Publishing Second World War History as Linked Data Events on the Semantic Web

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

Semantic Computing Research Group (SeCo), Aalto University, Finland
`http://www.seco.tkk.fi/, firstname.lastname@aalto.fi`

Abstract. Data about wars is typically heterogeneous, distributed in the data silos of the fighting parties, multilingual, and often controversial depending on the political point of view. It is therefore hard for the historians to get a global picture of what has actually happened, to whom, where, and how. This paper argues that Semantic Web and Linked Data technologies are a very promising approach for modeling, harmonizing, and aggregating data about war history. This makes it possible, for both historians and laymen, to study history in a contextualized way where linked datasets enrich each other. This paper introduces the WarSampo project, where massive collections of heterogeneous data about the history of the Second World War are harmonized using an event-based approach, and provided as a Linked Open Data service for applications to use. As a use case, a semantic portal WarSampo providing different perspectives to the war based on events is discussed.

1 Second World War on the Semantic web

There are several projects publishing data about the World War I on the web, such as Europeana Collections 1914–1918[1], 1914–1918 Online[2], WW1 Discovery[3], Out of the Trenches[4], CENDARI[5], Muninn[6], and WW1LOD [7]. War history makes a promising use case for Linked Data because war data is heterogeneous, distributed in different countries and organizations, and written in different languages [4]. Using metadata, also different opinions and conflicting information about the war can be represented.

Many websites also publish information about the World War II, the largest global tragedy in human history, such as the World War II Database[7] to name just one. However, such portal data is typically meant for human consumption, and there are only few works that deal with *machine readable data* about the WW2 for applications to use, such as [2,1].

---

[1] http://www.europeana-collections-1914-1918.eu

[2] http://www.1914-1918-online.net

[3] http://ww1.discovery.ac.uk

[4] http://www.canadiana.ca/en/pcdhn-lod/

[5] http://www.cendari.eu/research/first-world-war-studies/

[6] http://blog.muninn-project.org

[7] http://ww2db.com

This paper introduces the WarSampo Linked Data service and a semantic portal on top of it. Our work contributes to the related research above by initiating and fostering large scale LOD publication of WW2 data on the web, based on event-based data modeling. The idea is to publish Linked Open data, aggregated from distributed data silos, for Digital Humanities applications to use. In our case study, the data is related to the Finnish Winter War 1939–1940 against the Soviet attack, the Continuation War 1941–1944, where the occupied areas of the Winter War were temporarily regained by Finns, and the Lapland War 1944–1945, where the Finns pushed the German troops away from Lapland.

We first present and discuss the data modeling approach developed for the WarSampo LOD service. After this an application perspective in the WarSampo portal is presented where events are linked to related resources, such as photos, persons, and historical places. In conclusion, some lessons learned thus far are discussed and directions for further research pointed out.

2 The Data Service: Modeling War Events as Linked Data

Dataset	Name	Providing organization	Size
1	Casualties of WW2	National Archives	93,000 death records
2	War diaries	National Archives	23,000 war diaries of troops
3	Photos & films	Defence Forces & Military Museum	160,000 photos & films
4	Kansa taisteli magazine articles	Bonnier & The Assoc. for Military History in Finland	3,360 articles of veteran soldiers
5	Karelian places	National Land Survey	35,000 places of the annexed Karelia
6	Karelian maps	National Land Survey	War time maps of Karelia
7	Organization cards	National Archives	War events of ca 500 major units
8	Audio & films	National Broadcasting Company YLE	250 recordings and films

Table 1. Central datasets to be linked in WarSampo.

Data The project deals initially with the datasets presented in Table 1. The casualties data (1) includes data about the deaths in action during the wars. War diaries (2) are digitized authentic documentations of the troop actions in the frontiers. Photos and films (3) were taken during the war by the troops of the Defense Forces. The Kansa taisteli magazine (4) was published in 1957–1986; its articles contain mostly memories of the men that fought on the fronts. Karelian places (5) and maps (6) cover the war zone area

in pre-war Finland that was finally annexed by the Soviet Union. Organization cards (8), written after the war, document events of military units during the war. YLE's audio and film material about the wars (7) ("Living Archive") was recorded during the war, or is related to it.

Metadata Models CIDOC CRM[8] is used as the harmonizing basis for modeling data, with events providing the semantic glue for data linking [3]. Our data model for WW1, presented in [7], is used as the metadata model to start with.

Domain Ontologies The data is annotated using a set of domain ontologies, including: 1) an ontology of the troops and their hierarchies, 2) persons with their ranks and roles, 3) place ontology of historical places, 4) event ontology of battles, politics, and other war time incidents, 5) an ontology of time periods, 6) ontology of weapons, 7) ontology of vessels, and 8) a subject matter ontology.

WarSampo data is published as a SPARQL endpoint service using the "7-star" Linked Data Finland platform[9] [6], based on Fuseki[10] with a Varnish Cache[11] front-end for serving LOD. The casualty data from the National Archives is already publicly available for everyone to use[12], other datasets are at the moment used internally in application development, and will be made open to the public later.

