

Linked Death – representing, publishing, and using Second World War death records as Linked Open Data

Mikko Koho, Eero Hyvönen, Erkki Heino, Jouni Tuominen, Petri Leskinen, and Eetu Mäkelä

Semantic Computing Research Group (SeCo),
Aalto University, School of Science, Finland
<http://seco.cs.aalto.fi/>, first.last@aalto.fi

Abstract. War history of the Second World War (WW2), humankind’s largest disaster, is of great interest to both laymen and researchers. Most of us have ancestors and relatives who participated in the war, and in the worst case got killed. Researchers are eager to find out what actually happened then, and even more importantly why, so that future wars could perhaps be prevented. The darkest data of war history are casualty records—from such data we could perhaps learn most about the war. This paper presents a model and system for representing death records as linked data, so that 1) citizens could find out more easily what happened to their relatives during WW2 and 2) digital humanities (DH) researchers could (re)use the data easily for research.

1 Introduction

Lots of information about the WW2 is available on the Web.¹ However, this information is typically meant for human consumption only. The underlying *data* is not available in machine-readable, i.e., “semantic” form for Digital Humanities research and use [5, 3] and for end-user applications to utilize. 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 hopefully promotes peace in the future [7].

For the case of the First World War, the situation has started to change, with several projects publishing linked data on the web, such as Europeana Collections 1914–1918,² 1914–1918 Online,³ WW1 Discovery,⁴ Out of the Trenches,⁵ CENDARI,⁶ Muninn,⁷ and WW1LOD [11]. A few works have used the linked

¹ <http://ww2db.com>, <http://www.world-war-2.info>, different Wikipedias, etc.

² <http://www.europeana-collections-1914-1918.eu>

³ <http://www.1914-1918-online.net>

⁴ <http://ww1.discovery.ac.uk>

⁵ <http://www.canadiana.ca/en/pcdhn-lod/>

⁶ <http://www.cendari.eu/research/first-world-war-studies/>

⁷ <http://blog.muninn-project.org>

data approach to WW2 data, too, such as [2, 1], the Open Memory Project,⁸ and WarSampo [7].

This paper discusses the publication and use of casualty (death) records as linked data, as one part of the larger WarSampo system. Here, a dataset of some 95,000 deaths in military action in the Finnish frontiers is concerned. We first present the data, its modeling, and the Linked Open Data (LOD) service. After this two use case applications are presented: 1) Analyzing the data for DH research and 2) reassembling the biographical war history of individual soldiers and military units. The latter use case serves, e.g., laymen in trying to figure out what happened to their relatives in WW2. The WarSampo system⁹ was published on Nov 27, 2015 and has had tens of thousands of end users indicating a large public interest in such applications.

2 Dataset, Data Model, and Data Service

Information about all known Finnish casualties of WW2 has been gathered in a relational database at the National Archives. This database contains 94,696 records of persons that fought on the Finnish side, and died in 1939–1945 in the Winter War, the Continuation War, or in the Lapland War, or died of injuries obtained in those wars.

For use in the WarSampo project, the casualty database was first converted to CSV format, which was then converted to RDF format. Because the objective was to develop interactive applications directly on top of the large RDF dataset, it was important to keep the amount of RDF triples as low as possible without losing information and still linking the death records to ontological concepts. Thus, a simple data model was created for representing the data as linked data.

The data model is based on representing each death record as an instance of the Document class (*E31_Document*) of the CIDOC Conceptual Reference Model (CRM) vocabulary, which is designed for information exchange and integration of various cultural heritage information [4]. The graveyard vocabulary describes graveyards around Finland and is also linked to the municipality vocabulary listing Finnish municipalities during WW2.

