Military History on the Semantic Web: Lessons Learned from Developing Three In-use Linked Open Data Services and Semantic Portals for Digital Humanities

Eero Hyvönen

Aalto University, Department of Computer Science, Semantic Computing Reserarch Group (SeCo) and University of Helsinki, Helsinki Centre for Digital Humanities (HELDIG)

ARTICLE HISTORY
Compiled June 24, 2024

ABSTRACT
This paper reviews three in-use systems for publishing and studying military history on the Semantic Web: WarSampo, WarWictimSampo 1914–1922, and WarMemoirSampo. Targeted to Digital Humanities researchers, application developers, and the public, these systems include a Linked Open Data (LOD) service with a live SPARQL endpoint that can be used with modest programming skills, and a ready-to-use semantic portal on top of it. Lessons learned in developing these systems are discussed based on the so-called Sampo Model, a set of principles evolved during the work for creating and publishing Knowledge Graphs (KG) as LOD services and semantic portal User Interfaces (UI). The UIs support semantic faceted search, data browsing and exploration, as well as seamlessly integrated tools for data analyses.

KEYWORDS
digital humanities; military history; semantic web; linked open data; portal; data service; data analysis

1. Introduction

Georg Friedrich Hegel (1770–1831) has maintained that *we learn from the history that we learn nothing from the history*. He seems to be right, given, e.g., the currently raging war in Ukraine not long after the Second World War (WW2), arguable the most devastating catastrophe of human history—and the First World War only twenty years before that. In spite of this, the work discussed in the paper maintains that *the more we learn about the price of the war, the less there will be wars*. It is argued that the evolving Semantic Web (Bizer, Heath, & Berners-Lee, 2009; Hitzler, 2021) and Linked Open Data (LOD) technologies (Heath & Bizer, 2011; Hyvönen, 2012) offer a promising way to expose data about the wars, such as the deaths, to the researchers and the public to study and learn.

As a proof of concept, a series of three LOD services and semantic portals listed in Table 1 have been designed, implemented, and published for public use in Finland on the Semantic Web:

(1) WarSampo aggregates and publishes data about the Second World War (WW2)

CONTACT Eero Hyvönen Email: eero.hyvonen@aalto.fi; homepage: http://seco.cs.aalto.fi/u/eahyvone/

Table 1. Three Sampo portals and LOD services for Digital Humanities on Military History; distinct user counts, based on site visits, using Google Analytics by July 2023

Portal	**Publish**	**Domain**	**# Users**	**# Triples**	**Data owners**
WarSampo[a]	2015–2019	World War II	1 200 000	14M	National Archives, Defense Forces, and others, Finland
WarVictimSampo 1914–1922[b]	2019	Finnish Civil and Kindred wars	53 000	10M	National Archives of Finland
WarMemoirSampo[c]	2021	War memoirs on videos	4400	0.32 M	Tammenlehvän Perinneiitto ry and National Archives, Finland

[a]Portal: `https://sotasampo.fi`; project: `https://seco.cs.aalto.fi/projects/sotasampo/`

[b]Portal: `https://sotasurmat.narc.fi`; project: `https://seco.cs.aalto.fi/projects/sotasurmat-1914-1922/`

[c]Portal: `https://sotamuistot.narc.fi`; project: `https://seco.cs.aalto.fi/projects/war-memoirs/`

in Finland from some 20 data sources and several collaborating organizations.

(2) WarVictimSampo 1914–1922 is a related system based on the death records and battles of the Civil War in Finland and Kindred Wars during 1914-1922, with a focus on data analytic Digital Humanities research.

(3) WarMemoirSampo demonstrates a novel way of publishing and watching videos on the Semantic Web, with a focus on publishing memoirs of WW2 veterans on the Semantic Web.

These prototype portals have become quite popular in Finland suggesting feasibility of the approach. For example, WarSampo has had 1.2 million distinct users in a small country like Finland of only 5.5 million inhabitants. WarSampo and WarVictimSampo 1914–1922 are based on death records of their topical wars that were originally provided by the National Archives of Finland via separate online services. In 2022, the provision of these legacy systems was terminated, and their users were directed to using the corresponding Sampo systems.

The three portals of the table share not only their domain of discourse, military history, but also the underlying semantic infrastructure, the model of aggregating and publishing LOD as a service, and the user interface (UI) logic of the portals. These general commonalities, based on the *Sampo Model* (Hyvönen, 2023a) and the *Sampo-UI Framework* (Ikkala, Hyvönen, Rantala, & Koho, 2022; Rantala, Ahola, Ikkala, & Hyvönen, 2023), as well as on the Finnish LOD infrastructure (Hyvönen, 2023b), are the topics of this paper. It is suggested that the model and the tools presented could be of use for new similar applications and beyond in other countries, too.

In the following, these three systems on military history are first briefly introduced (Section 2). After this lessons learned in creating and publishing LOD on the Semantic Web are discussed (Section 3), followed by issues related to UI design (Section 4). It is shown, how the principles of the Sampo Model have been applied to the three portals. In conclusion (Section 5), related works are discussed, contributions summarized, and directions for further research suggested.

2. Three LOD Services and Semantic Portals for Military History

The applications of Table 1 demonstrate how to aggregate and enrich heterogeneous, distributed datasets into harmonized Knowledge Graphs (KG), based on a shared ontology infrastructure. In this paper, a knowledge graph is defined as a semantic network that represents real-world entities (objects, concepts, events, etc.) based on an ontological data model. The KGs can be published as LOD services and then be used for 1) data analyses in Digital Humanities (DH) with tools such as the Yasgui editor[1] (Rietveld & Hoekstra, 2017), Google Colab[2], and Jupyter Notebooks[3], and for 3) developing ready-to-use applications. The Sampo portals, where faceted search and browsing are integrated seamlessly with data analytic tools, test and demonstrate the application use.

2.1. WarSampo – Finnish Second World War on the Semantic Web

WarSampo (Hyvönen et al., 2016) is a system for publishing and studying collections of heterogeneous, distributed data about the Second World War in Finland on the Semantic Web. Its KG (Koho et al., 2021) is based on aggregating and harmonizing some twenty war-related data sources, which makes it possible to enrich the datasets semantically with each other's contents. The key dataset originates from a database of some 95 000 WW2 death records of the National Archives of Finland; rich metadata about everybody killed in action during the war in Finland is included in the data.

As customary in Sampo systems, WarSampo has two components: First, there is a LOD service for DH research and for creating web applications. Second, a semantic portal has been created to test and demonstrate the usability of the data service. The portal allows both historians and laymen to study military history and destinies of the soldiers in the war from different interlinked perspectives. The data model used is event-based extending the CIDOC CRM standard[4] to event types of war (e.g., battle, bombardment, etc.) and personal history (e.g., getting wounded, getting killed, promotion in rank, etc.).

WarSampo was published first in 2015 with six application perspectives into the data: Events, Persons, Army Units, Places, Casualties, and Memoir Articles. After this, new datasets and perspectives have been added into the system, including a search view for 160 000 authentic photographs provided by the Defence Forces of Finland, a perspective about over 672 war cemeteries (Ikkala et al., 2017), and yet another one about the 4200 Finnish prisoners of war in the Soviet Union 1939–1945 (Koho, Ikkala, & Hyvönen, 2020).

A key innovation of WarSampo is to try to re-assemble the life stories of the soldiers automatically by data linking. For example, if one's relative was killed in action, he can be found in the death records of the KG telling, e.g., in which army unit he served. For army units, data about their battles, movements, etc. can be found and therefore also the actions of a person as part of the unit, assuming that he was serving in the unit. The events related to a person can be illustrated on maps and timelines. Additional data, such as photographs, can be linked not only to persons and units mentioned, but also to events based on places and times mentioned in the metadata using Named Entity Linking (Heino et al., 2017). Furthermore, links to the actual war diaries can be provided based on the army unit data. In addition to reassembling personal war histories, the data has also be used for data analyses (Koho & Hyvönen, 2023; Koho et al., 2017).

