly transformed into a computer database. In 1991 the project started to save the names intended for the book series into the database.

[2]https://www.yadvashem.org/archive/hall-of-names/database.html
[3]https://base.memo.ru/
[4]https://www.carlapedret.cat/first-central-database-victims-spanish-civil-war-franco-regime/

Preparations for the publication of the database's web version started in 1998 and the website was published in September of 2000.

In 1998 Finnish Government started a project to research war-related deaths in the years 1914-1922 [29]. Project created a database of the victims and the online version of the database was published in 2001 by the National Archives of Finland. The database consisted of 39 500 victims of which about 93 percent died in 1918 during the Finnish Civil War. The database was updated last time in year 2004.

In Finland both of these databases have served as important sites of commemoration of the war victims. The database of the Victims of WW2 was visited by more than 500 thousand people during the first month after its publication and the interest has stayed high also after that. The database of the war victims of 1914-1922 has also been very popular. From the day of its publication, the database has had more than 1,4 million visitors. To stay relevant, databases should not merely be places for storing the data, and the technological aspects and the contents should be updated regularly.

1.2 Linked Open Data Approach

In our work, the war victims database was converted into LOD and updated with data about 1590 new, previously unknown victims, and new information concerning the old records. An important contribution of the new system is making the access to the data easier and more versatile by integrating ready-to-use DH tooling seamlessly with a faceted search and browsing [11] interface for data exploration [18]. The data is now more Findable, Accessible, Interoperable, and Re-usable as suggested by the FAIR Guiding Principles for scientific data management and stewardship[5]. This paper extends our earlier demo [20] and workshop papers [21] on the topic with description of the background of the project, it's underlying principles, data models, and technical aspects.

The old web application[6] that has been in use since 2001 includes options for exploring the data with simple search functionality and a homepage for each person. The end users of the system have deemed the search interface fairly inflexible and with too few options to choose from for exploring the data or even accessing all the data. Also means of exporting the data from the database have been asked for. WarVictimSampo 1914–1922 has been developed as a solution to these shortcomings. The data of the old database has not been openly available. This paper shows how LOD and modern web technologies can be used to enhance and update a traditional web service for historical data. It is also demonstrated how modern data analytic tools can be used with LOD in Digital Humanities (DH) research and applications.

This paper is structured as follows: First, we motivate in more detail why the linked open data was chosen as the basis of WarVictimSampo 1914–1922 and present the data publishing model Sampo underlying the application. After this, the data conversion into a Linked Open Data service is in focus. To test and demonstrate the possibilities of the proposed approach for Digital Humanities research, tools for data exploration, analysis, and visualization are then presented. In conclusions, related works are discussed and directions for the future are outlined.

2 SAMPO MODEL FOR SEMANTIC PORTALS IN DIGITAL HUMANITIES

There are many reasons why Semantic Web (SW) technologies and Linked Data [5] are a promising approach for creating semantic portals for Cultural Heritage applications [7]:

(1) **Interoperable data aggregation**. The idea of using ontologies [1, 24] with well-defined semantics in data modeling and metadata descriptions addresses problems of semantic interoperability. This is a challenge when

[5] https://www.go-fair.org/fair-principles/
[6] http://vesta.narc.fi/cgi-bin/db2www/sotasurmaetusivu/main?lang=en

aggregating CH data from separate data silos of memory organizations in a distributed content creation environment

(2) **Data enrichment**. Semantic data, i.e., data that can be comprehended by the machine, can be used for reasoning new data [6] and for knowledge discovery.

(3) **Intelligent applications**. When the machine "understands" the data it is dealing with, more intelligent applications and services can be created with less human effort.

(4) **Reusing data**. The Linked Data paradigm includes standards and best practices[7] for publishing content as operational data services with standardized Application Programming Interfaces (API), including the notions of SPARQL[8] endpoints and resolving URI references. The same data service can flexibly be used to support different applications without modifying the data, and be separated from the user interface logic, which helps application development.

These ideas are very much in line with the FAIR principles. They form a basis of our own work on developing the Sampo model [10], a framework for designing, implementing, and using semantic portals in Digital Humanities. WarVictimSampo 1914–1922 is based on the Sampo model, and a member in the growing Sampo series of portals[9]. The underlying linked open data of the portal is hosted at the Linked Data Finland platform[10] [8], where it can be queried using SPARQL, and where various services are available to support re-use of the data. The semantic portal makes queries to this publicly open endpoint, and a researcher can also query the endpoint for her own purpose using tools, such as YASGUI [23] for SPARQL, or Jupyter[11] and Google CoLab[12] by Python scripting. It is also possible to download the dataset for other kind of applications.

The Sampo model includes three components related to 1) the model for creating and publishing heterogeneous, distributed Linked Data, 2) to providing the end user with multiple *application perspectives* to the contents, and 3) how the application perspectives can be used in two basic steps. This model is called "Sampo" according to the Finnish epic Kalevala, where Sampo is a mythical machine giving riches and fortune to its holder, a kind of ancient metaphor of technology[13]. The idea of collaborative content creation by data linking is a fundamental idea behind the Linked Data movement [14] and has been developed also in various other settings, e.g., in ResearchSpace[15].

1. Data creation and publishing model. The ideas of the Semantic Web and Linked Data can be applied to address the problems of semantic data interoperability and distributed content creation at the same time. A shared semantic ontology infrastructure, including mutually aligned shared domain ontologies and core metadata is used for hamonizing and interlinking dataset, based on SW standards[16]. If content providers outside of the circle provide the system with metadata about their contents, the data is automatically linked and enriched with each other and forms a knowledge graph represented using RDF[17]. For example, if metadata about a painting created by Picasso comes from an art museum, it can be enriched (linked) with, e.g., biographies from Wikipedia and other sources, photos taken of Picasso, information about his wives, books in a library describing his works of art, related exhibitions open in museums, and so on. At the

[7] https://www.w3.org/standards/semanticweb/
[8] https://www.w3.org/TR/rdf-sparql-query/
[9] https://seco.cs.aalto.fi/applications/sampo/
[10] http://www.ldf.fi/dataset/siso
[11] https://jupyter.org/
[12] https://colab.research.google.com/notebooks/intro.ipynb#recent=true
[13] https://en.wikipedia.org/wiki/Sampo
[14] https://lod-cloud.net
[15] https://www.researchspace.org/
[16] https://www.w3.org/standards/semanticweb/
[17] https://www.w3.org/RDF/

same time, the contents of any organization in the portal having Picasso related material get enriched by the metadata of the new artwork entered in the system. This is a win-win business model for everybody to join; collaboration pays off. The knowledge graph can be published as a data service in a SPARQL end-point using the principles of Linked Data [5].

2. Multiple perspective interface design. On top of the data service different applications can be created by re-using the data service, without modifying the data. For example, in WarSampo [12] the data about the Second World War can be accessed from nine points of view: historical events, people, units, places, articles, death records, photographs, cemeteries, and prisoners of war. In Sampo portals the application perspectives are provided on the landing page of the system. By selecting one of them the corresponding application is opened.