A metadata schema was created that defines the properties used to describe each casualty with the information from the original database. The schema consists of OWL properties which have *crm:E31_Document* as the domain. A list of the properties and their *rdfs:range* constraints are shown in Table 1. The namespace prefixes used in this paper are:

: <http://ldf.fi/schema/narc-menehtyneet1939-45/>
crm: <http://www.cidoc-crm.org/cidoc-crm/>
skos: <http://www.w3.org/2004/02/skos/core#>

⁸ <http://www.bygle.net/wp-content/uploads/2015/04/Open-Memory-Project.3-1.pdf>

⁹ Including a semantic portal in use at <http://sotasampo.fi> and the underlying LOD SPARQL service at <http://www.ldf.fi/dataset/warsa/>.

wat: http://ldf.fi/warsa/actors/actor_types/
wrank: <http://ldf.fi/warsa/actors/ranks/>

The default namespace corresponds to the casualty schema namespace. RDF Schema (RDFS), Web Ontology Language (OWL) and XML Schema namespaces are omitted. The properties are used only when there is a value for the property. Municipalities are currently linked to two distinct ontologies, which is why their range is not defined.

Table 1. Casualty metadata schema of all properties used for describing the death records.

| Property description | Property name | Range |
|----------------------------------|----------------------|-------------------|
| mother tongue | :aeidinkieli | :Aeidinkieli |
| occupation | :ammatti | xsd:string |
| principal abode | :asuinkunta | |
| first names | :etunimet | xsd:string |
| date of becoming wounded | :haavoittumisaika | xsd:date |
| municipality of becoming wounded | :haavoittumiskunta | |
| place of becoming wounded | :haavoittumisaikka | xsd:string |
| burial place | :hautapaikka | xsd:string |
| burial graveyard | :hautausmaa | :Hautausmaa |
| military unit | :joukko_osasto | xsd:string |
| military unit code | :joukko_osastokoodi | xsd:string |
| known military unit | :osasto | wat:MilitaryUnit |
| citizenship | :kansalaisuus | :Kansalaisuus |
| nationality at time of death | :kansallisuus | :Kansallisuus |
| date of becoming missing | :katoamisaika | xsd:date |
| municipality of becoming missing | :katoamiskunta | |
| place of becoming missing | :katoamisaikka | xsd:string |
| place of domicile | :kotikunta | |
| date of death | :kuolinaika | xsd:date |
| municipality of death | :kuolinkunta | |
| place of death | :kuolinpaikka | xsd:string |
| number of children | :lasten_lukumaeaerae | xsd:integer |
| perishing class | :menehtymisluokka | :Menehtymisluokka |
| marital status | :siviilisaaety | :Siviilisaaety |
| military rank | :sotilasarvo | wrank:Rank |
| last name | :sukunimi | xsd:string |
| gender | :sukupuoli | :Sukupuoli |
| municipality of birth | :synnyinkunta | |
| date of birth | :syntymaiaika | xsd:date |
| full name | skos:prefLabel | rdfs:Literal |
| WarSampo person instance | crm:P70_documents | crm:E21_Person |

The Simple Knowledge Organization System (SKOS)¹⁰ was used to define vocabularies to present the information found in the original database in RDF. The created SKOS vocabularies for describing the death records in the casualty dataset are listed in table 2.

Table 2. SKOS vocabularies for describing the death records.

| Vocabulary | Number of concepts |
|-------------------|--------------------|
| citizenships | 10 |
| genders | 3 |
| graveyards | 802 |
| marital statuses | 5 |
| mother tongues | 11 |
| municipalities | 632 |
| nationalities | 11 |
| perishing classes | 7 |

2.1 Interlinking with WarSampo datasets

The RDF dataset has been enriched by linking it to other parts of WarSampo like military ranks, military units, information about persons found from other sources, and municipalities of wartime Finland. The dataset is published in the Linked Data Finland (LDF) [6] platform, where it is freely available¹¹ for use via an open SPARQL endpoint.