[1] https://yasgui.triply.cc
[2] https://colab.research.google.com/notebooks/intro.ipynb
[3] https://jupyter.org
[4] https://www.cidoc-crm.org/

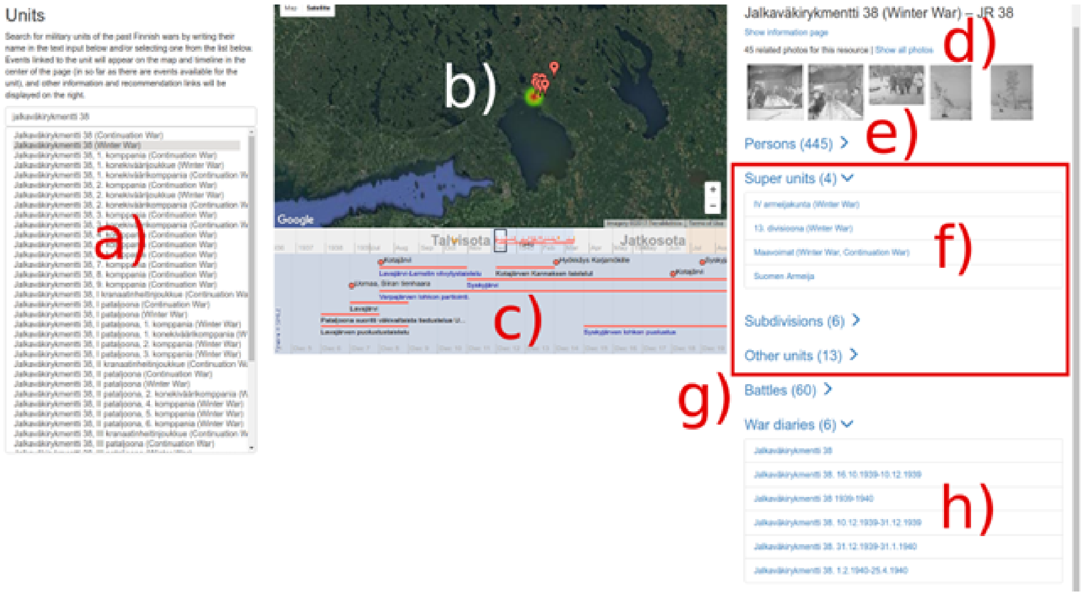

Figure 1. A view from the Army Units application perspective of WarSampo showing aggregated spatio-temporal data about the history of the Infantry Regiment 38.

For example, Figure 1 illustrates data about a unit selected on the left (a). Related events are shown on a timeline (c), and those in the particular time window in the middle are visualized on the map (b) with a heat map about the casualties within the currently visible time window. On the right, information regarding the unit is shown, including photographs related to it (d). Additional links are shown to 445 persons in the unit (e), related units (f), battles (g), and authentic war diaries (h) of the unit for primary source data.

2.2. WarVictimSampo – Casualties of Finnish wars 1914–1922

WarVictimSampo 1914–1922 is a follow-up system of WarSampo, a semantic portal and LOD service about the war victims, battles, and prisoner camps in the Finnish Civil War in 1918 and casualties of the Kindred Wars (Rantala et al., 2020; Rantala, Jokipii, Ikkala, & Hyvönen, 2022). The system contains detailed information about some 40 000 deaths extracted from several data sources, data about over 1000 battles of the Civil War between the “Reds” and “Whites”, and some additional related data created, compiled, and linked during the project.

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. This allows, e.g., studying data about wider prosopographical groups in addition to individual war victims.

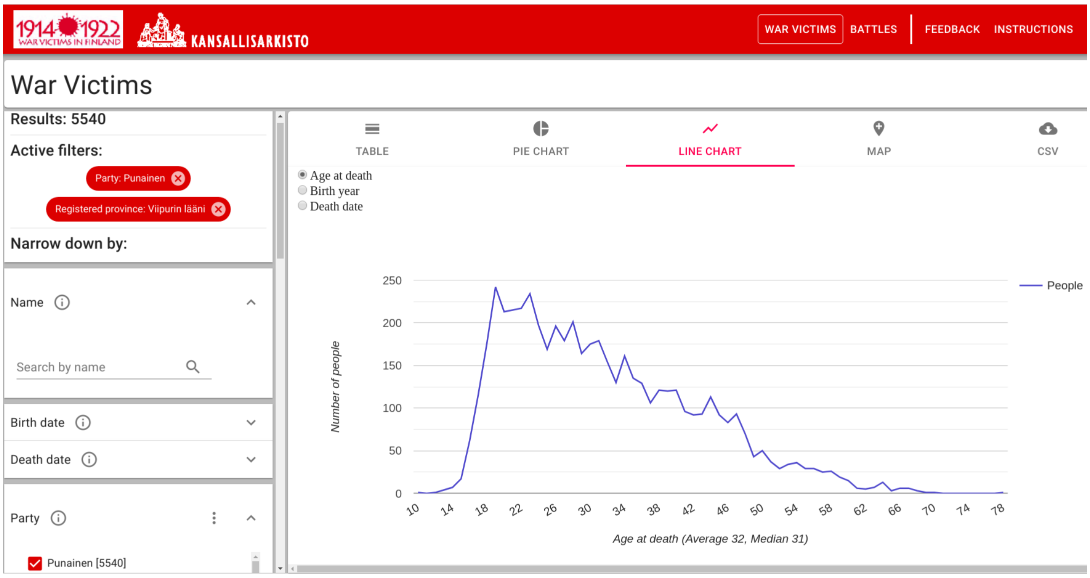

Figure 2. The age distribution of people who supported the Reds side in the Finnish Civil War and who were from the Viipuri Province as shown in WarVictimSampo 1914–1922 portal. The LINE CHART tab can be used for visualizing the birth and death years, too.

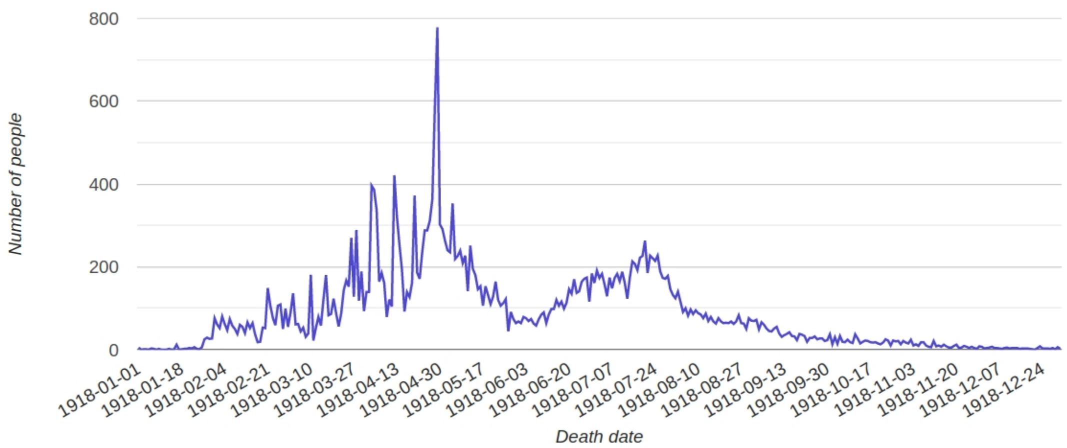

Figure 3. Distribution of death dates in 1918 in the knowledge graph of WarVictimSampo 1914–22

The portal includes two application perspectives, one for studying the victims of war and one for the Civil War battles. As customary in the Sampo model these perspectives correspond to classes in the underlying KG and allow searching, browsing, and analyzing their individuals, here victims and battles. Figure 2 illustrates how the victims can be studied. On the left there are 21 facets for filtering out the casualties; only four first ones are visible in the figure due to limited space, i.e., Name, Birth date, Death date, and Party. The user has made two selections: Party = 'Reds', and the Place of registry = 'Viipuri province'. The result set can then be studied and visualized on the right using five tabs that can be selected form the top: TABLE lists the filtered people, PIE CHART shows their distributions along different facets, LINE CHART visualizes the people using a few numerical property values, the MAP tab shows the deaths of the selected people on a map, the CVS tab makes it possible to download the results in CSV form. The last option was considered important by history researchers as it allows to re-use the data in various other tools, such as spreadsheet programs.

In the Figure 2, the line chart visualization was selected for the filtered victims. Currently there are three different options for the x-axes of this visualization: Age at death, Birth year, and Death date. For example, the Age at death option was selected on the figure drawing automatically a line chart where the x-axis represents age in years and y-axis represents the number of victims.

The average and median values are also shown under the visualization graph. The user can then easily and quickly compare distributions and average values between certain subsets of people in the data. For example, in Figure 3 we have used the facets to select victims who supported the Reds side in the Civil War and who were registered to the Viipuri province. Comparing this distribution to people from other provinces shows that war victims who supported the Reds side of the Civil from the Viipuri Province tended to be older than others according to the data, with the median age of over thirty years. Explaining this would require more detailed analysis and close reading the data, but this demonstrates how faceted search combined with data analytic tools can be effective at finding interesting phenomena in the data.