2. Filter-analyze two-step usage cycle. In many Sampos, the application perspectives can be used by a two-step cycle for research: First the focus of interest, the target group, is filtered out using faceted semantic search [25]. Second, the target group is visualized or analyzed by using ready-to-use DH tools of the application perspectives. For example, in BiographySampo [13] a group of people, such as the clergy of the 19th century Graph Duchy of Finland, can be filtered out first. After this, the life charts of the priests from places of birth to death can be visualized for analyzing their mobility, their mutual network be visualized, various statistics of the group be viewed, and so on.

To develop, test, and demonstrate the model, a series of Sampo portals have been created and are in use on the Semantic Web in Finland. These living lab prototypes and applications have been created as part of research projects at the Semantic Computing Research Group (SeCo) active at Aalto University and the University of Helsinki, Helsinki Centre for Digital Humanities (HELDIG), and are based on collaborations with a large network of Finnish memory and other organizations as data providers and cultural heritage domain experts. The systems are examples of utilizing a national level FinnONTO ontology and Linked Open Data infrastructure [9] that has been developed in conjunction with the portals.

3 DATA, ONTOLOGIES, AND DATA SERVICE

3.1 Data Transformation

The data was delivered by the National Archives of Finland as three separate CSV files taken from the old relational database with a documentation. The data is too complex to easily be represented as a one simple table, and for example the sources had their own separate table. All the victims had identifiers in the data and they were preserved as part of the new URI identifiers created for the victims and their death records. The type of data was represented with two number codes in the old database. For example number 24 would represent information relating to the military position of the person, and a second number 4 with that previous number would represent company that he or she served with. There are over 150 different information types in the data. We represent information types with resources that have the original number codes in their name, but also a plain language explanation, for example here 24_4_company. This is meant to make the data easier to read and understandable even without reading any documentation, and also make it easy to trace the data to the original documentation if necessary at some point in the future. Some pieces of data included codes that were not explained in the documentation. In some cases it was possible to deduce the meaning based on the context. In some other cases the unexplained codes may be errors in the data, when one of the codes has been entered wrong by an updater. For those cases a separate information type was created, and the original codes are preserved as properties of the resource. This allows future researchers to possibly try to fix the errors.

Some elements of the original data were not available to us. For example the original data included a lot of information relating to the societies that the person belonged to. That could be important information relating to the Civil War and persons motivations. This data was mostly not available in the data that we have. The data has few thousands of clarifying additional pieces of information relating to societies, but without the main information, or documentation, it is difficult to understand the meaning of such information. These pieces of information have also been preserved to the converted data with their codes. Researchers may find that data still interesting. It is also possible that the main data relating to the societies is found in the future and made available to convert. Because the conversion was made in a consistent manner preserving the original identifiers, it should be possible to include this data to the RDF database.

3.2 Updating the Data

In the years following the publication of the War Victims Database, the National Archives of Finland has received hundreds of corrections to the data from the public. Most of the corrections have been quite small but some people have also sent completely new information about their deceased relatives. Before updating the database, all of the corrections sent by the citizens were verified from the archival material.

In addition to active citizens and the relatives of the deceased, the historians have also continued the research around the victims of the Civil War. The update project received datasets consisting of more than 1500 new victims and more than 2000 important corrections for the data.

We chose to use CSV table as the format of adding the updates to the data. This was because a table is a format of information that is easy to understand even by people without computer science background, and can be filled by history experts without considerable training. The update data from the table was then converted with mainly a Python script to RDF format. We do not update the database directly, but instead modify the text files in Turtle format that include the data. When data is updated the whole server is then rebuilt from the updated files. This also allows using existing ontology editors to update the ontologies in the data.

The obvious draw back of our approach of using CSV files and a custom built script to run the update is that the process is not reusable for other data sets without modifications. The script may also need to be modified if data model is changed.

3.3 Data Model

The main content of the WarVictimSampo knowledge graph is the detailed information about the war victims. When transforming the old database to RDF form we aimed at creating a data model that is as simple as possible and crucially includes all the information in the original database. We have not tried to use any harmonizing data models, such as CIDOC CRM. Some parts of the data, such as the birth places of the people, could be presented with a harmonizing data model to make connecting to other materials easier. This can be done in the future. We have generally used project specific properties and classes. The namespace of the project is <http://ldf.fi/siso/> and the namespace of the project schema is <http://ldf.fi/siso/schema>. The schema namespace has been shortened to ':' in the examples of this paper. In few places we have utilized certain commonly used properties especially from the SKOS[18] vocabulary.

This data is represented in two ways in the WarVictimSampo knowledge graph. The pieces of information, like birth place or occupation, are represented as individual information resources. These information resources have properties that include value, type of information, and sources. These pieces of information can some times be conflicting if

[18] https://www.w3.org/TR/swbp-skos-core-spec/

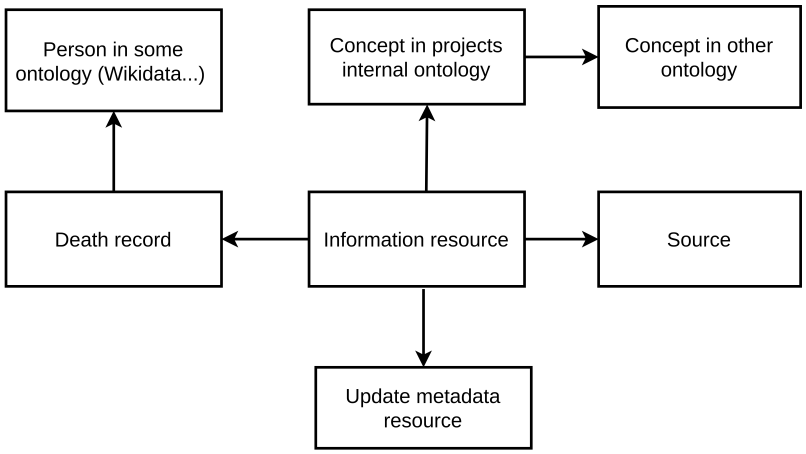

Fig. 2. A simplified model shoving the basic relations in the data model of the WarVictimSampo 1914–1922.

sources disagree. For each victim there is a single death record in the data that has a simple table like structure and only one value for each property. The death records have properties for the most important pieces of information, and represent the information that is considered most reliable if there are conflicting sources. The Figure 2 illustrates the basic relations in the data model.