A figure displaying the external linking of the death records is shown in Fig. 1. Each casualty is linked to four common WarSampo ontologies in addition to the vocabularies that are part of the casualty schema. Military ranks are linked to a common WarSampo military rank ontology according to the ranks of persons at the time of death. Persons that have been found and recognized from other sources are linked from the death records to their more general person instances via *crm:P70_documents* from their death records. New information about persons are not added directly to the casualty dataset, in order to maintain the integrity of the dataset as a whole.

The death records and military units are programmatically linked with persons and military units found from other WarSampo sources. The linking is implemented using the automatic annotation service ARPA [10], and a fuzzy logic algorithm to calculate a score for the similarity of two persons based on each person's name, birth date, death date and military rank. If the score is high enough, the persons are expected to be the same and are linked.

For 118 death records, we were able to automatically link them to WarSampo persons from other sources. The amount is quite low because the person infor-

¹⁰ <https://www.w3.org/2009/08/skos-reference/skos.html>

¹¹ <http://www.ldf.fi/dataset/narc-menehtyneet1939-45>

mation from other sources currently contains mostly information about high ranking officers and persons who survived the war. The found person links have been manually validated and found links seem to be depicting the same persons. Manual validation was also done to person pairs that were close to the score threshold but not linked, and these seem to be either not depicting the same persons or don't have enough information to make an assumption either way.

Automatic linking to military units was implemented using the ARPA service. Some 66,700 death records we linked to WarSampo military units, so this counts for 70% of all the casualties.

Municipalities in the data are linked to the historical wartime municipalities registry [8] of WarSampo automatically based on the municipalities' labels. As shown in Table 1, there are six properties that relate to municipalities for each death record. This leads to 98% of all death records having at least one link to the known wartime municipalities. All municipalities do not link as the wartime municipalities dataset is currently incomplete. The burial municipality property links to municipalities in the Finnish Geographic Names Registry,¹² as it refers to contemporary municipalities.

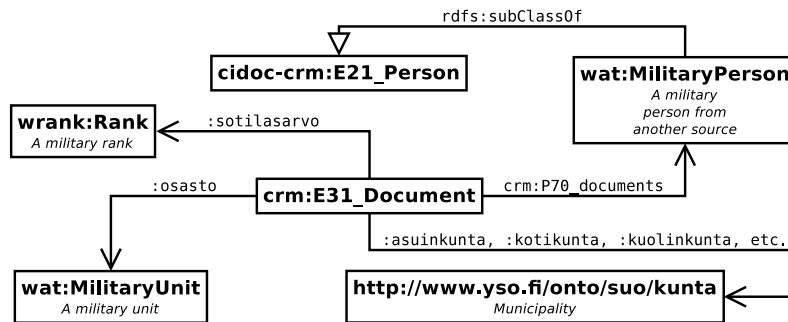


Fig. 1. External links from the death records, which are instances of *crm:E31_Document*. The properties are explained in Table 1. All classes are aligned with the CIDOC CRM framework used in WarSampo by *rdfs:subClassOf* relations.

3 Use Case 1: Studying Death Records

The dataset graph consists of almost 2.4 million RDF triples. Presenting the data in an online service for users to search and browse is not straightforward due to the large size of the dataset. Furthermore, there are lots of links to related data in other WarSampo datasets (persons, places, military units, etc.).

¹² Described at <http://www.ldf.fi/dataset/pnr>

Faceted search provides effective support for interactive information-seeking in information systems [12], and this approach was chosen to create an application for searching and browsing the dataset. The created application¹³ is part of the WarSampo portal, and provides the casualty perspective as one of the portal’s different perspectives.

Fig. 2 shows a screenshot of the faceted search application. The data is laid out in a table-like view. Facets are presented on the left of the interface with string search support. The number of hits on each facet is calculated dynamically and shown to the user, and selections leading to an empty result set are hidden. In addition, there is a text search facet for finding persons directly by name: the user just enters a person’s name or a part of it into the search box, and also a date range selector to filter the results by date of death.

Fig. 2 shows seven facets and the results where the user has selected “widow” in the marital status facet, focusing the search down to 278 killed widows of war that are presented in the table with links to further information, represented by blue text.