The MAP tab view shows the death places of the victims 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. A victim's information is not shown on the map if there is no death geo-data available about him/her.

The other perspective of the portal is the Battles of the Finnish Civil War. This application perspective works in a way similar to the war victims perspective. The user can search and filter the battles using facets and the results can be visualized in different ways on tabs, in this case as a table, on a map, or as an animation. The temporal animation visualization is unique to the battles. It shows the battle sites at different times on a map as markers. A marker appears on the map as red when the animation reaches the starting date of the battle and then stays on the map as grey turning into black when the animation progresses in time. Figure 4 shows the situation on March 10, 1918. You can see how a front line has been formed across Finland from east to west, and how the battles are raging along it.

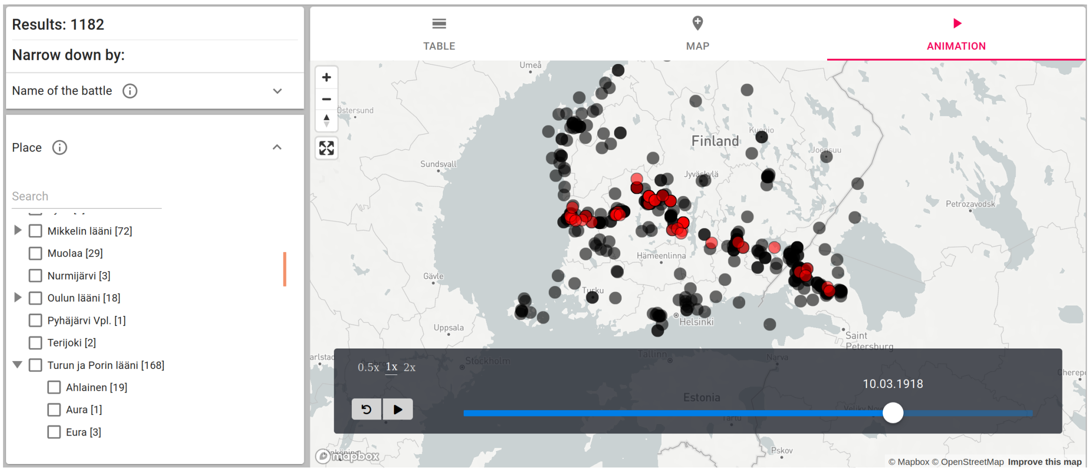

Figure 4. The battles animation view of the WarVictimSampo 1914–1922 portal, stopped on March 10, 1918. A front line can be seen as red markers from east to west. The battles can be filtered using the facets on the left.

2.3. WarMemoirSampo – Memoir Interviews of WW2 Veterans

WarMemoirSampo focuses on preserving and publishing memoirs of the WW2 veterans for the coming generations (Hyvönen et al., 2022). The dataset here is a set of video interviews created and stored by the Finnish WW2 veteran association *Tammenlehvän perinneliitto* and the National Archives of Finland.

A technical novelty of the system is to enable scene segments in videos to be searched by their semantic textual content (Koho et al., 2022; Leal et al., 2022). This makes is possible, for example, to find particular points in long videos, where a person, place or other entity is mentioned. Another key idea of WarMemoirSampo is to enhance video watching experience by data linking: while watching a video, additional contextual information is provided dynamically based on the underlying LOD (Hyvönen et al., 2022).

The system is based on and demonstrates the idea of re-using the WarSampo infrastructure and KG in other applications. The LOD for WarMemoirSampo has been extracted automatically from timestamped textual natural language descriptions of the video contents; the data is interlinked not only internally but also externally with the WarSampo KG.

Based on the Sampo model and Sampo-UI framework, the landing page of WarMemoirSampo portal provides three application perspectives to the underlying KG with faceted semantic search:

(1) *Interviews perspective* is used for searching whole videos based on their nine key properties: Interview content (text facet for traditional search), Interviewee, Interviewee Gender, and mentioned Places, Persons, Military units, Organizations, Events, Other entities, and Topic.
(2) *Scenes perspective* is used for searching video scenes inside interviews using the same facets as in the Interviews perspective.
(3) *Directory perspective* contains all ca. 3000 entities mentioned in the texts with direct links to the scenes where the entities were mentioned. It is a kind of semantic index of the underlying dataset.

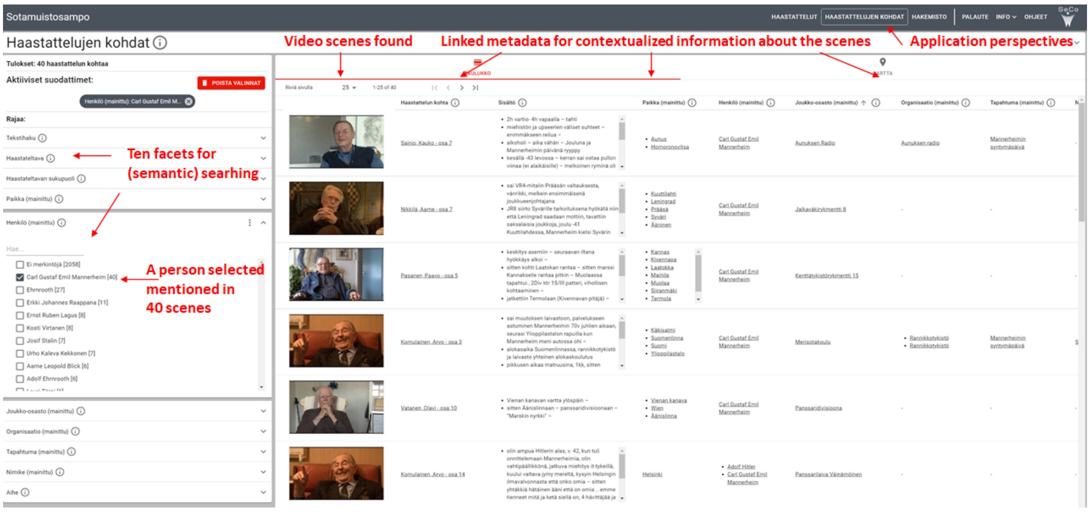

Figure 5. Faceted search for scenes inside videos.

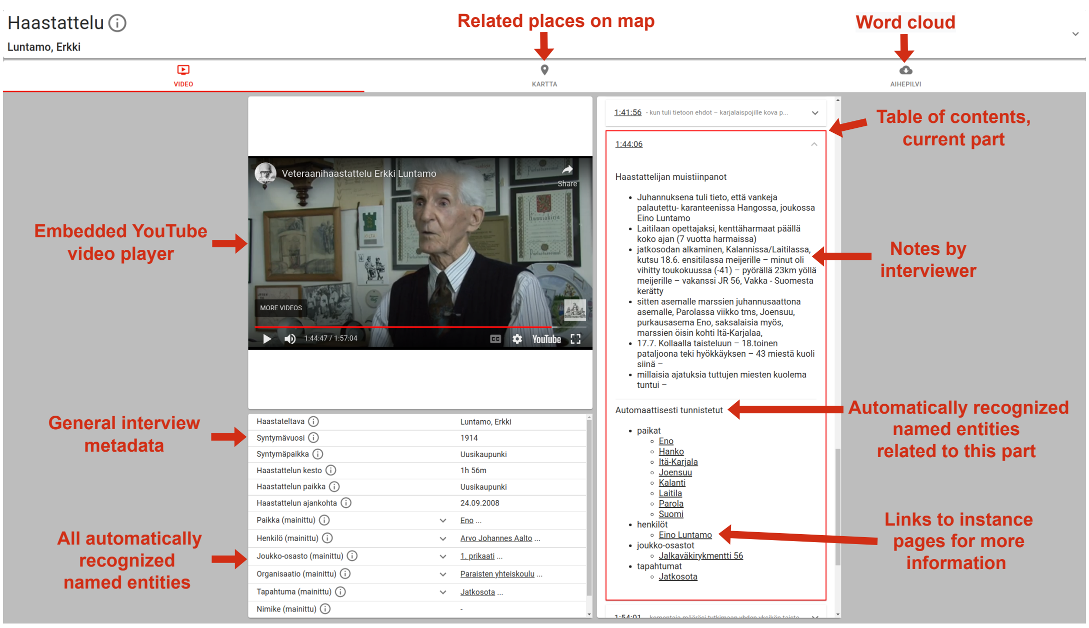

Figure 6. Video viewing page with a dynamic table of contents for contextual linked data.