For example, the following extract in Turtle[19] format includes the information expressed in a single information resource. The resource has properties for referred death record, information type and the value of information, which can be compared to the subject, predicate, and object of an RDF triple. In addition there is a property for the source of this particular information, and a property referencing metadata relating to the creation of this particular resource, such as time, creator, and information is this piece of information meant to overrule a previous piece of information. The properties are project specific, but one could, for example, define the projects source property as subproperty of the commonly used <http://purl.org/dc/elements/1.1/source> property of the Dublin Core[20] metadata model. This would allow simple connection to sources of other possible similar materials.

```
<http://ldf.fi/siso/death_records/information/10141_tt_26> a :Information_resource ;
    :creation_project <http://ldf.fi/siso/projects/t2019> ;
    :information_type siso-schema:0_26_party ;
    :referred_death_record <http://ldf.fi/siso/death_records/p_10141> ;
    :source <http://ldf.fi/siso/sources/tt> ;
    :value <http://ldf.fi/siso/parties/punainen> .
```

3.4 Ontologies for Content Description

Ontologies [24] are an important part of describing content with linked data. The ontologies provide the vocabulary to use in the particular data and they can be used to connect different data sets if the ontologies are either shared or linked to each other. The ontologies can also include machine understandable data about the concepts. This can be used to make automatic deductions. The unique identifiers of ontologies can be useful in differentiating between concepts. For

[19] https://www.w3.org/TR/turtle/
[20] https://www.dublincore.org/

example our place ontology includes a concept for Liverpool because one Finnish sailor died there according to the data in 1915. When using the identifier of Liverpool from the projects place ontology, <http://ldf.fi/siso/places/m_liverpool>, we can be sure that the meant place is not perhaps one of the various Liverpools in USA. We might also automatically deduce that this Liverpool is in England, and we can connect the concept to other relevant ontologies, and through those to other data that is linked to same ontologies.

To describe the war victims and battles, we use a wide variety of ontologies. These include currently ontologies for place, occupation, marital status, gender, nationality, language, military organization, form of recruitment, military rank, military position, combatant status, party in the civil war, armed status, family welfare, prison camp, manner of death, burial method, and religion. Some of these ontologies are very small, perhaps only two concepts such as with genders, while there are 1391 occupation concepts and 1247 concepts for municipality level places alone. The ontologies were created based on the vocabularies used in the original database, but we made some corrections and harmonization especially with places. We also created a hierarchy for places, that allows automatically deducing, for example, that certain municipality is part of a certain province. We have used project specific ontologies, instead of using some generic ontologies, because it would be difficult to ensure that the specific historical nuances of the data would not be lost if they would be directly harmonized to some outside ontology. These ontologies were then linked to harmonizing ontologies where appropriate.

3.5 Enriching the Data with Linking and Reasoning

The place and occupation ontologies are the largest and arguably most important ontologies of the data, and they are linked to outside ontologies. The places are linked to the place ontology of BiographySampo[21] [13] and through that to the general Finnish ontology place ontology YSO places[22]. The places are currently ontologized only at municipality, or higher, level. The data includes often also more exact information about, for example, very specific death places, but we have not yet had the resources to ontologize and link these literal terms. This remains future work, but would be very useful, and could be used to find exact coordinates for the death places. Currently we use municipality coordinates extracted through links to BiographySampo ontology to show death municipalities on map in the WarVictimSampo 1914–1922 portal.

Occupations are linked to AMMO [16] ontology. This is an ontology of Finnish historical occupations, and it includes, for example, mappings to international HISCO [27] and HISCLASS [26] classification of occupations. These provide connections to international materials and may be used to deduce the social status of people in the data. The main limitation is that the most common occupation in the data is worker (työmies). This is too vague definition to have a HISCLASS classification for social class, and that needs to be considered when creating analysis based on those classifications.

The war victims are linked to certain other person ontologies. We have created mappings to Wikidata[23], Wikipedia, BiographySampo[24], student data of University of Helsinki[25], and the Norssi Alumni web service[26]. We currently have 167 links to Wikipedia and Wikidata, 27 links to student records, 14 links to biographies in Biographysampo, and 26 links to people from Norssi Alumni service. Wikidata is the semantic version of Wikipedia, and includes mainly quite

[21] https://seco.cs.aalto.fi/projects/biografiasampo/en/
[22] https://finto.fi/yso-paikat/en/
[23] https://www.wikidata.org/
[24] https://seco.cs.aalto.fi/projects/biografiasampo/en/
[25] https://seco.cs.aalto.fi/projects/yo-matrikkelit/
[26] https://www.norssit.fi/semweb/

important people, so only a limited number of war victims have corresponding resources there. The identifiers used in the data allow future linking to, or from, genealogy sites or other materials. The data also includes some victims who may be of international interest, such as Russian and German soldiers who died in Finland. Linking their data to international materials may be interesting future work.

To lessen the human workload when updating the data, we use automated reasoning in the update process. For example, we use a place ontology to automatically deduce provinces based on municipalities, and we automatically calculate ages based on birth and death dates. The automatically deduced values have a separate source resource to indicate to the users, that these values we not entered by a human expert, but were instead automatically filled by a computer program.

3.6 Linked Open Data Service

The linked data underlying WarVictimSampo 1914–1922 is hosted by the Linked Data Finland platform LDF.fi [8]. LDF.fi automates the process of publishing linked datasets as far as possible in the following way: The publisher is expected to create an RDF dataset with minimal metadata about it and its schemas using the W3C Service Description recommendation[27] and the VoID vocabulary[28], and the data is stored into a SPARQL endpoint. It is also possible to give an example URI pointing into the dataset, a SPARQL query example for querying the data, and optionally a link to possible visualizations of the dataset. Based on such metadata, LDF.fi generates for each dataset a home page on which the following functionalities are available to support re-use of data:

(1) Links for downloading datasets and graphs are provided (if licensing permits it).
(2) Schemas can be downloaded if provided with the data, and links to their documentation are provided (when available).
(3) Following forms are created for inspecting the dataset in more detail: 1) Given a URI the corresponding RDF description can be read in various formats (Turtle, RDF/XML, RDF/JSON, N3, N-triples) for human consumption in a browser. The example URI is used as a first choice to try out. 2) Given a URI, Linked Data browsing can be started from it, with the example URI as a starting point.
(4) There is a SPARQL query form for querying the service with the given query used as a first example.
(5) Links providing analysis reports of the graphs in the dataset are provided. They tell the end-user what schemas (vocabularies) are used in the data, and how they have been used. Issues on data quality are pointed out.
(6) SPARQL Service Descriptions of the datasets are provided, if available. LDF uses W3C SPARQL Service Description recommendation for this.
(7) Links to visualizations of the data that may give the re-user more insight on how the dataset can be used in applications.
(8) Licensing conditions of the dataset are provided as well as a label of 1–7 stars. LDF.fi generalizes the traditional linked data 5-star model[29] of Tim Berners-Lee into a 7-star model, with the goal of encouraging data publishers to provide their data with explicit metadata schemas (6. star) and to validate their data for better quality (7. star).

