Modeling and Using an Actor Ontology of Second World War Military Units and Personnel

Petri Leskinen[1(✉)], Mikko Koho[1], Erkki Heino[1,2], Minna Tamper[1,2], Esko Ikkala[1], Jouni Tuominen[1,2], Eetu Mäkelä[1,2], and Eero Hyvönen[1,2]

[1] Semantic Computing Research Group (SeCo), Aalto University, Espoo, Finland
{petri.leskinen,mikko.koho,erkki.heino,minna.tamper,esko.ikkala,
jouni.tuominen,eetu.makela,eero.hyvonen}@aalto.fi
[2] HELDIG – Helsinki Centre for Digital Humanities,
University of Helsinki, Helsinki, Finland
http://seco.cs.aalto.fi, http://heldig.fi

Abstract. This paper presents a model for representing historical military personnel and army units, based on large datasets about World War II in Finland. The model is in use in WarSampo data service and semantic portal, which has had tens of thousands of distinct visitors. A key challenge is how to represent ontological changes, since the ranks and units of military personnel, as well as the names and structures of army units change rapidly in wars. This leads to serious problems in both search as well as data linking due to ambiguity and homonymy of names. In our solution, actors are represented in terms of the events they participated in, which facilitates disambiguation of personnel and units in different spatio-temporal contexts. The linked data in the WarSampo Linked Open Data cloud and service has ca. 9 million triples, including actor datasets of ca. 100 000 soldiers and ca. 16 100 army units. To test the model in practice, an application for semantic search and recommending based on data linking was created, where the spatio-temporal life stories of individual soldiers can be reassembled dynamically by linking data from different datasets. An evaluation is presented showing promising results in terms of linking precision.

Keywords: Semantic web · Linked open data · Actor ontology · Digital humanities · Biographic representation

1 Introduction

Authority files [18], vocabularies (e.g., ULAN[1]), and actor ontologies (e.g. FOAF[2], REL[3], BIO[4], schema.org [5]) are used for (1) identifying people, groups,

---

[1] http://www.getty.edu/research/tools/vocabularies/ulan/about.html.
[2] http://xmlns.com/foaf/spec/.
[3] http://vocab.org/relationship/.
[4] http://vocab.org/bio/.

and organizations and (2) for representing data about them. They constitute a central resource for cataloging and information management in museums, libraries, and archives, but also a challenge for data linking due to alternative names, homonyms, spelling variations, different languages, transliteration rules, and changes in time. Although actor ontologies play an essential part in modeling historical information, there are still very few published scientific articles about the subject.

Historical military units and personnel is a particularly challenging domain for creating an actor ontology: the structures of units are large and change rapidly, different codes can be used for actors in order to confuse the enemies, and people come and go due to the violent actions of war. For example, during the phases of WW2 in Finland (The Winter War, The Continuation War, and The Lapland War) different units have used the same name, and during Winter War in Finland the names of major units were changed just to bluff the enemy. Furthermore, the data about the actors is often incomplete and uncertain, involving lots of "unknown soldiers" of whom little is known.

From a Linked Data viewpoint this poses two major problems: (1) Data linking (based on named entity linking [2,6]) is difficult, because it has to be done in a changing and vague domain specific contexts [7]. For example, to tell whether a mention *captain Smith* and *colonel Smith* can refer to the same person, and to which *Smith* in the first place, data about different Smiths and their ranking history in time is needed. (2) It is difficult to aggregate and enrich data about actors that come from different sources and in different documentary forms, such as death records, diaries, magazine articles, or photographs, and to compile the global biographical history of the actors to the end users [9].

We argue that to address the problems above, a semantically rich spatio-temporal model for representing actors in relation to the events of the war is needed. This paper contributes to the state-of-the-art by presenting such an ontological actor model for historical military units and personnel. The model is in use in end-use application perspectives of the semantic portal WarSampo[5], where the idea is to reassemble automatically the biographical war history of individual soldiers and units. The model enables disambiguation of names in spatio-temporal contexts as well as combining contents from various sources, and publishing them in a harmonized format. The ontology and related data has been published as a Linked Open Data service[6] that can and has been used in digital humanities research and as well for developing online portals. For example, the community portal Sotapolku[7], provided by a commercial company, makes use of the WarSampo actor data.

The work is done as part of the WarSampo project[8], and builds upon our previous publications [7,9,11,12], which focus on the architecture, named entity

---

[5] Sotasampo in Finnish; available with an English GUI at http://www.sotasampo.fi/en, but the content is in Finnish.

[6] http://www.ldf.fi/dataset/warsa.

[7] http://sotapolku.fi.

[8] http://seco.cs.aalto.fi/projects/sotasampo/en/.

linking, and end-user views of the application. In contrast, this paper represents the underlying ontology model and dataset regarding army units and people in detail, as well as the actor related application perspectives in use.