The faceted search is used not only for searching but also as a flexible tool for researching the underlying data. In Fig. 2, the hit counts immediately show distributions of the killed widows along the facet categories. For example, the facet “Number of children” shows that one of the deceased had 10 children and most often (in 88 cases) widows had one child. If we next select the category “one child” on its facet, we can see that two of the deceased are women and 86 are men in the gender facet.

| Name | Rank | Unit | Time of Death | Punishing Class | Death Municipality | Place of Death | Gender | Marital Status | Occupation | Mother Tongue | Nationality | Number of Children |
|------------------------|----------------|---|---------------|--|--------------------|--------------------|--------|----------------|------------------|---------------|-------------|--------------------|
| AARNE, ERKKI AARNE | Sotamies | 3.OP-7 | 26.10.1941 | kaatui, suruttu ja haudattu | | Munkkiniemi | Mies | Leski | matkustajäläinen | Suomi | Suomi | 0 |
| ALECHT, FANN | Sotilaskokkoon | Sotilaskokkoon | 05.07.1942 | kaatui, suruttu ja haudattu | Ruukola | Mutanvaara | Nainen | Leski | työväkijä | Suomi | Suomi | 1 |
| AHLQVIST, YLVO EDVARD | Sotamies | Jalkaväkirykmentti 55 | 03.08.1941 | kaatui, suruttu ja haudattu | Ruukola | | Mies | Leski | maastopäivä | Suomi | Suomi | 3 |
| ALONEN, OSKARI | Sotamies | Vaasa Regimentti, Jalkaväkirykmentti 22, Jalkaväkirykmentti 5, Regimentti Hagberg, Vaasa Regimentti, II Pääpatolija, Jalkaväkirykmentti 9, III Pääpatolija, Regimentti Hagberg, III Pääpatolija, Jalkaväkirykmentti 22, III Pääpatolija | 06.11.1941 | kaatui, suruttu ja haudattu | | Syväni voutalaitos | Mies | Leski | puuseppä | Suomi | Suomi | 1 |
| ALÉN, KAUNO KALEVVO | Korpraali | Jalkaväkirykmentti A, Jalkaväkirykmentti 12, III Pääpatolija, 7. Komppania | 13.11.1941 | kuoli haavoittumiseen, suruttu ja haudattu | | 32 Sotais | Mies | Leski | työmieh | Suomi | Suomi | 0 |
| AARPE, EERO OLAVI | Sotamies | 71 Antenninviivienestekomppania | 30.07.1941 | kuoli myrky myrky sodasta väkivallasta aiheutunut (aerava, laupattua) | | 36 KS | Mies | Leski | esimies | Suomi | Suomi | 0 |
| AITTO, MATIAS AUKUSTI | Sotamies | Jalkaväkirykmentti 23 | 24.12.1939 | kaatui, suruttu ja haudattu | Munkkiniemi | Talpoisjärvi | Mies | Leski | maamyllynä | Suomi | Suomi | 1 |
| AITTO, VÄNO | Sotamies | Jalkaväkirykmentti 21, Jalkaväkirykmentti 4 | 22.12.1939 | kaatui, ruuvin jäljiltä, haudattiin kaivokseen, suruttu, keuhkoihin jäämässä oli pöyhäntömyrkkyä | Munkki | | Mies | Leski | maamyllynä pöykä | Suomi | Suomi | 1 |
| BERGSTRÖM, KARL FJALAR | Kersantti | Raid-Pöli | 02.12.1941 | kuoli haavoittumiseen, suruttu ja haudattu | | 36 KS | Mies | Leski | välikomendantti | Ruotsi | Suomi | 0 |
| ERIKSSON, ERIK VILHELM | Sotamies | AufKV 1 | 20.05.1942 | kuoli myrky myrky sodasta | | 16 KS | Mies | Leski | autonostaja | Ruotsi | Ruotsi | 1 |

Fig. 2. The faceted search interface of death records with one selected facet.