[27] http://www.w3.org/TR/sparql11-service-description/
[28] http://rdfs.org/ns/void
[29] http://5stardata.info/

4 WARVICTIMSAMPO 1914–1922 PORTAL

A semantic portal was developed to allow different user groups to access the data easily. The user groups include researchers, students, and the wider public interested in either the Finnish Civil War in general or the fates of their relatives. The portal is also used as the main exploration tool of the data by the maintainers of the data at the National Archives of Finland.

4.1 Using the Portal

The user interface is built around the concept of faceted search [4, 25]. With faceted search, the user can easily narrow the search step by step by making selections based on predetermined orthogonal hierarchies of property values called facets. Facets also show the number of available items with each possible selection. This allows the user to immediately see the number of solutions of each possible selection. Combined with selections on other facets like occupation, party, and age, the user may also draw interesting conclusions by observing the hit distributions on the facets. Faceted search can therefore be used to not only find individuals that fit certain criteria, such as relatives, but it can also be used to find information about the distributions of different kind of the casualties. The faceted search paradigm is an example of exploratory search.

The user interface currently includes two main perspectives for exploring the underlying knowledge graph: 1) The main perspective is based on searching and exploring the casualties. 2) There is also a perspective based on the battles of the Finnish Civil War, covering currently 1182 geo-coded battles. Other views may be added later in the same way as in other "Sampo" series semantic portals[30]. At the front page of the portal, the user is presented with links to the different perspectives. The perspectives can also be navigated using the to bar of the application.

The user can navigate to the War Victims perspective by clicking the picture representing this perspective. There the user is presented with facets that can be used to filter the data on the left, and results view on the right. The user can switch the result view using tabs above the view. As default the user is presented with a table view of the war victims in the selection. At Figure 3 selections have been made in the facets so that the group has been narrowed down to 78 people that are shown as a list.

The user can make selections facets and the selection is updated after each selection. Facets also show the number of hits available for each selection. This can be used to immediately see relative numbers in the data. For example if no other selections are made, the selection Tampere in birthplace facet shows the number 458 after it. This means that there are currently 458 people who were born in Tampere in the data. Selections in other facets may change this number, as the selected group changes. For example, after selecting the Reds from the party facet, this number would change to 357. This reflects that there are currently in the data 357 war victims who were born in Tampere and who supported the Red side in the civil war.

The facets include text facets, checkbox facets, date facets and slider facets. The free text tool can be used to search with the name of the victim, and this may be the most common search tool used by the public. Check box facets are used to filter ontologized values such as the party in the Civil War. Date facets van be used to filter based on death or birth date, and slider facets can be used to filter based on numeric information such as age or the number of children. For example in Figure 3 the user has selected Helsinki from the birthplace facet and selected May 1918 as a time interval from the death date facet.

[30] https://www.europenowjournal.org/2019/09/09/linked-data-in-use-sampo-portals-on-the-semantic-web

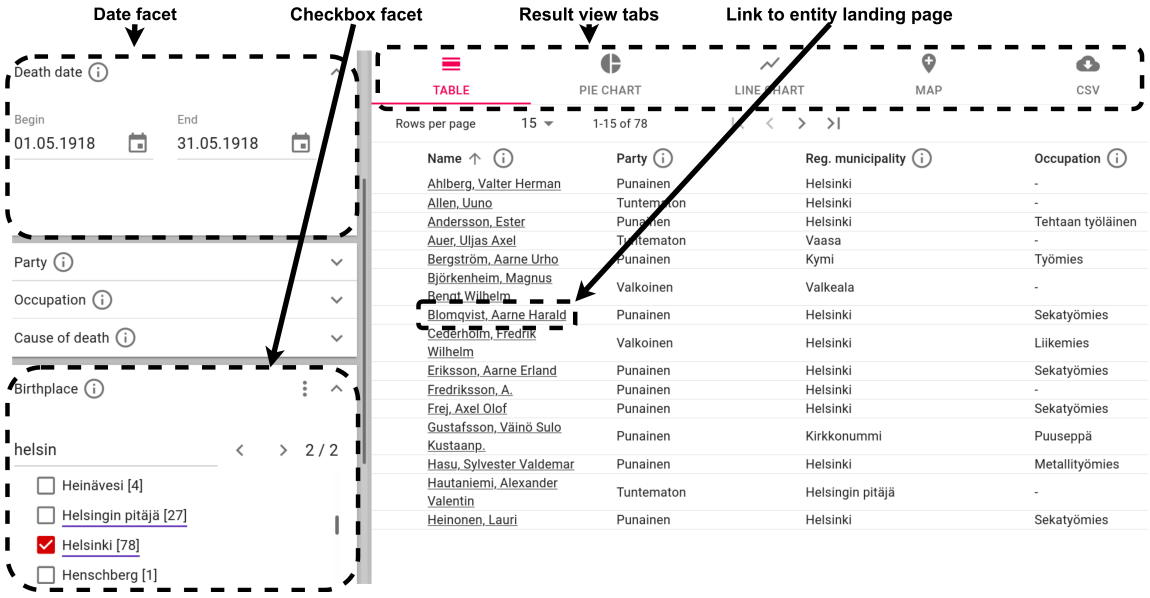

Fig. 3. A faceted search view of the WarVictimSampo 1914–1922914–1922 portal. By using the facets on the left, the initial result set of some 40 000 war victims have been narrowed down into 78 victims who were born in Helsinki and died in May 1918.

The name of the person in the table view has a link to the individual landing pages of each victim. These pages show the detailed information relating to the person, including the sources of each individual piece of information. Sometimes there can be conflicting information about certain thing in the different sources, and these conflicting pieces of information are shown as a list with all their separate sources.

The user can switch the result view using the tabs above the result view. In addition to the default table view, the War Victim perspective includes a pie chart view, line chart view, a map view of death places, and an option to download the selection as a CSV file.

When selecting the pie chart view, the user is presented with a pie chart visualization of the current selection. The pie chart visualizes the relative numbers of different values of within a certain type of information. As default the chart visualizes the relative numbers of victims from different parties of the Civil War. Without any selections in the facets the user can see here that 71 percent of the victims in the data supported the red side. By making selections in the facets the user can narrow down the group to, for example, people from certain town. This visualized value can be changed from menu. The pie chart can visualize, for example, gender, occupation, or manner of death distributions.

The line chart view shows a line chart visualization based on the current selection. The distribution of death dates shown in Figure 1 was created using this visualization view. Currently there are three different options for this visualization: age at death, birth year, and death date. For example the age at death option automatically draws a line chart where the x axis represents age in years and y axis represents the number of victims with that age at death. The average and median values are also shown under the visualization. The user can then easily and quickly compare distributions and average values between certain subsets of people in the data. For example in Figure 4 we have used the facets to select a subset of victims that represents those that supported the Red side in the Civil War and were registered to the Viipuri province. Comparing this distribution to people from other provinces shows that war victims

who supported the Red side of the Civil from the Viipuri Province tended to be older than others according to the data, with median age over thirty. Explaining this would require more detailed analysis, but this demonstrates how faceted search combined with simple analysis tools can be effective at finding interesting phenomenons in the data.