The paper is structured as follows: First, ontology model for representing army units, and military personnel, is presented. After this the collecting of WarSampo actor dataset is represented, and a brief look on person and unit perspectives at WarSampo portal is taken. In conclusion, contributions of the work are summarized and some directions for further research are suggested.

2   Use Case and Datasets

The use case for our work is the WarSampo semantic portal[9] [9]. It provides the end user with richly interlinked data about the WW2 in Finland via application perspectives in the Sampo model [8]. An illustration of the WarSampo datasets is represented in Fig. 1. In total, the WarSampo data cloud contains data of more than a dozen different types (e.g. casualty data, photographs, events, war diaries, and historical maps) from an even larger pool of sources (e.g. the National Archives, the Defense Forces, and scanned books, from which part of the data has been extracted semi-automatically).

The actor dataset contains ca. 100 000 soldiers, and ca. 16 100 army units. The data is enriched with ca. 488 000 links from events to actors. Actors have furthermore been linked to external resources in the LOD cloud databases on the Web.
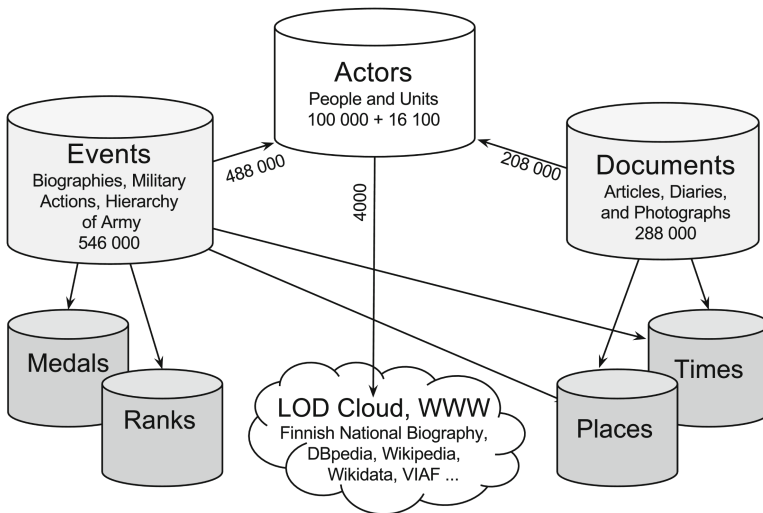

Fig. 1. Linkage in the actor-event based dataset

---

3   Actor Ontology Model

The ontology of actors is based on the CIDOC CRM[10] [4] model, where the
resources of actors are essentially described in terms of the spatio-temporal
events they participate in. An event represents any change of status that divides
the timeline into periods before and after the event. Using the actor-event-model
facilitates reconstructing the status of an actor at a specified moment. One main
reason for adapting the model is that the information regarding a single actor
varies a lot in both form and amount; in some cases we may have access to a
very detailed description of the actor's biography, in some other cases only sparse
pieces of information exist. All this data can be harmonized into a sequence of
events. The applied actor-event-model also allows us to easily add new event
types to the schema and new events the to database.
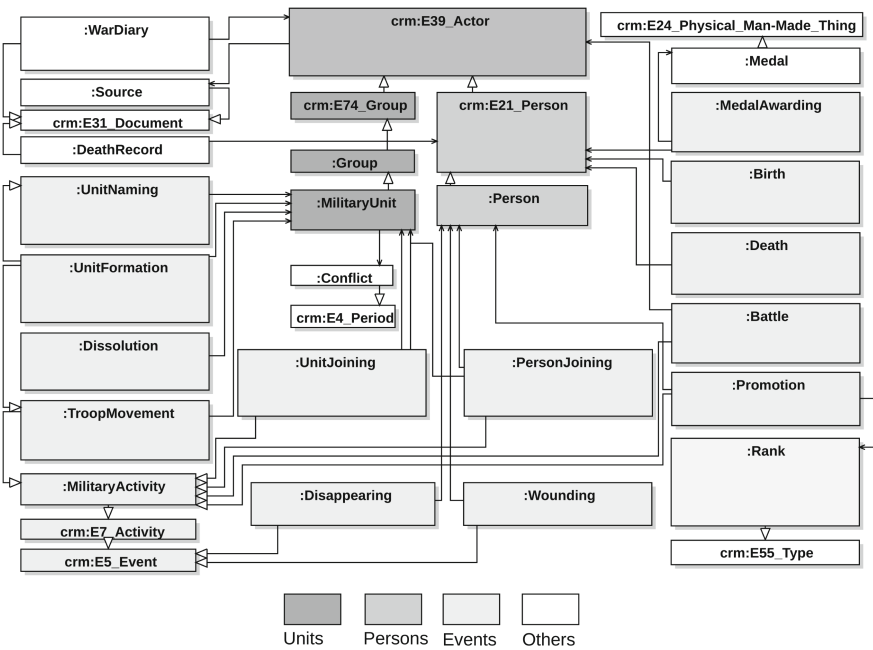
Fig. 2. Ontology schema of actors and events

Schema of the ontology is illustrated in Fig. 2. The schema is available at
http://ldf.fi/schema/warsa, the namespaces and prefixes in use are listed in
Table 1. The actor superclass **crm:E39_Actor**[11] is shown at center on the top.
There is one subclass for people, and two for groups. For various types of events
there are 19 classes with superclass **:Event**[12].