¹³ <http://www.sotasampo.fi/casualties/>

The application is developed in JavaScript as a Rich Internet Application (RIA) on the client side, using the open SPARQL endpoint to fetch data according to user selections. The application is open source,¹⁴ and is based on our *SPARQL Faceter* tool [9]. When the user's selections of the facets change, an asynchronous SPARQL query is sent from the user's web browser to the SPARQL endpoint. The SPARQL endpoint returns results of the query to the user's browser, which does additional processing of the data before displaying the new results to the user. The system works well even with the large casualty dataset, because pagination is used to limit the amount of results that are queried and displayed at a time.

The shown facets are configured in the application and they directly use the properties of each death record instance. The property values are annotation resources selected from the corresponding ontologies, such as places.

The event perspective of the WarSampo portal also makes use of the casualty data as seen in Fig. 3. The application displays wartime events on a timeline and map. The casualty data is visualized by a heat map layer on the map, showing an overview of where casualties occurred during different time frames, and also which events happened nearby. The application also displays statistics regarding the casualties during the selected time frame: the total amount of casualties, and the amount per perishing class. Persons mentioned in the event descriptions are linked to the WarSampo person ontology, and thus to the casualties where the person has died in the wars. The application provides hyperlinks to the linked entities through their perspectives.

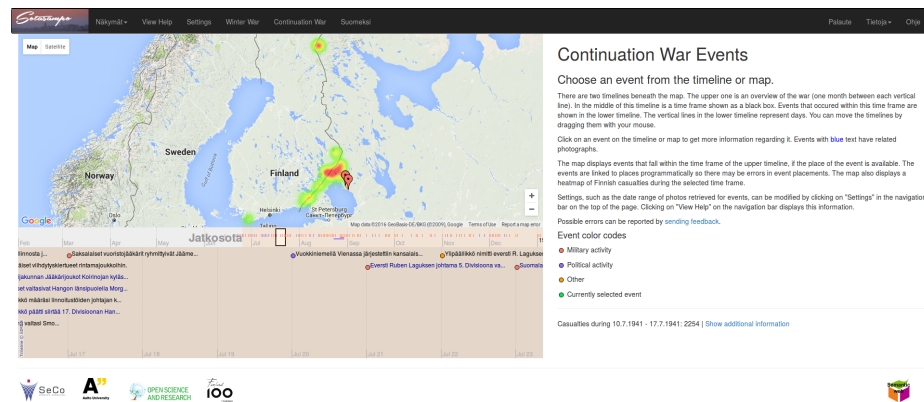


Fig. 3. The casualties of an 8 day time period visualized as a heat map on top of Google Maps, with events visualized as points.

Fig. 4 shows a histogram of the amount of casualties per day of a single military unit, the 33rd infantry regiment, and all of its sub units. The time span

¹⁴ <https://github.com/SemanticComputing/WarSampo-death-records>

covers the whole war, and shows also the war related deaths that occur after the war. Demonstrating the value of linking additional data to the death records, we have information of 7 events that are linked to the military unit and its sub units. They seem to explain quite well the casualties during the Winter War (30 November 1939 – 13 March 1940), as high peaks in casualty amounts occur when there are multiple concurrent battles. Information regarding this unit’s activities during the Continuation War (25 June 1941 – 19 September 1944) are currently lacking. The unit and its sub units are known to have participated in the following events:

1. Defensive battle at Pasuri, 30.11.1939 – 16.02.1940
2. Battle at Punnus, 14.02.1940 – 17.02.1940
3. Battle at Punnus, 15.02.1940 – 16.02.1940
4. Stalling phase from Pasuri sector to Salmenkaita, 16.02.1940, 17.02.1940
5. Defensive battle at Salmenkaita, 17.02.1940 – 28.02.1940
6. Stalling phase from Salmenkaita to Kaskiselkä sector, 28.02.1940 – 29.02.1940
7. Defensive battle at Kaskiselkä, 02.03.1940 – 13.03.1940

The histogram is created by reading data directly from the WarSampo SPARQL endpoint and visualizing it with YASGUI online SPARQL tool,¹⁵ and the following SPARQL query:

```
PREFIX atypes: <http://ldf.fi/warsa/actors/actor_types/>
PREFIX crm: <http://www.cidoc-crm.org/cidoc-crm/>
PREFIX casualties: <http://ldf.fi/schema/narc-menehtyneet1939-45/>

SELECT ?date (count(?cas) as ?casualties)
WHERE {
  { SELECT ?subunit
    WHERE {
      VALUES ?unit { <http://ldf.fi/warsa/actors/actor_1331> } .
      ?unit (^crm:P144_joined_with/crm:P143_joined)+ ?subunit .
      ?subunit a atypes:MilitaryUnit .
    }
  } UNION {
    VALUES ?subunit { <http://ldf.fi/warsa/actors/actor_1331> } .
  }
  ?cas casualties:osasto ?subunit .
  ?cas casualties:kuolinaika ?date .
} GROUP BY ?date ORDER BY ?date
```

4 Use Case 2: Reassembling Soldier Biographies and Military Unit Histories

Linking the death records to information about the same persons in other sources, events, military units, war diaries, and wartime places provides new information about the persons involvement in the war events, his whereabouts and movements during the war. By linking all these pieces of information together, we are able to construct partial biographies of individual soldiers, and the movements

¹⁵ <http://yasgui.org>

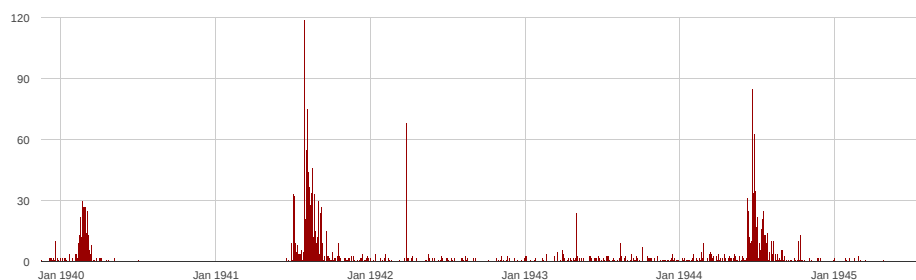


Fig. 4. A histogram visualization showing casualties of the 33rd infantry regiment.

and actions of their military units. This allows an individual who is interested in investigating the biography of a relative who took part in the war to look at where the person probably fought, with whom, and when, and in what events his military unit participated. Also, the interlinked dataset, together with applications to effectively use it, provides digital humanities researchers with new perspectives to study casualties, that would not be possible with a non-linked casualty dataset.

Fig. 5 depicts an integrated view of information related to a casualty in the persons perspective of the WarSampo portal. Basic information about a person (e.g., name, birth and death dates and places, occupation, marital status, military rank with promotion dates if available) is displayed on top of the page. After that, thumbnails of the linked photographs involving the person are shown. By clicking the thumbnails the user can explore the higher resolution versions of the photographs and their captions. In order to get further context to the person examined, the user can browse the army units the person belonged to during the war, and places related to his life events (e.g., birth and death municipalities on historical and contemporary maps). This way the user is able to track the person's participation in the war by investigating the movements of his army units and the durations of the battles the units fought.

The units perspective of WarSampo visualizes the troop actions both on a map and a timeline, and shows the casualties of the unit as a heat map in the same fashion as in the event perspective. For even deeper understanding of the history, links to digitized images of the war diaries of the army units are provided, containing rich primary source descriptions of the events. By following the municipalities links to the places perspective of WarSampo, one can, e.g., study what kind of war events took place in the person's birth place, see photographs taken at specific locations the troops were located in, or read magazine articles depicting wartime events that took place in some place.