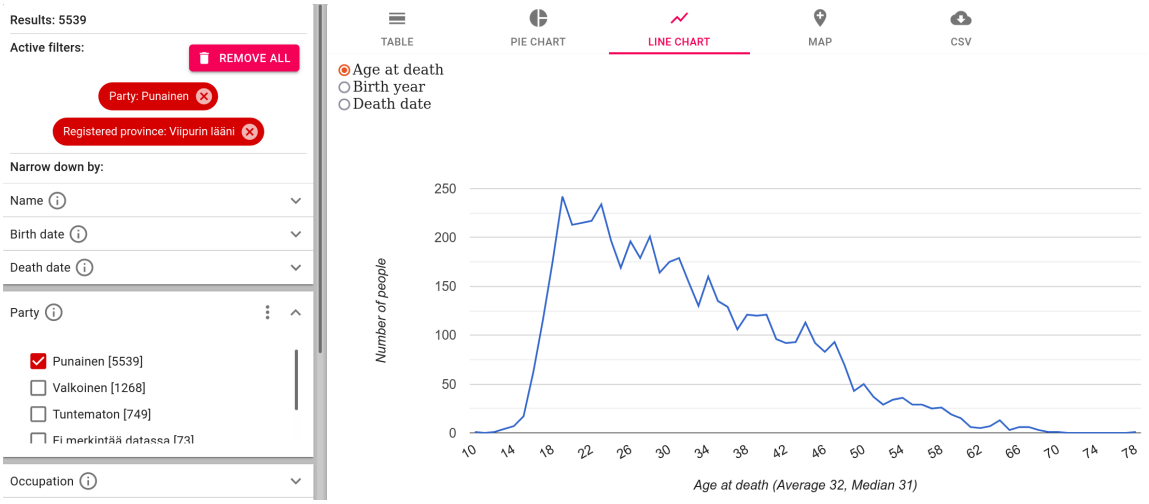

Fig. 4. The age distribution of people who supported the Red side in the Finnish Civil War and who were from the Viipuri Province as shown in WarVictimSampo 1914–1922 portal.

The map view shows the death places of the victims in selection on a map. This map is clustered so that nearby places are grouped together depending on the zoom level. Each cluster shows the number of victims that died in that area. The death places are shown on a municipality level. If there is no death municipality data, a victims information is not shown on the map.

Final option on the right is the CSV option that allows for downloading information of the selected war victims as a CSV file where one row corresponds to a one victim. This feature was considered important and was requested by history researchers. This allows the users to easily download raw data to use with their own analysis tools. Not all the data is however presented in this CSV file, as it is difficult to condense all the aspects of the knowledge graph to a single table. For example metadata such the information sources is left out from this file.

The other perspective of the portal is the Battles of The Finnish Civil War portal. This works in similar way as the war victims perspective. The user can search and filter the battles using facets and the results can be shown as table, map, or an animation. This animation view is unique to the battles. It shows battle sites at different times on a map. The marker appear on a map as a red marker when animation reaches the starting date of the battle and it then stays on the map as grey marker when the animation progresses in time. Figure 5 shows the situation on March of 1918. You can see how a clear front has formed across Finland and battles are going on along it. This view is mainly aimed for educational purposes.

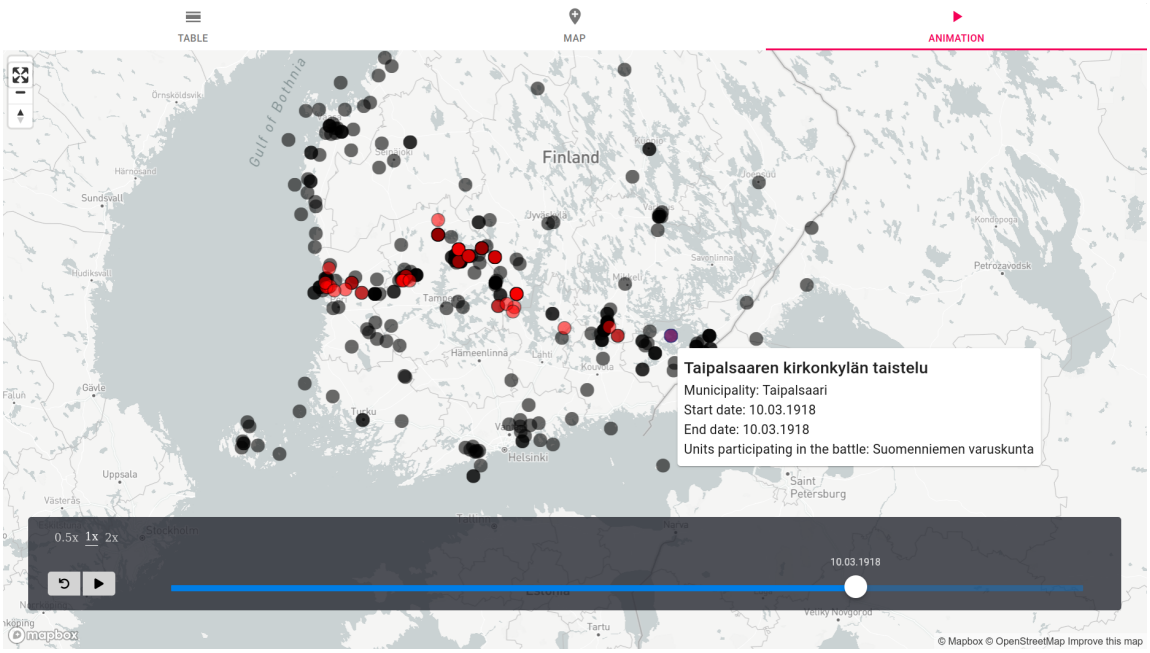

Fig. 5. The battles animation view of the WarVictimSampo 1914–1922 portal with the infobox of a single battle opened

.

Even though the data can be accessed by anyone with SPARQL queries this can be too technically demanding for many users. Even a researcher who is able to create her own SPARQL queries can find it useful to have an easy way to explore the data and to create simple visualizations quickly. The visualization tools provided by the portal are expected to be useful for both finding new data and for educating the public about history. These tools should not be expected to fully replace manual research and close reading. They are aimed to be used to spot interesting phenomena in the data that require more detailed analysis.

4.2 Technical Implementation

This subsection presents the how user interface[31] of WarVictimSampo 1914–1922 was implemented using the Sampo-UI framework[32] [14]. Sampo-UI provides software developers with a ready-to-use basis for a user interface of a semantic portal, which needs only minor modifications for deploying as a modern JavaScript web application into production.