---

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

[11] http://www.cidoc-crm.org/Entity/e39-actor/version-6.2.

[12] http://www.cidoc-crm.org/Entity/e5-event/version-6.2.

Table 1. Namespaces and prefixes used in actor ontologies

Namespace	Prefix
http://ldf.fi/schema/warsa/	:
http://www.cidoc-crm.org/cidoc-crm/	crm:
http://purl.org/dc/elements/1.1/	dc:
http://purl.org/dc/terms/	dct:
http://xmlns.com/foaf/0.1/	foaf:
http://rdf.muninn-project.org/ontologies/organization#	mil:
http://www.w3.org/2002/07/owl#	owl:
http://www.w3.org/1999/02/22-rdf-syntax-ns#	rdf:
http://www.w3.org/2004/02/skos/core#	skos:

The biographical representation of a person was modeled with events of birth (**:Birth**), and death (**:Death**), and his military career with events like promotion (**:Promotion**), serving in an army unit (**:UnitJoining**), participating in battles (**:Battle**), or getting awarded with a medal of honor (**:MedalAwarding**). Furthermore, there are classes for getting wounded (**:Wounding**) or disappearing (**:Disappearing**), which represent the data fields in Casualties database. The schema includes supporting classes for representing military ranks, war diary entries, medals of honor, documentation, and data sources.

Example of a person resource[13] (**:Person**) is shown in Table 2. The principle is to represent only constant information in a person resource; it has full name as a primary title, and the family and first names as separate fields. Property **owl:sameAs** links to a corresponding resource in external databases, and **foaf: page** to external web pages.

Examples of related events are shown in Table 3. During the war, the person in example has been promoted from lieutenant first to captain and finally to major. When the Winter War started in 1939 he served as a commander in an air force squadron, and shot down an enemy aircraft soon after.

In literature military personnel are ofter referred using a combination of current military rank and family name (e.g. *Captain Karhunen* or *Colonel Talvela*). So, to describe a person in detail, an ontology of military ranks was needed. The rank ontology is based on Muninn Military Ontology [19]. The hierarchy of ranks was constructed by interlinking the instances to equal and lower ranks. The **:Rank** instances in the datasets were enriched with additional information (e.g. countries or service branches in which the rank has been used, or categories like officer or non-commissioned officer). Event **:Promotion** was used to attach a rank to a person. Due to the variations in the amount of available data, a promotion event was created in all cases, even if a person is known to have only a single rank with no specific date of promotion.

---

[13] http://ldf.fi/warsa/actors/person_294.ttl.

Table 2. Properties of a resource describing pilot Jorma Karhunen

Property	RDF identifier	Value
Primary title	skos:prefLabel	"Jorma (Joppe) Karhunen"
Family Name	foaf:familyName	"Karhunen"
First name (Nickname)	foaf:firstName	"Jorma (Joppe)"
Text description	dc:description	"Jorma Karhunen was a Finnish Air ..."@en
External LOD-links	owl:sameAs	http://dbpedia.org/resource/Jorma_Karhunen http://wikidata.org/entity/Q5482501
Related websites	foaf:page	https://en.wikipedia.org/wiki/Jorma_Karhunen www.mannerheim-ristinritarit.fi/ritarit?xmid=38

Table 3. Examples of events describing pilot Jorma Karhunen

Event description/resource URI	RDF class	Date
Born at Pyhäjärvi http://ldf.fi/warsa/events/birth_294.ttl	:Birth	1913-03-17
Serving as a squadron commander in 24th Fighter Squadron http://ldf.fi/warsa/events/joining_294_459.ttl	:PersonJoining	1939-11-30
Aerial victory in Tainionkoski: enemy SB-2 shot down http://ldf.fi/warsa/events/event_lv2408.ttl	:Battle	1939-12-01
Promotion to captain http://ldf.fi/warsa/events/kapteeni_294.ttl	:Promotion	1941-08-04
Photograph of capt. Karhunen with his dog Becky Brown http://ldf.fi/warsa/photographs/sakuva_7265.ttl	:Photography	1942-06-01
Awarded with the Mannerheim Cross of Liberty http://ldf.fi/warsa/medals/medal_83_294.ttl	:MedalAwarding	1942-09-08
Died at Tampere http://ldf.fi/warsa/events/death_294.ttl	:Death	2002-01-18

An example of RDF resource of a military unit is shown in Table 4, the resource is also available in Turtle format[14]. Just like in the case of a person, the properties describe only constant information like unit's preferable name and abbreviation, description, conflicts participated in, and links to LOD cloud resources. The events (Table 5) describe the unit's position in the army hierarchy and the involved military activities. The lifespan of a unit spans from its formation **:UnitFormation** to dissolution **:Dissolution**. The changes of the unit

---

[14] http://ldf.fi/warsa/actors/actor_459.