3 Event-based Perspective to War History

The idea of the WarSampo portal is to provide a variety of different perspectives [5] to war history. For example, most datasets in Table 1 have their own perspectives, where the user can first search data of interest and then get linked data related to the other resources found. The idea is that the perspectives enrich each other via Linked Data.

Fig. 1 illustrates the WarSampo event perspective application to the WarSampo data. Events are displayed on a map, (a) in Fig. 1, and on a timeline (b) that shows here events of the Winter War. When the user clicks on an event, it is highlighted (c), and the historical place, time span, type, and description for the selected event are displayed (d). Photographs related to the event (e) are also shown. The photographs are linked to events based on location and time. Furthermore, information about casualties during the time span visible on the timeline is shown alongside the event description (f), and the map (a) features a heatmap layer for a visualization of these deaths.

The events can also be found and visualized through other perspectives. For example, in the military units perspective, the events in which a unit participated can be viewed on maps and in time, providing a kind of graphical activity summary of the unit. In the casualties perspective, military units of the dead soldiers are known, making it possible sort out and visualize the personal war history of the casualties e.g. on historical maps that come from a yet another dataset in WarSampo.

---

[8] http://cidoc-crm.org

[9] http://www.ldf.fi

[10] http://jena.apache.org/documentation/serving_data/

[11] https://www.varnish-cache.org

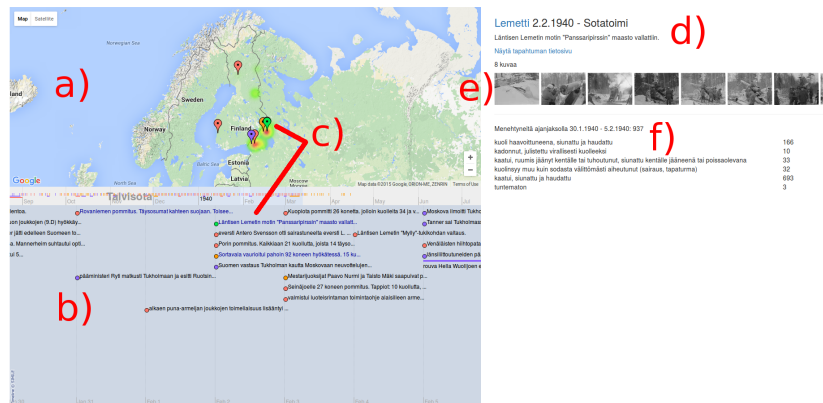[12] http://www.ldf.fi/dataset/narc-menehtyneet1939-45

Fig. 1. Event perspective featuring a timeline and map.

4 Discussion

Our first experiments, as illustrated in Section 3, suggest that heterogeneous datasets of war history really can be interlinked with each other through events in ways that provide useful insights for the historians. We have also learned that even in the rural northern parts of Europe, massive amounts of WW2 data can be found. We have initially dealt with tens of thousands of people killed in action. However, there is also data available about hundreds of thousands of soldiers who survived the war. In addition to historians, WarSampo data is very interesting to the layman, too: every soldier's history is of interest at least to e.g. his/her relatives. Managing the data, and providing it for different user groups, suggests serious challenges when dealing with e.g. the war 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 hard. However, it seems that Linked Data is a promising way to tackle these challenges.

References

1. de Boer, V., van Doornik, J., Buitinck, L., Marx, M., Veken, T.: Linking the kingdom: enriched access to a historiographical text. In: Proc. of the 7th International Conference on Knowledge Capture (KCAP 2013). pp. 17–24. Association of Computing Machinery, New York (2013)
2. Collins, T., Mulholland, P., Zdrahal, Z.: Semantic browsing of digital collections. In: Proc. of the 4th International Semantic Web Conference (ISWC 2005). Springer–Verlag (2005)
3. Doerr, M.: The CIDOC CRM – an ontological approach to semantic interoperability of metadata. AI Magazine 24(3), 75–92 (2003)
4. Hyvönen, E.: Publishing and using cultural heritage linked data on the semantic web. Morgan & Claypool, Palo Alto, CA, USA (2012)

5. Hyvönen, E., Lindquist, T., Törnroos, J., Mäkelä, E.: History on the semantic web as linked data – an event gazetteer and timeline for World War I. In: Proc. of CIDOC 2012 – Enriching Cultural Heritage (2012)
6. Hyvönen, E., Tuominen, J., Alonen, M., Mäkelä, E.: Linked Data Finland: A 7-star model and platform for publishing and re-using linked datasets. In: The Semantic Web: ESWC 2014 Satellite Events, Revised Selected Papers. pp. 226–230. Springer–Verlag (2014)
7. Mäkelä, E., Törnroos, J., Lindquist, T., Hyvönen, E.: World War I as Linked Open Data (2015), http://www.seco.tkk.fi/publications/submitted/makela-et-al-ww1lod.pdf, submitted for review