Fig. 5 depicts the Scenes perspective, where the user has selected "Carl Gustav Emil Mannerheim" on the facet Person: the 40 scenes mentioning this marshal are shown on the right with metadata links for further information. By clicking on a video, it is opened for dynamic viewing at the right point as depicted in Fig. 6. Links to additional information are provided on the fly. Selecting the tab Map on top shows the places mentioned on a map; Fig. 7 shows all 4566 of them. A click on a marker on the map opens a pop-up window with links to all related scenes. Finally, the tab Word Cloud summarizes the topics of the video interview as a word cloud based on the extracted subject matter concepts in the KG. The concepts were extracted by an automatical

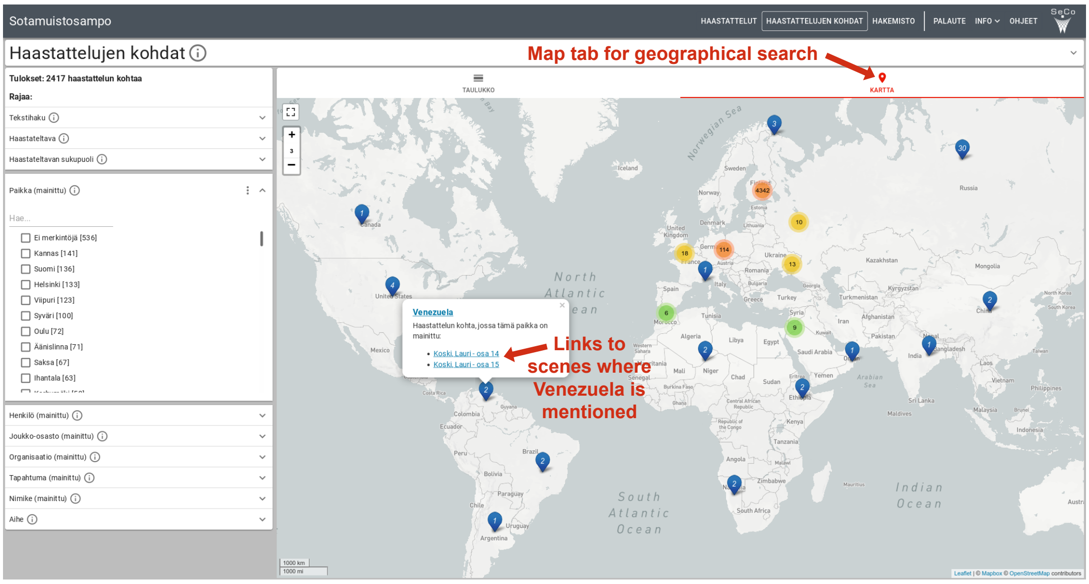

Figure 7. Map view to access scenes that mention a place.

annotation tool (Leal et al., 2022) using the subject matter ontologies of the national Finnish ontology infrastructure (Hyvönen, 2023b).

3. Lessons Learned: Creating and Publishing Cultural Heritage LOD

Since 2002, the Semantic Computing Research Group (SeCo)[5] at the Aalto University and University of Helsinki has been involved in creating a national semantic web infrastructure. To test and demonstrate its usability, over twenty LOD services and semantic portals based on them have been created, mostly in the domain of Cultural Heritage (CH) and DH. During this work, the experiences have gradually evolved into the Sampo model, a set a general principles on 1) how to create LOD services and 2) user interfaces on top of them.

Regarding LOD creation there are three major principles P1–P3 in the model.

3.1. Support collaborative data creation and publishing (PI)

Leonardo da Vinci has maintained[6]: *Learn how to see. Realize the everything connects to everything else.* This wisdom applies well to the general idea of LOD where mutually interlinked aggregated datasets are used to enrich each other. When creating search and data exploration systems based on data aggregation, there are two basic approaches available.

(1) *Distributed strategy*: federated search. The traditional way is to take the user's query, send it to distributed local data services hosting the data to be aggregated, collect the answers, and present them to the user.

[5]SeCo homepage: `https://seco.cs.aalto.fi`
[6]`https://philosophynow.org/issues/134/The_Mind_of_Leonardo_da_Vinci`

(2) *Centralized strategy*: aggregating global data. The other approach is to aggregate and harmonize the distributed heterogeneous datasets first into a global database or KG, and apply the query to its centralized data service.

In the distributed strategy, the burden of figuring out what the user wants can be distributed to the local data providers that transform the query for their local databases. Also the burden of actually executing the query can be distributed. However, it is difficult to transform the query and present results in a semantically interoperable way in local services whose data models and vocabularies[7] used in the metadata are different. This deteriorates precision and recall, and makes data-analyses challenging. For example, entities, such as persons and places are typically represented in different ways locally and therefore confused with each other. Furthermore, not having simultaneous access to the global data is a severe restriction on what can be analyzed from the global data. For example, finding out relations between entities in local datasets is hard. In the Sampo model, the centralized strategy was therefore selected, introduced already in the first Sampo system MuseumFinland (Hyvönen et al., 2005) (online since 2004). However, using the global strategy brings in its own challenges. These regard especially data model harmonization of the local datasets and disambiguating and linking the data instances for semantic interoperability. However, these challenges are not due to the centralized strategy, but to the heterogeneity of the local datasets and the ways they are created, and have to be addressed in any case when dealing with local data in a semantically proper interoperable way.

3.2. Use a shared open ontology infrastructure (P2)

Albert Einstein has said: *Intellectual solve problems. Geniuses prevent them.* This wisdom applies well to the idea of developing and using an infrastructure in creating CH and DH applications (Hyvönen, 2023b): it is arguably better to prevent interoperability problems already when creating data than fix problems afterwards when aggregating data (Hyvönen, 2010). According to our experiences, most of the time is "wasted" in projects like the Sampo systems in cleaning, aligning, and harmonizing distributed local datasets, problems that could have been avoided by using shared data models and vocabularies in populating them in databases in the first place.

Our work on developing the Sampos has therefore been supported by a systematic effort on creating a national semantic web infrastructure for vocabularies and by using standard data models, such as CIDOC CRM and Dublin Core. For example, ontologies developed from national keyword thesauri as well and ontologies for, e.g., historical places, actors, and occupations have been re-used and developed further step-by-step in the Sampo systems. This work started in 2003 as the series of FinnONTO projects[8] in which, e.g., a LOD cloud called KOKO of interlinked domain-specific ontologies was created (Frosterus, Tuominen, Pessala, & Hyvönen, 2015; Hyvönen, Viljanen, Tuominen, & Seppälä, 2008) and published as a national ontology service ONKI.fi (Tuominen, Frosterus, Viljanen, & Hyvönen, 2009), deployed by the National Library of Finland in 2014 as the contemporary Finto.fi service (Suominen et al., 2014). This work is now continuing as part of the national FIN-CLARIAH initiative for DH re-

[7]In this paper the term *vocabulary* is used to refer to (hierarchical) knowledge organization systems, such as thesauri, authority files, and geographical gazetteers, whose entries are used to fill in metadata element (property) values.

[8]FinnONTO project series: `https://seco.cs.aalto.fi/projects/finnonto/`

search infrastructures in Finland[9], with a goal to combine the European CLARIN[10] and DARIAH[11] infrastructures on a national level.

3.3. Make clear distinction between the LOD service and the user interface (UI) (P3)

This principle was tested first when developing the ontoly service ONKI Light for SKOS vocabularies (Suominen, Johansson, Ylikotila, Tuominen, & Hyvönen, 2012): is it possible to re-implement the original ONKI.fi ontology services (Tuominen et al., 2009; Viljanen, Tuominen, & Hyvönen, 2009) by using SPARQL queries only for data access? The answer was “yes”, and the Finto.fi ontology service was deployed based on ONKI Light. Another related step was to test whether it makes sense to apply this idea to implement faceted semantic search, too, used in the early Sampo systems starting from MuseumFinland. The answer was “yes” again, and this development lead to developing the tools SPAQRL Faceter (Koho, Heino, & Hyvönen, 2016), used in WarSampo, and later Sampo-UI in 2018. By 2023, Sampo-UI has been used in some 15 Sampos including WarVictimSampo 1914–1922 and WarMemoirSampo.

3.4. Maintaining Linked Open Data and Data Services

According to Heraclitus (fl. 500 BC) *everything changes and nothing remains still; and you cannot step twice into the same stream.* An important issue of using LOD is maintaining changes in the KG as time goes by and software evolves. However, the Sampo principles discussed above focus only on how to create and publish a LOD service.

A piece of good news regarding the challenges of change is that linked data formats are open, standardized by W3C recommendations, and are based on text. The data is therefore pretty sustainable and re-usable, but tools, such as triple stores and UI frameworks that use the data change more often and may support and extend the standards, such as the SPARQL query language, in different ways.

A more severe challenge is what to do, when either the metadata models (Zeng & Qin, 2022), vocabularies used in populating the models, and the data itself evolves. For example, the WarSampo KG has been extended with new data as new application perspectives were developed. Then new vocabularies or entries in old ones may be needed, and the entities in the new data need to be aligned with those already there in the underlying infrastructure. This problem is discussed, e.g., in (Koho, Ikkala, Heino, & Hyvönen, 2018; Koho, Ikkala, & Hyvönen, 2018).

There are two basic approaches depending on how the primary data is managed. If the data is maintained in a legacy system using traditional formats, it makes sense to design the LOD transformation in such a way that it can be re-run automatically from scratch. This means that there should preferably be no intermediate manual phases in the process, as their results would be wiped away by when the KG is reconstructed. The challenge in this approach is that the new data is likely to contain typos and linking textual descriptions may need manual work and fixes after all. For finding quality issues in the linked data, semantic validation languages and frameworks, such

[9]FIN-CLARIAH initiative LOD work package: `https://seco.cs.aalto.fi/projects/fin-clariah/`
[10]CLARIN infrastructure: `https://www.clarin.eu`
[11]DARIAH infrastructure: `https://www.dariah.eu/`

as SHACL[12] and ShEx[13] can be used.

A better way would be managing the KG in native linked data form. This would keep the data automatically consistent and ready to be uploaded into a LOD service. Unfortunately, there are still few tools for editing and managing RDF data. An exception to this are ontology editors, such as Protege[14] and Topbraid Composer[15]. In the case of the Sampo systems, the SPARQL SAHA editor (Mäkelä & Hyvönen, 2014) was developed and has been used in maintaining, e.g., the BookSampo KG (Mäkelä, Hypén, & Hyvönen, 2013) by the data owners, i.e., the public libraries of Finland. SAHA is also used for maintaining the RDF data of OperaSampo on historical opera and musical performances (Ahola, Hyvönen, Rantala, & Kauppala, 2023) by its data owner, the Sibelius Academy.

4. Lessons Learned: User Interfaces for Digital Humanities Research

After a LOD service has been established its data can be used in two ways:

(1) *Using Application Programming Interfaces (API)*. The LOD publication methodology provides different ways to access the data: 1) The data can be downloaded from the service as data dumps. 2) The data can be browsed in a human readable way using a linked data browser[16]. 3) The LOD service provides content negotiation where URIs can be resolved and either data for the machine or HTML for the human user can be returned[17]. 4) Most importantly, the data service can be queried in flexible ways using the SPARQL query language[18] and endpoint. There are easy to use tools, such as Yasgui (Rietveld & Hoekstra, 2017), for editing and executing SPARQL queries with some built-in visualization options for the results. The SPARQL endpoint can be accessed from any programming environment, such as Jupyter notebooks and Python scripting for querying and analyzing data.

(2) *Using portals and other applications*. Ready-to-use applications for accessing and using the data without programming skills can be developed on top of the LOD service, as exemplified by the Sampo portal series.

In the following some lessons learned are discussed in developing UIs for the three portals of this paper, as formulated in the Sampo model the Sampo-UI framework.

Fig. 8 illustrates the navigational structure of using a Sampo-UI-based portal. The user first lands on the *landing page* with several *application perspectives* to the data. The perspectives are based on classes of the underlying KG, such as Artefacts, Persons, Places, etc. The usage cycle of each perspective can be divided into two steps: 1) filter and 2) analyze. The user first filters the data by using the faceted semantic search (Hearst, 2006; Tunkelang, 2009) tools provided by the portal. The results as well as the facet option hit counts are updated after each category selection on a facet, making it possible for the user to precisely filter the end-result entities by different aspects, e.g., filtering by party and registration municipality as shown in Figure 2.

[12]https://www.w3.org/TR/shacl/
[13]https://shex.io/
[14]https://protege.stanford.edu/
[15]https://allegrograph.com/topbraid-composer/
[16]See, e.g., the browser for DBpedia: https://dbpedia.org/ontology/Browser
[17]Content Negotiation by Profile: https://www.w3.org/TR/dx-prof-conneg/
[18]SPARQL 1.1 Query Language: https://www.w3.org/TR/sparql11-query/

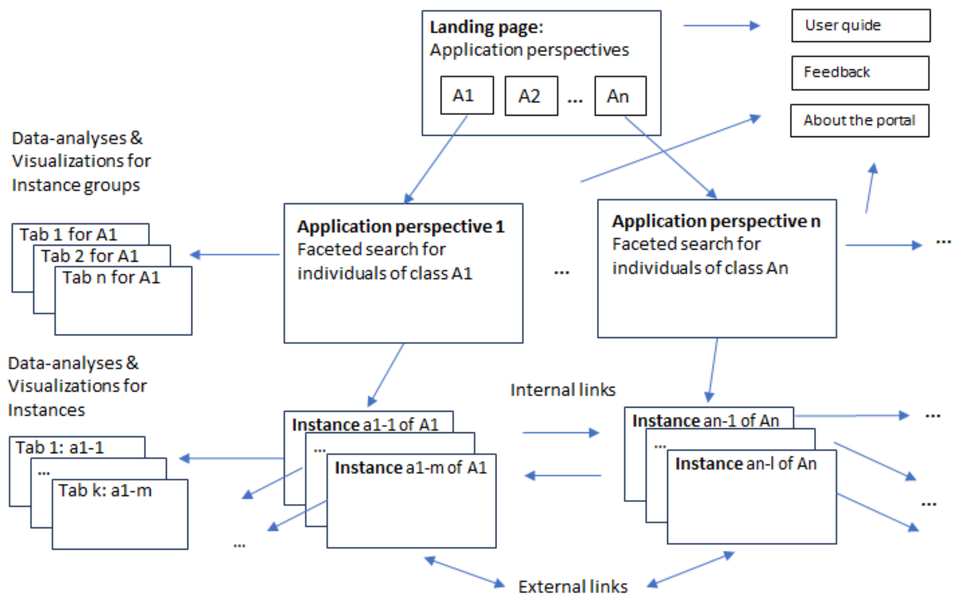

Figure 8. Navigational page structure of a portal based on Sampo-UI.

In faceted search, the hit counts direct the search and prevent ending up in dead-end situations where no results are found. Faceted search was developed already in the 90's and early 00's but under the name "view-based search" (Hyvönen, Saarela, & Viljanen, 2004; Pollitt, 1998) and also as "dynamic taxonomies" (Sacco, 2005).

After filtering the data to the wanted subset, the *target group*, the user can analyze the results set, i.e., a set of instances of the class corresponding to the application perspective, with integrated data-analytic tools available as tabs on the application perspective page. An example of a visualization can be seen in Figure 2.

It is also possible to select a particular instance of the result set for a closer look: each instance has an *instance page* that provides aggregated information about the individual with internal and external links for further information to browse. In addition, instance pages also may have a set of tabs that provide contextualized data-analyses of the individuals in the same way as for target groups. For example, BiographySampo (Hyvönen et al., 2019) and AcademySampo (Leskinen, Rantala, & Hyvönen, 2022) contain an application perspective corresponding to the class Person. Visualizations for studying prosopographical target groups (e.g., persons with the same occupation or place of birth) are available on different tabs of the application perspective page, and on instance pages of each particular person there are tabs for, e.g., studying ego-centric networks of individuals.

This filter-analyze two-step usage cycle allows an iterative approach to exploring the data (Marchionini, 2006; Tzitzikas, Manolis, & Papadakos, 2017). It is possible to find potentially interesting subsets and individuals in the data without having to be already familiar with the content. By providing a text facet, it is also possible to support use cases where the user is looking for a specific instance, say a person with a known name, and can formulate the search query easily.

The UI logic above is based on the three principles P4–P6 articulated in the Sampo model:

(1) *Provide multiple perspectives to the same data (P4)* The idea here is the same as in the FAIR principles[19], but adapted to UI design: reusing the data even within one UI. The class structure of KGs provide for this a natural approach: classes (e.g., Person, Place, Battle, etc.) can be used as a basis for searching their individuals (people places, battles, etc) in the application perspectives. In our case, WarSampo has nine application perspectives, WarVictimSampo 1914–1922 two, and WarMemoirSampo three.
(2) *Standardize portal usage by a simple filter-analyze two-step cycle (P5)* This idea, discussed above, was inspired by the prosopograhical research method on groups of people (Verboven, Carlier, & Dumolyn, 2007), where a target group of people sharing some common features is first filtered out and then analyzed in more detail.
(3) *Support data analysis and knowledge discovery in addition to data exploration (P6)* Finally, in addition to semantic faceted search and data exploration, one should consider providing the user with intelligent tools for analyzing the data, or even with intelligent agents trying to find interesting pattern of knowledge in the data by themselves, solving research problems, and explaining the results to the user, leading to "third generation" systems in DH (Hyvönen, 2020).

5. Contributions, Related Work, and Discussion

In this section, the overall contributions of using LOD in the three Sampo systems presented are summarized, followed by a review on related works. Finally, some critical issues on using LOD are discussed and a view to the future is given.

5.1. Contributions

There are three main technical reasons for using LOD in publishing and using contents about Military History, and more generally about Cultural Heritage:

(1) It is possible to enrich everybody's data collaboratively from separate data silos. Everybody can win by collaboration.
(2) By creating Findable, Accessible, Interoperable, Re-usable data, as suggested by the FAIR principles for scientific data management and stewardship, the value of data increases in the four FAIR dimensions.
(3) By using Semantic Web Linked Data semantics (Hitzler, Krötzsch, & Rudolph, 2010), based on first order logic, the machine can "understand" the data and, for example, enrich the data by reasoning and solve problems. This means that creating more intelligent applications for the public, curators, and researchers is possible and more cost-efficient.

From a DH perspective the idea of using LOD services and semantic portals on top of them is promising in filtering our patterns of possibly interesting phenomena in Big Data using distant reading (Moretti, 2013). However, also close reading can be supported for verifying and interpreting the data-analytic results.

[19]https://www.go-fair.org/fair-principles/

5.2. Related Works

There are several projects that have published linked data about wars, e.g., concerning the World War I: Europeana Collections 1914–1918[20], 1914–1918 Online[21], WW1 Discovery[22], Out of the Trenches[23], CENDARI[24], and Muninn[25]. Our work on WarSampo was inspired by our earlier system WW1LOD on WW1 data that was created in collaboration with the University of Colorado, Boulder (Mäkelä, Törnroos, Lindquist, & Hyvönen, 2017).

In addition to WarSampo, there are a few works that have used the Linked Data approach to WW2, such as (Collins, Mulholland, & Zdrahal, 2005; de Boer, van Doornik, Buitinck, Marx, & Veken, 2013), Open Memory Project[26] on holocaust victims, and the Dutch project Netwerk Orloogsbronnen[27]. In (Wang, 2023), knowledge graph technologies are studied for integrating heterogeneous data related to wars, in this case the Second Sino-Japanese war 1937–1945 as a case study.

The 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[28] and has been developed also in various other settings, e.g., in ResearchSpace[29]. The idea of providing multiple analyses and visualizations to a set of filtered search results has been used in other portals, such as the ePistolarium[30] (Ravenek, van den Heuvel, & Gerritsen, 2017) for epistolary data, and using multiple perspectives have been studied as an approach in decision making Linstone (1989). Faceted search (Hearst, 2006; Tunkelang, 2009), known also as view-based search (Hyvönen et al., 2004; Pollitt, 1998) and dynamic taxonomies (Sacco, 2005), is a well-known paradigm for explorative search and browsing (Marchionini, 2006) in computer science and information retrieval, based on S. R. Ranganathan's original ideas of faceted classification in Library Science in the 1920's (Ferreira, dos Santos Maculan, & Naves, 2017). The two step study model used in our work has been used, e.g., in prosopographical research (Verboven et al., 2007) (without the faceted search component). The novelty of the Sampo Model lies in combining several ideas and operationalizing them for developing LOD services and applications in Digital Humanities.

The research area of video indexing is surveyed in (Hu, Xie, Li, Zeng, & Maybank, 2011). Indexing can be done by analyzing the frames and/or audio of the recording to find, e.g., the spots where goals are made in a football match. Another option is to use the textual subtitles (dialogues, commentaries) of the video. In some cases, e.g., in historical film archives, manually curated textual descriptions or commentaries of the videos may be available for preserving cultural heritage—they can be used for annotations and indexing, too, as in our case study. Various methods and tools are available for extracting linked data from texts (Martinez-Rodriguez, Hogan, & Lopez-Arevalo, 2020). Providing contextual information and ads while watching videos

[20] `http://www.europeana-collections-1914-1918.eu`
[21] `http://www.1914-1918-online.net`
[22] `http://ww1.discovery.ac.uk`
[23] `http://www.canadiana.ca/en/pcdhn-lod/`
[24] `http://www.cendari.eu/research/first-world-war-studies/`
[25] `http://blog.muninn-project.org`
[26] `http://www.bygle.net/wp-content/uploads/2015/04/Open-Memory-Project_3-1.pdf`
[27] `https://www.oorlogsbronnen.nl`
[28] `https://lod-cloud.net`
[29] `https://www.researchspace.org`
[30] `http://ckcc.huygens.knaw.nl`

has been suggested already in the 80's in systems such as Hypersoap[31]. Works on enriching video watching experience using linked data-based recommendations include, e.g., (Nixon, Bauer, & Bara, 2013).

5.3. Discussion

We have learned that even in the rural northern parts of Europe, massive amounts of WW2 data can be found and opened for public use. In WarSampo some 100,000 people, mostly casualties of the WW2 events are considered. However, there are also data available about hundreds of thousands of soldiers who survived the war only in Finland. On a European level, these numbers are counted in tens of millions. Managing the data, and providing it for different user groups, suggests serious challenges when dealing with, e.g., the war events in the central parts of Europe, where the amount of data is orders of magnitude larger than in Finland, multilingual, and distributed in different countries. For example, solving entity resolution problems regarding historical place names and person names can be difficult. However, it seems that using Linked Data is a promising way to tackle these challenges.

A major challenge in creating data analyses like the ones shown in this paper is related to the quality of the data produced. Historical (meta)data is typically incomplete and our knowledge about it is uncertain. Also using more or less automatic means for transforming and linking the data leads to problems of incomplete, skewed, and erroneous data (Mäkelä et al., 2020). In general, more data literacy (Koltay, 2015) is usually needed from the end-user when using data analytic tools. The methods of network analysis, for instance, can be very sensitive to even small errors in the data or biases in the sampling schemes.

This as well as conceptual difficulties in modeling complex real world ontologies, such as historical geographical gazetteers, become sometimes embarrassingly visible when using and exposing the knowledge structures to end-users. The same problems exist in traditional systems but are often hidden in the non-structured presentations of the data and UIs.

Arguably three successive generations of semantic portals for cultural heritage can be identified (Hyvönen, 2020): In the begin of the millennium, the research focus in semantic portal development was on data harmonization, aggregation, search, and browsing ("first generation systems"). The rise of Digital Humanities research then started to shift the focus to providing the user with tools for solving data-analytic research problems in interactive ways ("second generation systems"). The portals presented in this paper are examples of this generation. The next step ahead to "third generation systems" is based on Artificial Intelligence: future portals not only provide tools for the human to solve problems but are used for finding research problems in the first place, for addressing them, and even for solving them automatically under the constraints set by the human researcher. Such systems should preferably be able to explain their reasoning, which is an important aspect in the source critical humanities research tradition. In the novel *Hitchhiker's Guide to the Galaxy* by Douglas Adams the supercomputer came to the conclusion that the meaning of life, the universe, and everything is "42". In addition to knowing that it would be nice to know why so.

---

[31] www.media.mit.edu/hypersoap/

Acknowledgements

Lots of people at the Semantic Computing Research Group (SeCo)[32] have contributed to the work presented in this paper, funded by several of organizations in Finland. Thanks to Marcia Zeng for commenting an earlier version of this paper. Thanks also to the Finnish Cultural Foundation for an Eminentia Grant for reflecting our work. CSC – IT Center for Science has provided us with computational resources.

References

Ahola, A., Hyvönen, E., Rantala, H., & Kauppala, A. (2023). Publishing and studying historical opera and music theatre performances on the Semantic Web: case OperaSampo 1830–1960. In *Proceedings of SWODCH 2023. Semantic Web and Ontology Design for Cultural Heritage. Co-located with the 22nd International Semantic Web Conference (ISWC 2023) in Athens, Greece.* CEUR Workshop Proceedings, Vol-3540. Retrieved from `https://ceur-ws.org/Vol-3540/paper8.pdf`

Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked Data – The Story So Far. *International Journal on Semantic Web and Information Systems (IJSWIS)*, *5*(3), 1–22. doi: 10.4018/jswis.2009081901

Collins, T., Mulholland, P., & Zdrahal, Z. (2005). Semantic browsing of digital collections. In Y. Gil, E. Motta, V. Richard Benjamins, & M. A. Musen (Eds.), *The Semantic Web – ISWC 2005* (Vol. 3729, pp. 127–141). Springer. doi: 10.1007/11574620˙12

de Boer, V., van Doornik, J., Buitinck, L., Marx, M., & Veken, T. (2013). Linking the kingdom: Enriched access to a historiographical text. In *Proceedings of the Seventh International Conference on Knowledge Capture (K-CAP 2013)* (pp. 17–24). Association for Computing Machinery, New York, NY, USA. doi: 10.1145/2479832.2479849

Ferreira, A. C., dos Santos Maculan, B. C. M., & Naves, M. M. L. (2017). Ranganathan and the faceted classification theory. *TransInformação, Campinas*, *29*(3), 279–295. doi: 10.1590/2318-08892017000300006

Frosterus, M., Tuominen, J., Pessala, S., & Hyvönen, E. (2015). Linked open ontology cloud: managing a system of interlinked cross-domain light-weight ontologies. *International Journal of Metadata, Semantics and Ontologies*, *10*(3), 189–201. Retrieved from `http://dx.doi.org/10.1504/IJMSO.2015.073879`

Hearst, M. (2006). Design recommendations for hierarchical faceted search interfaces. In *ACM SIGIR workshop on faceted search* (pp. 1–5).

Heath, T., & Bizer, C. (2011). *Linked data: Evolving the web into a global data space* (1st ed.). Morgan & Claypool, Palo Alto, CA, USA.

Heino, E., Tamper, M., Mäkelä, E., Leskinen, P., Ikkala, E., Tuominen, J., ... Hyvönen, E. (2017). Named entity linking in a complex domain: Case Second World War history. In J. Gracia, F. Bond, J. P. McCrae, P. Buitelaar, C. Chiarcos, & S. Hellmann (Eds.), *Language, Data, and Knowledge: First International Conference, LDK 2017* (Vol. 10318). Springer. doi: 10.1007/978-3-319-59888-8˙10

Hitzler, P. (2021, January). A review of the semantic web field. *Commun. ACM*, *64*(2), 76–83. doi: 10.1145/3397512

Hitzler, P., Krötzsch, M., & Rudolph, S. (2010). *Foundations of semantic web technologies.* Springer.

Hu, W., Xie, N., Li, L., Zeng, X., & Maybank, S. (2011). A survey on visual content-based video indexing and retrieval. *IEEE Transactions on Systems, Man, and Cybernetics*, *41*(6), 797-819. doi: 10.1109/TSMCC.2011.2109710

Hyvönen, E. (2010). Preventing interoperability problems instead of solving them. *Semantic Web – Interoperability, Usability, Applicability*, *1*(1–2), 33–37. doi: 10.3233/SW-2010-0014

[32] `https://seco.cs.aalto.fi/`

Hyvönen, E. (2012). *Publishing and using cultural heritage linked data on the semantic web.* Morgan & Claypool, Palo Alto, CA, USA.

Hyvönen, E., Mäkelä, E., Salminen, M., Valo, A., Viljanen, K., Saarela, S., ... Kettula, S. (2005). MuseumFinland—Finnish museums on the semantic web. *Journal of Web Semantics*, *3*(2), 224–241. doi: 10.1016/j.websem.2005.05.008

Hyvönen, E., Saarela, S., & Viljanen, K. (2004). Application of ontology techniques to view-based semantic search and browsing. In *The Semantic Web: Research and Applications. Proceedings of the First European Semantic Web Symposium (ESWS 2004).* Springer. doi: 10.1007/978-3-540-25956-5˙7

Hyvönen, E., Viljanen, K., Tuominen, J., & Seppälä, K. (2008). Building a National Semantic Web Ontology and Ontology Service Infrastructure – The FinnONTO Approach. In *The Semantic Web: Research and Applications. ESWC 2008* (pp. 95–109). Springer. doi: 10.1007/978-3-540-68234-9˙10

Hyvönen, E. (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), 187–193. doi: 10.3233/SW-190386

Hyvönen, E. (2023a). Digital humanities on the semantic web: Sampo model and portal series. *Semantic Web – Interoperability, Usability, Applicability*, *14*(4), 729-744. doi: 10.3233/SW-223034

Hyvönen, E. (2023b). How to create a national cross-domain ontology and linked data infrastructure and use it on the semantic web. *Semantic Web – Interoperability, Usability, Applicability.* Retrieved from `https://seco.cs.aalto.fi/publications/2022/hyvonen-infra-2022.pdf` (Under review)

Hyvönen, E., Heino, E., Leskinen, P., Ikkala, E., Koho, M., Tamper, M., ... Mäkelä, E. (2016). WarSampo Data Service and Semantic Portal for Publishing Linked Open Data about the Second World War History. In H. Sack, E. Blomqvist, M. d'Aquin, C. Ghidini, S. P. Ponzetto, & C. Lange (Eds.), *The Semantic Web. Latest Advances and New Domains: 13th International Conference, ESWC 2016* (pp. 758–773). Springer. doi: 10.1007/978-3-319-34129-3˙46

Hyvönen, E., Ikkala, E., Koho, M., Leal, R., Rantala, H., & Tamper, M. (2022). How to search and contextualize scenes inside videos for enriched watching experience: Case stories of the Second World War veterans. In *The semantic web: ESWC 2022 satellite events* (pp. 163–167). Springer. doi: 10.1007/978-3-031-11609-4˙31

Hyvönen, E., Leskinen, P., Tamper, M., Rantala, H., Ikkala, E., Tuominen, J., & Keravuori, K. (2019). BiographySampo – Publishing and enriching biographies on the Semantic Web for digital humanities research. In *The Semantic Web. 16th International Conference, ESWC 2019* (pp. 574–589). Springer. doi: 10.1007/978-3-030-21348-0˙37

Ikkala, E., Hyvönen, E., Rantala, H., & Koho, M. (2022, January). Sampo-UI: A full stack JavaScript framework for developing semantic portal user interfaces. *Semantic Web – Interoperability, Usability, Applicability*, *13*(1), 69–84. doi: 10.3233/SW-210428

Ikkala, E., Koho, M., Heino, E., Leskinen, P., Hyvönen, E., & Ahoranta, T. (2017). Prosopographical views to Finnish WW2 casualties through cemeteries and linked open data. In A. Adamou, E. Daga, & L. Isaksen (Eds.), *Proceedings of the Second Workshop on Humanities in the Semantic Web (WHiSe II)* (pp. 45–56). CEUR Workshop Proceedings, Vol. 2014. Retrieved from `http://ceur-ws.org/Vol-2014/`

Koho, M., Heino, E., & Hyvönen, E. (2016). SPARQL Faceter – Client-side Faceted Search Based on SPARQL. In *Joint proc. of the 4th international workshop on linked media and the 3rd developers hackshop.* CEUR Workshop Proceedings, Vol. 1615. Retrieved from `http://ceur-ws.org/Vol-1615/semdevPaper5.pdf`

Koho, M., & Hyvönen, E. (2023). Studying occupations and social measures of perished soldiers in WarSampo linked open data. In *Biographical Data in a Digital World 2022 (BD 2022), Tokyo.* Retrieved from `https://seco.cs.aalto.fi/publications/2022/koho-et-al-warsampo-bd-2022.pdf` (Forth-coming)

Koho, M., Hyvönen, E., Heino, E., Tuominen, J., Leskinen, P., & Mäkelä, E. (2017). Linked death — representing, publishing, and using second world war death records as linked open data. In E. Blomqvist, K. Hose, H. Paulheim, A. Ławrynowicz, F. Ciravegna, & O. Hartig (Eds.), *The semantic web: Eswc 2017 satellite events* (Vol. 10577, pp. 369–383). Cham: Springer. doi: 10.1007/978-3-319-70407-4˙45

Koho, M., Ikkala, E., Heino, E., & Hyvönen, E. (2018). Maintaining a linked data cloud and data service for second world war history. In *Digital Heritage. Progress in Cultural Heritage: Documentation, Preservation, and Protection. 7th International Conference, EuroMed 2018* (Vol. 11196). Springer. doi: 10.1007/978-3-030-01762-0˙12

Koho, M., Ikkala, E., & Hyvönen, E. (2018). How to maintain a linked data cloud in a deployed semantic portal. In *Proceedings of the ISWC 2018 Posters & Demonstrations, Industry and Blue Sky Ideas Tracks.* CEUR Workshop Proceedings, Vol. 2180. Retrieved from `http://ceur-ws.org/Vol-2180/`

Koho, M., Ikkala, E., & Hyvönen, E. (2020, September). 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 the Digital Age (BD 2019).* (in press)

Koho, M., Ikkala, E., Leskinen, P., Tamper, M., Tuominen, J., & Hyvönen, E. (2021). WarSampo knowledge graph: Finland in the Second World War as linked open data. *Semantic Web – Interoperability, Usability, Applicability*, *12*(2), 265–278. doi: 10.3233/SW-200392

Koho, M., Leal, R., Ikkala, E., Tamper, M., Rantala, H., & Hyvönen, E. (2022). Building lightweight ontologies for faceted search with named entity recognition: Case WarMemoirSampo. In S. Tiwari, N. Mihindukulasooriya, F. Osborne, D. Kontokostas, J. D'Souza, & M. Kejriwal (Eds.), *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)* (pp. 19–35). CEUR Workshop Proceedings, Vol. 3184. Retrieved from `http://ceur-ws.org/Vol-3184/TEXT2KG_Paper_2.pdf`

Koltay, T. (2015). Data literacy for researchers and data librarians. *Journal of Librarianship and Information Science*, *49*(1), 3–14. doi: 10.1177/0961000615616450

Leal, R., Rantala, H., Koho, M., Ikkala, E., Merenmies, M., & Hyvönen, E. (2022, March). WarMemoirSampo: A semantic portal for war veteran interview videos. In *DHNB 2022 The 6th Digital Humanities in Nordic and Baltic Countries Conference.* CEUR Workshop Proceedings, long papers, Vol. 3232. Retrieved from `http://ceur-ws.org/Vol-3232/paper30.pdf`

Leskinen, P., Rantala, H., & Hyvönen, E. (2022). Analyzing the lives of Finnish academic people 1640—1899 in Nordic and Baltic countries: AcademySampo data service and portal. In *DHNB 2022 The 6th Digital Humanities in Nordic and Baltic Countries Conference.* CEUR Workshop Proceedings, long papers, Vol. 3232. Retrieved from `http://ceur-ws.org/Vol-3232/paper07.pdf`

Linstone, H. A. (1989). Multiple perspectives: Concept, applications, and user guidelines. *Systems practice*, *2*(3), 307–331. doi: 10.1007/BF01059977

Mäkelä, E., Lagus, K., Lahti, L., Säily, T., Tolonen, M., Hämäläinen, M., ... Nevalainen, T. (2020). Wrangling with non-standard data. In *Proceedings of the Digital Humanities in the Nordic Countries 5th Conferenc* (pp. 81–96). CEUR Workshop Proceedings, Vol. 2612. Retrieved from `http://ceur-ws.org/Vol-2612/paper6.pdf`

Marchionini, G. (2006). Exploratory search: from finding to understanding. *Communications of the ACM*, *49*(4), 41–46. doi: 10.1145/1121949.1121979

Martinez-Rodriguez, J. L., Hogan, A., & Lopez-Arevalo, I. (2020). Information extraction meets the semantic web: A survey. *Semantic Web – Interoperability, Usability, Applicability*, *11*(2), 255–335. doi: 10.3233/SW-180333

Moretti, F. (2013). *Distant reading.* Verso Books. doi: 10.1093/llc/fqu010

Mäkelä, E., Hypén, K., & Hyvönen, E. (2013). *Fiction literature as linked open data – the BookSampo dataset* (Vol. 4) (No. 3). doi: 10.3233/SW-120093

Mäkelä, E., & Hyvönen, E. (2014). SPARQL SAHA, a configurable linked data editor and browser as a service. In V. Presutti, E. Blomqvist, R. Troncy, H. Sack, I. Papadakis, & A. Tordai (Eds.), *The Semantic Web: ESWC 2014 Satellite Events. ESWC 2014* (pp. 434–438). Springer. doi: 10.1007/978-3-319-11955-7˙62

Mäkelä, E., Törnroos, J., Lindquist, T., & Hyvönen, E. (2017). WW1LOD: An application of CIDOC-CRM to World War 1 Linked Data. *International Journal on Digital Libraries*, *18*(4), 333–343.

Nixon, L., Bauer, M., & Bara, C. (2013). Connected media experiences: Web based interactive video using linked data. In *Proceedings of the 22nd international conference on world wide web* (p. 309–312). ACM, New York. doi: 10.1145/2487788.2487931

Pollitt, A. S. (1998). *The key role of classification and indexing in view-based searching* (Tech. Rep.). UK: University of Huddersfield. Retrieved from `http://www.ifla.org/IV/ifla63/63polst.pdf`

Rantala, H., Ahola, A., Ikkala, E., & Hyvönen, E. (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 VOILA 2023, Co-locatd with ISWC 2023, Athens.* CEUR Workshop Proceedings, Vol. 3508. Retrieved from `https://ceur-ws.org/Vol-3508/paper3.pdf`

Rantala, H., Ikkala, E., Jokipii, I., Koho, M., Tuominen, J., & Hyvönen, E. (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* (pp. 191–196). Springer. doi: 10.1007/978-3-030-62327-2˙33

Rantala, H., Jokipii, I., Ikkala, E., & Hyvönen, E. (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*(1). doi: 0.1145/3477606

Ravenek, W., van den Heuvel, C., & Gerritsen, G. (2017). The ePistolarium: Origins and techniques. In A. van Hessen & J. Odijk (Eds.), *CLARIN in the Low Countries* (pp. 317–323). Ubiquity Press. doi: 10.5334/bbi

Rietveld, L., & Hoekstra, R. (2017). The YASGUI family of SPARQL clients. *Semantic Web – Interoperability, Usability, Applicability*, *8*(3), 373–383. doi: 10.3233/SW-150197

Sacco, G. M. (2005). Dynamic taxonomies: guided interactive diagnostic assistance. In N. Wickramasinghe (Ed.), *Encyclopedia of healthcare information systems.* Idea Group.

Suominen, O., Johansson, A., Ylikotila, H., Tuominen, J., & Hyvönen, E. (2012). Vocabulary services based on SPARQL endpoints: ONKI Light on SPARQL. In *Poster proceedings of the 18th International Conference on Knowledge Engineering and Knowledge Management (EKAW 2012).* Retrieved from `https://seco.cs.aalto.fi/publications/2012/suominen-et-al-onkilight-2012.pdf`

Suominen, O., Pessala, S., Tuominen, J., Lappalainen, M., Nykyri, S., Ylikotila, H., ... Hyvönen, E. (2014). Deploying national ontology services: From ONKI to Finto. In *Proceedings of the Industry Track at the International Semantic Web Conference 2014.* CEUR Workshop Proceedings, Vol. 1383. Retrieved from `http://www.ceur-ws.org/Vol-1383`

Tunkelang, D. (2009). *Faceted search.* Morgan & Claypool, Palo Alto, CA, USA. doi: 10.1007/978-3-031-02262-3

Tuominen, J., Frosterus, M., Viljanen, K., & Hyvönen, E. (2009). ONKI SKOS server for publishing and utilizing SKOS vocabularies and ontologies as services. In *Proceedings of the 6th european semantic web conference (eswc 2009)* (pp. 768–780). Springer. doi: 10.1007/978-3-642-02121-3˙56

Tzitzikas, Y., Manolis, N., & Papadakos, P. (2017). Faceted exploration of RDF/S datasets: a survey. *Journal of Intelligent Information Systems*, *48*(2), 329–364. doi: 10.1007/s10844-016-0413-8

Verboven, K., Carlier, M., & Dumolyn, J. (2007). A short manual to the art of prosopography. In *Prosopography approaches and applications. a handbook* (pp. 35–70). Unit for Prosopographical Research (Linacre College). doi: 1854/8212

Viljanen, K., Tuominen, J., & Hyvönen, E. (2009). Ontology libraries for production use: The Finnish ontology library service ONKI. In *Proceedings of the 6th european semantic web* *conference (eswc 2009)* (pp. 781–795). Springer. doi: 10.1007/978-3-642-02121-3˙57

Wang, Q. (2023). *Forging a path to understanding: Integrating and analyzing war history data through knowledge graph technologies.* PhD Thesis manuscript submitted, The Australian National University, Canberra.

Zeng, M., & Qin, J. (2022). *Metadata, third edition.* ALA Neal-Schuman, Chicago.