Table 4. Properties of a resource describing 24th Fighter Squadron

Property	RDF identifier	Value
Preferred label	skos:prefLabel	"Lentolaivue 24"
Preferred abbreviation	skos:altLabel	"LLv 24"
Description	dc:description	"No. 24 Squadron was a fighter ..."@en
Conflict	:hasConflict	wcf:WinterWar, wcf:ContinuationWar, ...
Army postal code	:covernumber	"8523", "8524", "8567"
Unit category	:hasUnitCategory	"Flying Regiments and Squadrons"
External LOD-links	owl:sameAs	https://www.wikidata.org/wiki/Q4356342
Related websites	foaf:page	https://fi.wikipedia.org/wiki/Lentolaivue_24

Table 5. Examples of events describing 24th Fighter Squadron

Event description/resource URI	RDF class	date
Troop founded as 24th Squadron (abbrev. LLv 24) http://ldf.fi/warsa/events/formation_971.ttl	:Formation	1934-10-10
Troop Movement to Immola Air Base http://ldf.fi/warsa/events/concentration_491.ttl	:TroopMovement	1939-10-12
Aerial victory in Tainionkoski: enemy SB-2 shot down http://ldf.fi/warsa/events/event_lv2408.ttl	:Battle	1939-12-01
Being part of Flying Regiment 2 http://ldf.fi/warsa/events/joining_458.ttl	:UnitJoining	1940-01-10
Written War Diary document http://ldf.fi/warsa/diaries/diary_c26701.ttl	:WarDiary	1941-06-19–1941-09-02
Changing the name to 24th Fighter Squadron (HLeLv 24) http://ldf.fi/warsa/events/form_459.ttl	:UnitNaming	1944-02-14

name were modeled as **:UnitNaming** events. Also the army hierarchy, including the temporal changes made in it, was modeled using the event schema: the hierarchy was represented as a tree graph where the army units are the nodes and the events of joining into a superior unit **:UnitJoining** form the edges. The events also included the military activities taken (e.g. movements **:TroopMovement** and battles **:Battle**). The event **:PersonJoining** was used to combine a

person to the unit, in which he has served. The event could also announce a role in the unit (e.g. being a commander or a squadron pilot).

4   Warsampo Actor Data

Currently the actor dataset contains ca. 100 000 people. The data has been collected from various sources: lists of generals and commanders, lists of recipients of honorary medals, the Casualties database[15], Finnish National Biography[16], photographers mentioned in Finnish Wartime Photograph Archive[17], Wikidata[18], and Wikipedia. Besides military personnel, an extract of 580 Finnish or foreign civilians from the National Biography database and Wikidata was included. This set consisted of people with political or cultural significance.

The unit dataset consists of over 16 100 Finnish wartime units, including Land Forces, Air Forces, Navy, Medical Corps, stations of Anti-Aircraft Warfare and Airwarning, Finnish White Guard, and Foreign Volunteer Corps. At this stage Soviet and German troops were excluded. The main sources of information have been the War Diaries, Army Postal Code list[19], and Organization Cards, all of which provided the information as datasheets in CSV format.

In general, the method to produce the data depended on the format of data source. The biographies of the National Biography and the Casualties Database had been transformed into LOD in our earlier projects, and therefore the information extraction process was to convert the existing data into new actor entries and relating events. Transformation was mostly done by using specific SPARQL construct queries. More than 95 000 entries were generated from the Casualty Database to actor dataset [12].

The organization cards (Fig. 3) were written by Finnish Defense Forces shortly after the WW2. The cards contain the major part of units in Finnish Army, unfortunately not those of Navy and Air Force. An example of organization card is shown in Fig. 3. The proper name and abbreviation of the unit is shown at the upper left corner (a), in this case *Jalkaväkirykmentti 7* (7th Infantry Regiment), abbreviated as *JR 7*. The regiment has been part of *3. divisioona* (3rd Division), which is told at the upper right corner (b). The card provides further information about the foundation (c) and the military district (d) of the unit. Changes considering the unit, like different names, are shown at part (e). During the Winter War *JR 7* participated in four battles (f). The three columns on each line show the location or a short description of the battle, battle's duration, and the name of the commanding officer.

The organization cards were provided as scanned booklets in PDF format, and converting to RDF had several steps. Firstly each page in PDF booklet

---

[15] http://kronos.narc.fi/menehtyneet/.