5 Discussion

War history data is usually scattered in many isolated silos and may exist in totally different formats, e.g., books, paper archives, and databases. In this paper

Persons
Search for known persons from the past Finnish wars by writing their name in the text input below and/or selecting a person from the list below. Information regarding the person and recommended links will appear on the right.

niilo korhonen
Niilo Antero Korhonen
Niilo Ensio Korhonen
Niilo Juhani Korhonen
Niilo Juho Korhonen

Niilo Juhani Korhonen
15.12.1915 Nilsä - 26.9.1943
Kuolinny muu kuin sodasta välittömästi aiheutunut (sairaus, tapaturma). Buried in Kuopio
Profession: opettaja
Marital status: naimisissa, 0 children
Served in 9./JR 10
Luutnantti
Source: Mannerheim-ristin ritarit. <http://www.mannerheim-ristin/ritarit.fi/ritarit>
Show information page

7 photos

Units (2)

Jalkaväkirykmentti 10, III Pataljoona, 9. Komppania
Jalkaväkirykmentti 10, Jalkaväkirykmentti 10

Ranks (1)

Luutnantti

Places (3)

Nilsä
Kuopio
Kuopio

Fig. 5. A screenshot of the Persons perspective of the WarSampo portal depicting linked information related to a casualty.

we have examined the benefits and challenges of linking casualty records of war to multiple related datasets, and publishing them as linked data for DH research and applications to use. Two use cases were studied related to supporting DH research and services for the public.

In the future, linking with other WarSampo data will be developed further as new datasets are added to the whole. We plan to develop tools for statistical analysis of the data, and collaborate with humanities researchers in studying how linked data and our tooling can help to solve their research problems. An hierarchical occupation ontology is planned to be created and linked to the death records to provide insight into the social status of each casualty.

Acknowledgements Tomi Ahoranta and Jérémie Dutruit from the National Archives contributed in publishing the original casualty database as LOD.

Our research is partially funded through the Finnish Open Science and Research Initiative by the Finnish Ministry of Education and Culture.

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. ACM (2013)
2. Collins, T., Mulholland, P., Zdrahal, Z.: Semantic browsing of digital collections. In: Proceedings of the 4th International Semantic Web Conference (ISWC 2005). pp. 127–141. Springer-Verlag (2005)

3. Crymble, A., Gibbs, F., Hegel, A., McDaniel, C., Milligan, I., Posner, M., Turkel, W.J. (eds.): *The Programming Historian*. 2 edn. (2015), <http://programminghistorian.org>
4. Doerr, M.: The CIDOC CRM—an ontological approach to semantic interoperability of metadata. *AI Magazine* 24(3), 75–92 (2003)
5. Graham, S., Milligan, I., Weingart, S.: *Exploring big historical data. The historian’s macroscope*. Imperial College Press (2015)
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. Hyvönen, E., Heino, E., Leskinen, P., Ikkala, E., Koho, M., Tamper, M., Tuominen, J., Mäkelä, E.: *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 (2016)
8. Hyvönen, E., Ikkala, E., Tuominen, J.: *Linked data brokering service for historical places and maps*. In: *Proceedings of the 1st Workshop on Humanities in the Semantic Web - WHiSe (2016)*
9. Koho, M., Heino, E., Hyvönen, E.: *SPARQL Faceter - client-side faceted search based on SPARQL (2016)*, submitted for review
10. Mäkelä, E.: *Combining a REST lexical analysis web service with SPARQL for mashup semantic annotation from text*. In: *The Semantic Web: ESWC 2014 Satellite Events*, pp. 424–428. Springer (2014)
11. Mäkelä, E., Törnroos, J., Lindquist, T., Hyvönen, E.: *World War I as Linked Open Data (2015)*, <http://seco.cs.aalto.fi/publications/>, submitted for review
12. Tunkelang, D.: *Faceted search. Synthesis lectures on information concepts, retrieval, and services* 1(1), 1–80 (2009)