Figure 6 presents the overall architecture of the WarVictimSampo 1914–1922 portal, provided by the Sampo-UI framework. The main parts are 1) a client based on the widely used and established React[33] and Redux[34] libraries, and 2) a Node.js[35] backend built with Express framework[36]. The primary task of the client is to display data to the end-user, and react to user's selections. The business logic of fetching the data using various search paradigms is placed on the

[31] The source code can be viewed at https://github.com/SemanticComputing/sotasurmat-web-app, and the portal tried at https://sotasurmat.narc.fi/en/.
[32] https://github.com/SemanticComputing/sampo-ui
[33] https://reactjs.org
[34] https://redux.js.org
[35] https://nodejs.org/en
[36] https://expressjs.com

backend. The client sends API requests to the backend, which queries the SPARQL endpoint and processes and merges the response rows into arrays of potentially nested JavaScript objects.

For geographic visualizations, the client is integrated with external vector and raster data sources. In WarVictimSampo 1914–1922 a raster version and a vector version of the the Mapbox Light basemap[37] is used, depending on the requirements set by the underlying geographical visualization library.

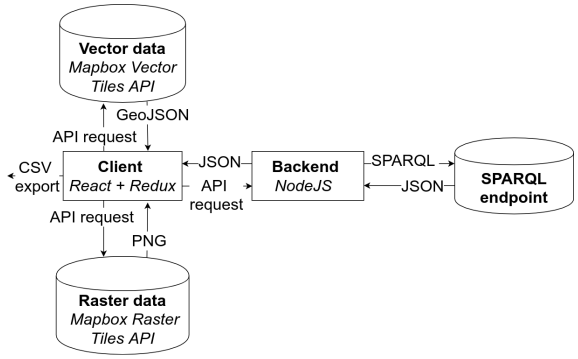

Fig. 6. The architecture of WarVictimSampo 1914–1922 portal.

The development of the user interface started by forking the Sampo-UI GitHub repository. Sampo-UI provides a pre-configured environment for full stack JavaScript development. Babel[38] is used for converting the latest features of JavaScript, such as arrow functions and the async/await syntax, into a backwards compatible version of the language for current and older browsers. Webpack[39] handles the automatically restarting development server for the client, and bundling all source code and dependencies into static assets. The Node.js backend is run concurrently with the client, and is automatically restarted using Nodemon[40] when the source code is changed. Uniform coding style is enforced by using the JavaScript Standard Style[41] package.

The user interface of WarVictimSampo 1914–1922 was developed on the basis of the default structure and components provided by Sampo-UI. The three main views, the landing page of the portal, faceted search perspective, and entity landing page, are presented in Figure 7.

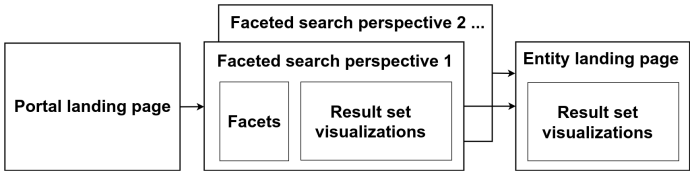

Fig. 7. Main views of the WarVictimSampo 1914–1922 portal, based on the default structure provided by the Sampo-UI framework. The arrows depict navigation links between the views.

[37] https://www.mapbox.com/maps/light
[38] https://babeljs.io
[39] https://webpack.js.org
[40] https://nodemon.io
[41] https://standardjs.com

For implementing these three main views, the Sampo-UI framework provides the developer with approximately 120 ready-to-use user interface components. The portal landing page component was configured to display links to two faceted search perspectives: War Victims and Battles. The faceted search perspectives were implemented using a combination of Sampo-UI's facet and result set visualization components. Entity landing pages were created for war victims and information sources by extending the general entity landing page component of Sampo-UI. The extension was needed for showing the sources for different pieces of information.

From a technical perspective, the sustainability of the WarVictimSampo 1914–1922 portal is fostered by the open source code, extensive documentation, and modular architecture of Sampo-UI. A significant part of the logic related to various search paradigms and the processing of search results are carried out in the backend using pure JavaScript, instead of integrating this functionality inside client-side frameworks or libraries, which are known to become deprecated considerably sooner than pure JavaScript code. The core functionality provided by the framework includes robust patterns and tools for processing the data and delivering it to the components in a predictable and uniform way, which makes is straightforward to include new features and result set visualizations in the future.

5 USING SPAQRL ENDPOINT FOR RESEARCH

The web portal allows the user to user to search and visualize the data in many ways. There are however always limitations. For serious research it is often useful to query the database directly. This allows the researcher to create custom visualizations and to find more obscure data. For example the following query[42] was created to get all the death records from the data that have no other source than the archives of the Social Democratic Party of Finland. The purpose of this was to find victims in the data that have been entered as war victims based on limited evidence, and may therefore have in fact survived.

```
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX : <http://ldf.fi/siso/schema/>
PREFIX sources: <http://ldf.fi/siso/sources/>

SELECT DISTINCT ?death_record ?homepage ?label ?id
WHERE {
    ?death_record a :Death_record ;
    skos:prefLabel ?label ;
    :identifier ?id .
    FILTER NOT EXISTS {
        ?information :referred_death_record ?death_record .
        ?information :source ?source.
        FILTER (?source != sources:sdp)
    }
    BIND(URI(CONCAT('https://sotasurmat.narc.fi/fi/victims/page/p_', ?id)) AS ?homepage)
}
```

[42]The query can be tested at Yasgui service through this link: https://api.triplydb.com/s/XZxBm87h1

The main desire expressed by the history researchers for the new database and application was the ability to download the data in CSV spreadsheet format. Spreadsheets are a data format that history researcher like to use and they have their own accustomed tools for that. We have tried to answer to that need in the web application and the user can easily download a presentation of data in CSV format using the application. It is however difficult to know what kind of data a researcher would like to have from the data. Often the researcher would simply say that he would like to get all the information. However the table like format of CSV is not the database, but only a certain representation of some parts of it. To print all the data from the database as a table is trivial and required query would take only one line. However creating a well structured and easy to read table, where all the information relating to a certain person is on one line, would be very difficult. Even in the best case the result would be difficult to understand and use. One use of SPARQL queries is to define the desired form that the researcher would like to see the data in.

Researchers can query the data directly from the endpoint to analyze the data. Such analysis can be easily made and shared with tools such as Yasgui[43]. This requires the ability to use SPARQL query language, but not necessarily programming expertise. For example the visualization in Figure 8 was created with using a simple query and visualization tools of Yasgui. This visualization show the relative numbers of war victims in the data registered to the same province from both sides of the civil war. One can see there, that Vaasa Province (Vaasan lääni) has more victims from the White side of the war, which is in clear contrast to most of the other provinces. The query used to create the analysis can be shared with a simple link[44]. Other researchers can inspect the query through the link and use that as a basis to create their own analysis. More complicated analysis can be done by, for example, using Python scripts that query and process the data.