[16] http://www.kansallisbiografia.fi/english/.

[17] http://sa-kuva.fi/neo?tem=webneoeng.

[18] https://www.wikidata.org/.

[19] http://www.arkisto.fi/uploads/Aineistot/kopsa[1].pdf.

was written as an individual PNG image. Images were preprocessed by adjusting the contrast and image rotation, and removing the compression artifacts. Next an *Optical character recognition (OCR)* process was applied. The resulting text was however very erroneous, and plenty of post-processing was required. The structured format of the cards, and the recurring use of military terms in the vocabulary however eased the automated error fixing. From the resulting text, the fields a-f (in Fig. 3) were extracted, and converted into RDF. The produced resources consisted of military units (**:MilitaryUnit**), their commanders (**:Person**) with ranks (**:Promotion**), and events like unit formations (**:UnitFormation**), joinings of units (**:UnitJoining**), movements (**:TroopMovement**), renamings (**:UnitNaming**), and battles (**:Battle**).

Fig. 3. Information on an organization card

Although the Wikipedia may not be considered as the most reliable source of information, it provided a way to connect data with external LOD cloud databases Wikidata, DBpedia[20], and VIAF[21]. The material regarding personnel was widely available, but for units, specially those of Finnish Army during the WW2, the information was sparse. Information was extracted from Wikipedia pages of e.g. Finnish high-ranking officers, politicians, wartime casualties, and foreign volunteers. The pages of Wikipedia follow a structured layout which facilitated extracting the information. In case of military units, detailed information for events like unit foundation, troop movements, battles, and for names of commanding officers were available. In total 2500 people and 480 units with 5000 events were generated from corresponding Wikipedia pages.

Characteristic sentences picked from Wikipedia were descriptions like "*1st Artillery Group was founded in Pori with Captain Paavo Suominen as the first commander*", "*10th July 1941 Regiment was moved to Kitee, from where it begun attacking towards Lake Ladoga*", or "*Regiment participated in the occupation of Prääsä September 7–8, 1941*". Each sentence was converted to an event, and

---

[20] http://wiki.dbpedia.org/.
[21] https://viaf.org/.

the named entities of personnel, places and dates were recognized and linked to database resources. The data retrieval was done using Python scripts utilizing MediaWiki API[22], and Wikipedia API for Python[23]. Entity linking was done with ARPA service [13].

The datasets of conflicts, war diaries, medals, and ranks are in separate graphs. Conflicts[24] contain four main periods of WW2 in Finland. The War Diary graph[25] has 26 400 entries. There are 200 medal types[26] and 200 rank entries[27]. The data includes ranks used by the Finnish Military with most common German and Soviet ranks, among with some civil titles (e.g. the ones used by women's voluntary association *Lotta Svärd*) [9].

5   The WarSampo Portal

The perspectives at WarSampo portal[28] visualize the linkage between the various datasets (e.g. military unit, personnel, casualties, events, places) etc [9,11]. WarSampo portal is a Rich Internet Application (RIA), where all functionality is implemented on the client side using JavaScript with AngularJS framework, only data is fetched from the server side SPARQL endpoints.

5.1   The Person Perspective

The WarSampo person perspective application[29] is illustrated in Fig. 4. A typical use case is someone searching for information about a relative who served in the army. On the left, the page has an input field (a) for a search by person's name. The matching query results are shown in the text field (b) below the input. After making a selection, information about the person is shown at the center top of the page (c). The tabs (d) allow the user to switch between this information page or a map-timeline application. In the example case, the page shows description of the person (e), photograph gallery (f), lists linking to related events (g), military units (h), battles (i), ranks (j), medals (k), related people (l), places (m), Wikipedia page (n), related Kansa Taisteli magazine articles (o), and a Finnish National Biography widget (p).

As an example of SPARQL query, the query fetching related people[30] defines a similarity measure between two people. The more events, medals, units, and the higher ranks the two share in common, the higher the similarity gets. The list of related people (l) shows the results sorted in descending order.

---

[22] https://en.wikipedia.org/w/api.php.
[23] https://pypi.python.org/pypi/wikipedia.
[24] See, e.g., http://ldf.fi/warsa/conflicts/LaplandWar.
[25] See, e.g., http://digi.narc.fi/digi/hae_ay.ka?sartun=319.SARK.
[26] See, e.g., http://ldf.fi/warsa/medals/medal_83.
[27] See, e.g., http://ldf.fi/warsa/actors/ranks/Majuri.
[28] http://www.sotasampo.fi/en/.
[29] http://www.sotasampo.fi/en/persons.
[30] http://yasgui.org/short/B1w2O71gb.

Fig. 4. Information on person perspective

WarSampo military unit perspective application[31] is illustrated in Fig. 5. In a typical use case someone searches for information about an army unit, where perhaps an elder relative has served during the wartime. On the left

there is an input field (a) for a search by unit's name. The matching results are shown in the text field (b) below the input. The map (c) depicts the known locations of the unit. The heatmap shows the casualties of the unit, and the timeline (d) the events (e), e.g. dates of unit foundations, troop movements, and durations of fought battles. On the right there are unit names and abbreviations (f), description (g), and a collection of related photographs (h). Three lists of related units are shown: larger groups in which the unit has been as a member (i), subdivisions being parts of the unit (j), and units at the same hierarchical level (k). Below there are fields for related battles (l), links to Kansa Taisteli magazine articles (m), Wikipedia page (n), and War Diaries (o). The number of casualties during the specified time is shown at the bottom of page (p).

5.2   The Military Unit Perspective

See Fig. 5.
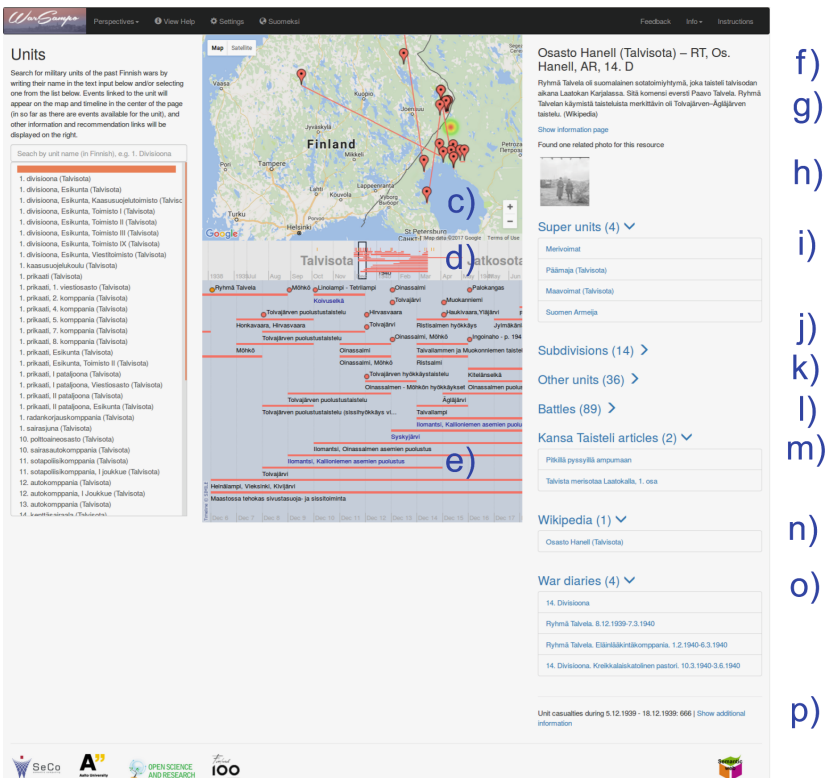

Fig. 5. Information on unit Perspective

5.3   The Kansa Taisteli Magazine Perspective

Kansa Taisteli is a magazine published by Sanoma Ltd and Sotamuisto association between 1957 and 1986. The magazine articles cover the memoirs of WW2

from the point of view of Finnish military personnel and civilians. The articles contain mentions of people, military units, and places. From these the military units and personnel have been linked to Actor ontology. The magazine perspective[32] can be used for searching and browsing articles relating to WW2. Military units and personnel are used as separate facets to search for articles. In addition, writers have been linked to Actor ontology as well.

Fig. 6. The Contextual Reader interface targeting the Kansa Taisteli magazine articles

The purpose of the perspective is two-fold: (1) to help a user find articles of interest using faceted semantic search and, (2) to provide context to the found articles by extracting links to related WarSampo data from the texts. The start page of the magazine article perspective is a faceted search browser. Here, the facets allow the user to find articles by filtering them based on author, issue, year, related place, army unit, or keyword. Some of the underlying properties, such as the year and issue number, are hierarchical and represented using SKOS. The hierarchy is visualized in the appropriate facet, and can be used for query expansion: by selecting an upper category in the facet hierarchy one can perform a search using all subcategories.

After the user has found an article of interest, she can click on it, and the article appears on the screen in the CORE Contextual Reader interface [14]. Depicted in Fig. 6, CORE is able to automatically and in real time annotate PDF and HTML documents with recognized keywords and named entities, such as army units, places, and person names. These are then encircled with colored boxes indicating the linked data source. By hovering the mouse over a box, data is shown to the user, providing contextual information for an enhanced read-

---

ing experience. In Fig. 6, for example, detailed data are shown about *Raymond August Ericsson*, one of the battalion commanders discussed in the article.

Solving the technical issues, however still left the problem of semantic disambiguation; in this case this concerned named entity recognition of correct people and military units. The identification was made by customizing the SPARQL queries, the order of the queries, and the article metadata. Each magazine article was identified and firstly references to people were searched from the text. The identification of people was done by using name and possibly a rank. Secondly the linking of the military units was performed from the remaining text. The article metadata was also used to identify the war to which the events of the article are related to. Afterwards the military units were linked based on the war into the corresponding units. A detailed description and evaluation of the process is available at [17].

5.4 Photographs

WarSampo contains a dataset of the metadata of ca. 160 000 historical photographs taken by Finnish soldiers during WWII. The data contains e.g. captions of the photographs. The actor ontology was used to automatically disambiguate and link people and military units mentioned in the metadata. Information in the actor ontology was used extensively in linking: For example, when disambiguating people, names, ranks, promotion dates, military units, sources, medals, and death dates were used to rank otherwise ambiguous mentions in the photograph captions [7].

The results of the linking can be seen in the person and unit perspectives of the WarSampo portal, as well as in the photograph perspective itself[33] which provides a faceted search interface for the photographs.

6 Related Work, and Discussion

There are several projects publishing linked data about the World War I on the web, such as Europeana Collections 1914–1918[34], 1914–1918 Online[35], WW1 Discovery[36], Out of the Trenches[37], Muninn [19], and WW1LOD [15]. There are few works that use the Linked Data approach to World War II, such as [1,3], Defence of Britain[38], and Open Memory Project[39]. The main focus on our work

---

[33] http://www.sotasampo.fi/en/photographs.
[34] http://www.europeana-collections-1914-1918.eu.
[35] http://www.1914-1918-online.net.
[36] http://ww1.discovery.ac.uk.
[37] http://www.canadiana.ca/en/pcdhn-lod/.
[38] http://thesaurus.historicengland.org.uk/thesaurus.asp?thes_no=365&thes_name=Defence%20of%20Britain%20Thesaurus.
[39] http://www.bygle.net/wp-content/uploads/2015/04/Open-Memory-Project_3-1.pdf.

is on representing an actor as a biographical life story, unlike databases like Getty ULAN or Smithsonian American Art Museum [16] that have actor vocabularies.

Our research group, Semantic Computing Research Group (SeCo), has produced several projects with highly interlinked actor ontologies: The National Biography, CultureSampo[40], BookSampo[41], and Norssit—High School Alumni [10] datasets. Bio CRM model[42] is developed to facilitate and harmonize the representation of an actor in semantic web, and therefore deals with the same problematics as the WarSampo actor ontology.

We have considered combining the different datasets like articles and photographs to actor ontology as one of the use-cases of the actor ontology. The evaluation of the ontology and actor dataset, has been work- and data-driven e.g. it has developed to the needs of semantically representing the data and of rendering the data at the end-user portal. 94% of users come from Finland and 25% of them are returning visitors. We have received feedback via the user interface, and we have considered their comments e.g. on misidentified people.

Main requirement for the ontology was to represent changes in spatiotemporal context as described in Introduction. Constant actor resources are enriched with events marking the changes in spatio-temporal continuity, adding details to the semantic biographical representation, and connecting the otherwise separate datasets of personnel, units, places, articles, photographs etc. The unit model had to be capable of representing even more dynamical changes than with people; identifiers like name and abbreviation may change in the time domain. The army hierarchy is represented as a tree graph where the groups are connected by the events of joining.

The actor ontology is based on CIDOC CRM standard which provides a clear framework and basis for actor-event schema. The Muninn Military Ontology offered an example of modeling military concepts semantically. In conclusion, there was no obvious basis for the ontology. On the contrary, it was constructed by combining principles of several solutions all serving different needs.

In a similar way Warsampo project has collected historical, wartime information from Finland. There is abundance of information about the WW2 in different countries, written in local languages, and published in various formats; often even having divergent points of view. Collecting the data and publishing it as LOD forms a tremendous field of work, but aims at constructing a comprehensive, worldwide database. In the events of history, individual people and groups are at the focal center; it is from their point of view that we build our notion of history.

The ontology model represented in this article may not be all-purpose suitable, but we encourage and hope to inspire the researchers to develop the ideas further.

---

[40] http://seco.cs.aalto.fi/applications/kulttuurisampo/.
[41] http://seco.cs.aalto.fi/applications/kirjasampo/.
[42] http://seco.cs.aalto.fi/projects/biographies/.

References

1. de Boer, V., van Doornik, J., Buitinck, L., Marx, M., Veken, T.: Linking the kingdom: enriched access to a historiographical text. In: Proceedings of the 7th International Conference on Knowledge Capture (KCAP 2013), pp. 17–24. ACM, June 2013

2. Bunescu, R.C., Pasca, M.: Using encyclopedic knowledge for named entity disambiguation. In: EACL, vol. 6, pp. 9–16 (2006)

3. Collins, T., Mulholland, P., Zdrahal, Z.: Semantic browsing of digital collections. In: Gil, Y., Motta, E., Benjamins, V.R., Musen, M.A. (eds.) ISWC 2005. LNCS, vol. 3729, pp. 127–141. Springer, Heidelberg (2005). doi:10.1007/11574620_12

4. Doerr, M.: The CIDOC CRM - an ontological approach to semantic interoperability of metadata. AI Mag. **24**(3), 75–92 (2003)

5. Guha, R.V., Brickley, D., Macbeth, S.: Schema.org: evolution of structured data on the web. Commun. ACM **59**(2), 44–51 (2016)

6. Hachey, B., Radford, W., Nothman, J., Honnibal, M., Curran, J.R.: Evaluating entity linking with Wikipedia. Artif. Intell. **194**, 130–150. http://dx.doi.org/10.1016/j.artint.2012.04.005

7. Heino, E., Tamper, M., Mäkelä, E., Leskinen, P., Ikkala, E., Tuominen, J., Koho, M., Hyvönen, E.: Named entity linking in a complex domain: case second world war history. In: Gracia, J., Bond, F., McCrae, J.P., Buitelaar, P., Chiarcos, C., Hellmann, S. (eds.) LDK 2017. LNCS, vol. 10318, pp. 120–133. Springer, Cham (2017). doi:10.1007/978-3-319-59888-8_10

8. Hyvönen, E.: Cultural heritage linked data on the semantic web: Three case studies using the sampo model (2017). http://seco.cs.aalto.fi/publications/submitted/hyvonen-vitoria-2017.pdf. Invited talk, Proceedings of the VIII Encounter of Documentation Centres of Contemporary Art: Open Linked Data and Integral Management of Information in Cultural Centres Artium, Vitoria-Gasteiz, Spain, 2016. Forth-coming

9. 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: Sack, H., Blomqvist, E., d'Aquin, M., Ghidini, C., Ponzetto, S.P., Lange, C. (eds.) ESWC 2016. LNCS, vol. 9678, pp. 758–773. Springer, Cham (2016). doi:10.1007/978-3-319-34129-3_46

10. Hyvönen, E., Leskinen, P., Heino, E., Tuominen, J., Sirola, L.: Reassembling and enriching the life stories in printed biographical registers: Norssi high school alumni on the semantic web. In: Gracia, J., Bond, F., McCrae, J.P., Buitelaar, P., Chiarcos, C., Hellmann, S. (eds.) LDK 2017. LNCS, vol. 10318, pp. 113–119. Springer, Cham (2017). doi:10.1007/978-3-319-59888-8_9. http://ldk2017.org/

11. 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), CEUR Workshop Proceedings, vol. 1608, pp. 39–52 (2016). http://ceur-ws.org/Vol-1608/#paper-06

12. Koho, M., Hyvönen, E., Heino, E., Tuominen, J., Leskinen, P., Mäkelä, E.: Linked death - representing, publishing, and using second world war death records as linked open data. In: WHiSe@ ESWC, pp. 3–14 (2016)

13. Mäkelä, E.: Combining a REST lexical analysis web service with SPARQL for mashup semantic annotation from text. In: Presutti, V., Blomqvist, E., Troncy, R., Sack, H., Papadakis, I., Tordai, A. (eds.) ESWC 2014. LNCS, vol. 8798, pp. 424–428. Springer, Cham (2014). doi:10.1007/978-3-319-11955-7_60
14. Mäkelä, E., Lindquist, T., Hyvönen, E.: CORE - a contextual reader based on linked data. In: Proceedings of Digital Humanities 2016, long papers, July 2016
15. Mäkelä, E., Törnroos, J., Lindquist, T., Hyvönen, E.: WW1LOD - an application of CIDOC-CRM to World War 1 linked data. Int. J. Digital Libr. **2016**, 1–11 (2016)
16. Szekely, P., Knoblock, C.A., Yang, F., Zhu, X., Fink, E.E., Allen, R., Goodlander, G.: Connecting the Smithsonian American art museum to the linked data cloud. In: Cimiano, P., Corcho, O., Presutti, V., Hollink, L., Rudolph, S. (eds.) ESWC 2013. LNCS, vol. 7882, pp. 593–607. Springer, Heidelberg (2013). doi:10.1007/978-3-642-38288-8_40
17. Tamper, M., et al.: AATOS – a configurable tool for automatic annotation. In: Gracia, J., Bond, F., McCrae, J.P., Buitelaar, P., Chiarcos, C., Hellmann, S. (eds.) LDK 2017. LNCS, vol. 10318, pp. 276–289. Springer, Cham (2017). doi:10.1007/978-3-319-59888-8_24
18. Taylor, A.: Introduction to cataloging and classification. Library and Information Science Text Series, Libraries Unlimited (2006)
19. