

Invited article to be published in EuropeanaTech Insight, 2017:  
<http://pro.europeana.eu/page/europeanatech-insight>



## WarSampo: Publishing and Using Linked Open Data about the Second World War

Eero Hyvönen, Erkki Heino, Petri Leskinen, Esko Ikkala,  
Mikko Koho, Minna Tamper, Jouni Tuominen and Eetu Mäkelä

Aalto University and University of Helsinki  
Helsinki Centre for Digital Humanities (HELDIG) <http://heldig.fi>  
Semantic Computing Research Group (SeCo) <http://seco.cs.aalto.fi>

The article overviews the system *WarSampo – Finnish World War 2 on the Semantic Web*, the winner of the LODLAM Challenge 2017 Open Data Prize on June 29 in Venice, Italy.

### Learning from History is Hard

Many websites publish information about the Second World War (WW2), the largest global tragedy in human history. Such information is of great interest not only to historians but to hundreds of millions of citizens globally who themselves or whose relatives participated in the war actions, creating a shared trauma all over the world. However, WW2 information on the web is typically meant for human consumption only, and there are hardly any websites that serve machine-readable data about the WW2 for Digital Humanities (DH) and end-user applications to use. Our work on the WarSampo system is based on the belief that by making war data more accessible our understanding of the reality of the war improves, which not only advances understanding of the past but also promotes peace in the future. Georg Wilhelm Friedrich Hegel has said: “*we learn from history that we learn nothing from history*”. Hopefully this is not the case for the WW2 now that fighting has started again even within the borders of Europe in Ukraine.

The WarSampo system 1) initiates and fosters large scale Linked Open Data (LOD) publication of WW2 data from distributed, heterogeneous data silos and 2) demonstrates and suggests its use in applications and DH research. WarSampo is to our best knowledge the first large scale system for serving and publishing WW2 LOD on the Semantic Web for machine and human users. Its knowledge graph metadata contains over 9 million associations (triples) between data items including, e.g., a complete set of over 95,000 death records of Finnish WW2 soldiers, 160,000 authentic photos taken during the war, 32,000 historical places on historical maps, 23,000 war diaries of army units, and 3,400 memoir articles written by the veterans after the war. WarSampo data comes from several Finnish organizations and sources, such as National Archives, Defense Forces, Land Survey of Finland, Wikipedia/DBpedia, text books, and magazines.

WarSampo has two separate components: 1) WarSampo Data Service for machines and 2) WarSampo Semantic Portal with various applications for human users.

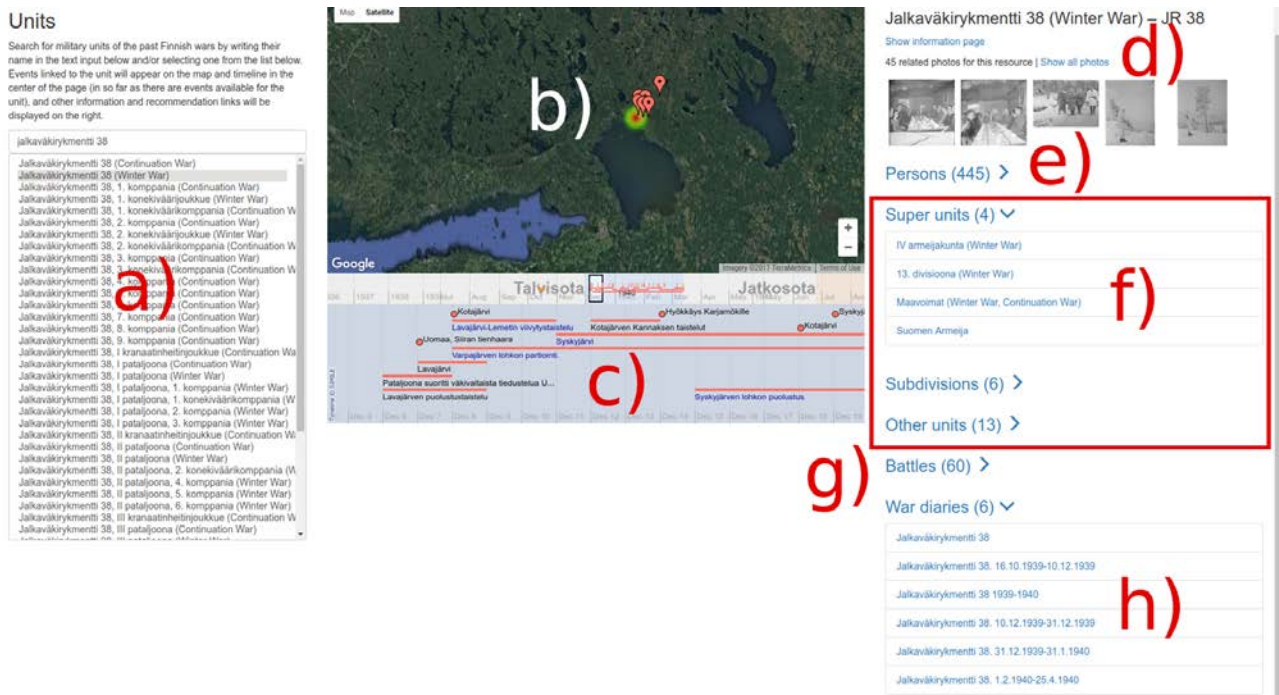
## WarSampo Data Service

WarSampo publishes massive collections of heterogeneous, distributed (meta)data about WW2 on the Semantic Web. The datasets are harmonized into LOD using event-based modeling, based on an extension of CIDOC CRM, and are enriched semantically with each other's contents and external data sources. The data is openly available for download, linked data browsing, and querying with SPARQL via the Linked Data Finland "7-star" platform (<http://www.ldf.fi/dataset/warsa>) for Digital Humanities research and for creating applications.

## WarSampo Semantic Portal

To test and demonstrate the usefulness of the data service, WarSampo Semantic Portal (<http://www.sotasampo.fi/en/>) was created, allowing both historians and laymen to study war history and the destinies of their family members in the war from 7 interlinked application perspectives: Historical events, Persons, Army Units, Places, Memoir Articles, Death Records, and Photographs. Two new application perspectives pertaining to some 650 war cemeteries and over 3,000 photographs of them, and to a database of some 4,500 prisoners of war, will be published by the end of 2017, the centennial of independence of Finland. The portal immediately attracted thousands of visitors after its publication in November 2015.

A typical use case of the system is automatic reconstruction of a soldier's life story based on data linking. For example, if one's relative was killed in action, he can be found in the death record data 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 in the unit be illustrated on maps and timelines (assuming that he was travelling with the unit). 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 and using Named Entity Linking. Furthermore, links to the actual war diaries can be provided based on the army unit data. 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 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 war diaries (h).



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

## Lessons Learned

Work on WarSampo argues and demonstrates that world war history makes a promising use case for Linked Open Data because war data is by nature heterogeneous, distributed in different countries and organizations, and written in different languages. WarSampo is based on the idea of creating a shared, open semantic data repository with a sustainable “business model” where everybody wins: When an organization contributes to the WW2 LOD cloud with a piece of information, say a photograph, its description is automatically connected to related data, such as persons or places depicted. At the same time, the related pieces of information, provided by others, are enriched with links to the new data.

## More Information: Videos and Scientific Publications

A short overview video about WarSampo has been published here:

<https://vimeo.com/212249404>

The project homepage (<http://seco.cs.aalto.fi/projects/sotasampo/en/>) describes the system in more detail and contains additional videos on using the application. Scientific publications as well as academic theses based on WarSampo research are available at the bottom of the project homepage, such as:

Eero Hyvönen, Erkki Heino, Petri Leskinen, Esko Ikkala, Mikko Koho, Minna Tamper, Jouni Tuominen and Eetu Mäkelä: 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, May, 2016.

<http://seco.cs.aalto.fi/publications/2016/hyvonen-et-al-warsa.pdf>