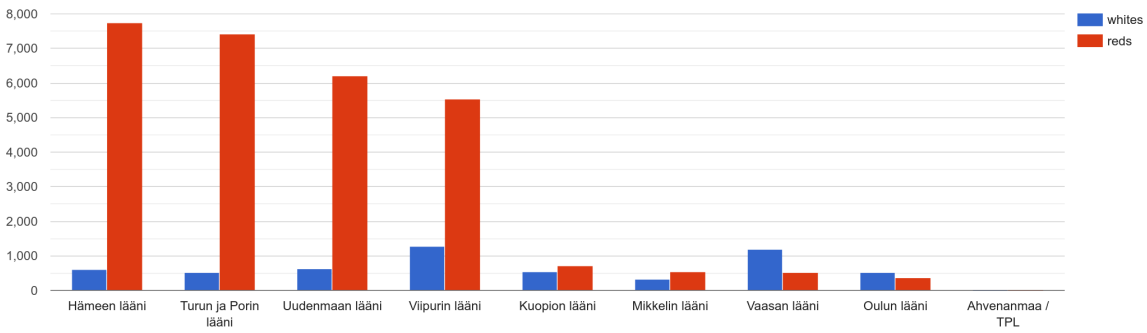

Fig. 8. A visualization showing the relative numbers of victims from both sides of the Finnish Civil War from each province.

6 DISCUSSION

In this paper we have presented ideas and technical aspects behind the WarVictimSampo 1914–1922 and how it updates and modernizes how the death records related to the Finnish Civil War and other wars of that period are maintained and published.

The results of our project will hopefully improve understanding of the Finnish Civil War by updating and correcting the data, and making it more readily available to both researchers and general public. We hope that our model will be useful in developing, maintaining, and publishing similar archives also internationally.

[43] https://yasgui.triply.cc/
[44] https://api.triplydb.com/s/414-FWdhq

Publishing the data as a Linked Open Data service allows the memory organization to better enrich their data and harmonize it with other materials. It allows the researchers an easier and direct access to the data. It also makes it possible for software developers to develop new services based on the data. These applications and tools make the data more accessible and understandable to the general public, and help historians. Faceted search integrated with visualizations makes it easy to compare groups within the data, in addition to searching information that relates to single person.

Updating the data relating to the war victims if the Finnish Civil War makes it possible for historians to create a more complete and exact picture of the Civil and it's background. The web application that we have developed in this project is open code and can be used as basis of building similar applications. In a more general sense, our project demonstrates what can be done with modern technologies to make archived data more accessible with various tools.

Linked Data makes it possible to easily make connections between different collections of data. For example, we have created links to other data sources for some of the more famous people, but in future it may be possible to connect most of the war victims to genealogy data in various sources. Linked open data will help research by making the data easily available, and it also allows using existing ontology infrastructure to lessen the workload. In our case this can be seen in practical terms in how the places are connected to an outside ontology and this allows using ready-made data such as the coordinates of places, so that they can be shown on a map, instead of having to create the coordinates just for our project.

As an example of how Linked Open Data can benefit research, we are developing an analysis of how the social status of the people fighting on different sides of the Finnish Civil War developed during the war, using the AMMO ontology of Finnish historical occupations. This can be done easily, because the occupation terms used in our war victims data are connected to the AMMO ontology and AMMO includes numeric data about the social status of the occupations. Having such ontology means also that some other research project that would be interested in doing computational analysis relating to the comparative social status of the Finnish historical occupations, can also use the data from AMMO instead of having to create completely new computer readable definitions. Using shared ontologies also makes comparing different data, such as the data about the victims of the Finnish Civil War and the Second World War in Finland, much easier.

Using Linked Open Data can also present challenges. For example when updating the data, the updater can't just type any value freely, but is restricted by the used ontologies. If new terms are required, the relevant ontologies need to be updated separately. This can add complexity to the process and require extra training. Using all the new possibilities in the most efficient way for research will also require the researchers to learn some new technical skills. However, even for researchers and general public without new technical skills, with WarVictimSampo 1914–1922 the data about the victims of the Finnish Civil War is now accessible more easily and completely than it was with the previous version of the data.

6.1 Related Work

These ideas behind the Sampo model have been explored and developed before in different contexts. For example, the notion of collaborative content creation by data linking is a fundamental idea behind the Linked Open Data Cloud movement[45] and has been developed also in various other settings, e.g., in ResearchSpace[46]. The idea of providing multiple analyses and visualizations to a set of filtered search results has been used in other portals, such

[45] https://lod-cloud.net
[46] https://www.researchspace.org

as the ePistolarium[47] [22] for epistolary data, and using multiple perspectives have been studied as an approach in decision making [17]. Faceted search [4, 25], also know as "view-based search" and "dynamic ontologies", is a well-known paradigm for explorative search and browsing [18] in computer science and information retrieval, based on S. R. Ranaganathan's original ideas of faceted classification in Libary Science. The two step usage model is used in prosopographical research [28] (without the faceted search component). The novelty of the Sampo model lies in combining several ideas and operationalizing them for developing applications in Digital Humanities.

WarVictimSampo 1914–1922 is a follow up project of WarSampo [12], which uses LD to present and publish information related to the Second World War in Finland, including death records. The novelty of WarVictimSampo 1914–1922 lays in the idea of developing new data-analytic tooling for research in war history, as well as in creating, cleaning, extending, and publishing the former War Victims 1914–22 database for open use on the the Semantic Web.

There have been several projects publishing linked data about the World War I on the Web, such as Europeana Collections 1914–1918[48], 1914–1918 Online[49], WW1 Discovery[50], Out of the Trenches[51], CENDARI[52], Muninn[53], and WW1LOD [19]. In addition to WarSampo, there are a few works that use the Linked Data approach to WW2, such as [2, 3], Open Memory Project[54] on holocaust victims, and the Dutch project Netwerk Orloogsbronnen[55].

6.2 Future Work

The National Archives of Finland still continue to receive new information about the war victims of the Finnish Civil War, for example from the family members of the victims or from research. It can be very important to the people that their relatives information is correct in these records. Therefore updating data continues in the future. Some mistakes in the data do also come up and do require fixing. In future we aim to develop the process of updating the data so that it will become easier and more convenient to the experts within a memory organization, requiring as little computer expertise as possible.

There is also a lot of possibilities to improve the ontologies related to the data, link them to more outside materials, and use them to enrich the data. For example the exact places and military units could be ontologized in the future. Linking the ontologies with other materials of the National Archives of Finland, and other both domestic and international sources, can be improved in the future.

New information about the victims of the Finnish Civil War may offer new information about the course of the Civil War and it's background. We aim to cooperate with history researchers to do a more detailed analysis of the data in the future.

ACKNOWLEDGMENTS

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

[47] http://ckcc.huygens.knaw.nl
[48] http://www.europeana-collections-1914-1918.eu
[49] http://www.1914-1918-online.net
[50] http://ww1.discovery.ac.uk
[51] http://www.canadiana.ca/en/pcdhn-lod/
[52] http://www.cendari.eu/research/first-world-war-studies/
[53] http://blog.muninn-project.org
[54] http://www.bygle.net/wp-content/uploads/2015/04/Open-Memory-Project_3-1.pdf
[55] https://www.oorlogsbronnen.nl

REFERENCES

[1] David Allemang and Jim Hendler. 2008. *Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL.* Morgan Kaufmann, San Francisco, CA, USA.

[2] T. Collins, P. Mulholland, and Z. Zdrahal. 2005. Semantic Browsing of Digital Collections. In *Proceedings of the 4th International Semantic Web Conference (ISWC 2005).* Springer–Verlag, Berlin, Heidelberg, 127–141.

[3] V. de Boer, J. van Doornik, L. Buitinck, M. Marx, and T. Veken. 2013. Linking the kingdom: enriched access to a historiographical text. In *Proceedings of the 7th International Conference on Knowledge Capture (KCAP 2013).* ACM, New York, NY, USA, 17–24. https://doi.org/10.1145/2479832.2479849

[4] Marti Hearst. 2006. Design recommendations for hierarchical faceted search interfaces. In *ACM SIGIR workshop on faceted search.* Seattle, WA.

[5] T. Heath and C. Bizer. 2011. *Linked Data: Evolving the Web into a Global Data Space (1st edition).* Morgan & Claypool, Palo Alto, California. http://linkeddatabook.com/editions/1.0/

[6] Pascal Hitzler, Markus Krötzsch, and Sebastian Rudolph. 2009. *Foundations of Semantic Web technologies.* Chapman & Hall/CRC, London, UK.

[7] Eero Hyvönen. 2012. *Publishing and Using Cultural Heritage Linked Data on the Semantic Web.* Morgan & Claypool, Palo Alto, California.

[8] Eero Hyvönen, Jouni Tuominen, Miika Alonen, and Eetu Mäkelä. 2014. Linked Data Finland: A 7-star Model and Platform for Publishing and Re-using Linked Datasets. In *The Semantic Web: ESWC 2014 Satellite Events.* Springer–Verlag, Berlin, Heidelberg, 226–230.

[9] E. Hyvönen, K. Viljanen, J. Tuominen, and K. Seppälä. 2008. Building a National Semantic Web Ontology and Ontology Service Infrastructure—The FinnONTO Approach. In *Proceedings of the 5th European Semantic Web Conference (ESWC 2008).* Springer–Verlag, Berlin, Heidelberg, 95–109.

[10] Eero Hyvönen. 2020. "Sampo" Model and Semantic Portals for Digital Humanities on the Semantic Web. In *Proceedings of the Digital Humanities in the Nordic Countries. 5th Conference.* CEUR Workshop Proceedings, Aachen, 373–378. http://ceur-ws.org/Vol-2612/ Vol-2612.

[11] Eero Hyvönen. 2020. Using the Semantic Web in Digital Humanities: Shift from Data Publishing to Data-analysis and Serendipitous Knowledge Discovery. *Semantic Web – Interoperability, Usability, Applicability* 11, 1 (2020), 187–19.

[12] 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).* Springer–Verlag, Berlin, Heidelberg, 758–773.

[13] 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).* Springer–Verlag, Berlin, Heidelberg.

[14] Esko Ikkala, Eero Hyvönen, Heikki Rantala, and Mikko Koho. 2021. Sampo-UI: A Full Stack JavaScript Framework for Developing Semantic Portal User Interfaces. *Semantic Web – Interoperability, Usability, Applicability* (2021). https://doi.org/10.3233/SW-210428 To appear. Available from http://www.semantic-web-journal.net/.

[15] Matti Karjalainen. 2014. Tietojen vertailua, korjaamista ja asiakaspalvelua – Suomen sodissa 1939–1945 menehtyneiden tietokanta. In *Kleion pauloissa*, Päivi Happonen, John Strömberg, and Aino Heikkinen (Eds.). Suomalaisen Kirjallisuuden Seura, Helsinki, 277–283.

[16] 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)* (Rome, Italy), Vol. 2375. CEUR Workshop Proceedings, Aachen, 91–96. Vol 2375.

[17] Harold A. Linstone. 1989. Multiple perspectives: Concept, applications, and user guidelines. *Systems practice* 2, 3 (1989), 307–331. https://doi.org/10.1007/BF01059977

[18] Gary Marchionini. 2006. Exploratory search: from finding to understanding. *Commun. ACM* 49, 4 (2006), 41–46.

[19] Eetu Mäkelä, 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 (2017), 333–343.

[20] Heikki Rantala, Esko Ikkala, Ilkka Jokipii, Mikko Koho, Jouni Tuominen, and Eero Hyvönen. 2020. WarVictimSampo 1914-1922: A Semantic Portal and Linked Data Service for Digital Humanities Research on War History. In *The Semantic Web: ESWC 2020 Satellite Events (Lecture Notes in Computer Science, Vol. 12124).* Springer-Verlag, Berlin, Hedelberg, 191–196. https://doi.org/10.1007/978-3-030-62327-2_33

[21] Heikki Rantala, Ilkka Jokipii, Mikko Koho, Esko Ikkala, Jouni Tuominen, and Eero Hyvönen. 2020. Building a Linked Open Data Portal of War Victims in Finland 1914–1922. In *Proceedings of Digital Humanities in Nordic Countries (DHN 2020).* CEUR Workshop Proceedings, Aachen, 310–317.

[22] Walter Ravenek, Charles van den Heuvel, and Guido Gerritsen. 2017. The ePistolarium: Origins and Techniques. In *CLARIN in the Low Countries*, Arjan van Hessen and Jan Odijk (Eds.). Ubiquity Press, London, United Kingdom, 317–323. https://doi.org/10.5334/bbi

[23] Laurens Rietveld and Rinke Hoekstra. 2017. The YASGUI family of SPARQL clients. *Semantic Web* 8, 3 (2017), 373–383.

[24] Staab S. and Studer R. (Eds.). 2009. *Handbook on Ontologies (2nd Edition).* Springer–Verlag, Berlin, Heidelberg.

[25] D. Tunkelang. 2009. *Faceted search.* Morgan & Claypool, Palo Alto, California.

[26] Marco H. D. Van Leeuwen and Ineke Maas. 2011. *HISCLASS: A historical international social class scheme.* Leuven University Press, Leuven, Belgium.

[27] Marco H. D. Van Leeuwen, Ineke Maas, and Andrew Miles. 2002. *HISCO: Historical international standard classification of occupations.* Leuven University Press, Leuven, Belgium.

[28] Koenraad Verboven, Myriam Carlier, and Jan Dumolyn. 2007. A short manual to the art of prosopography. In *Prosopography approaches and applications. A handbook.* Unit for Prosopographical Research (Linacre College), Oxford, United Kingdom, 35–70.

[29] Lars Westerlund (Ed.). 2004. *Sotaoloissa vuosina 1914—22 surmansa saaneet. Tilastoraportti.* Valtioneuvoston kanslia, Helsinki, Finland.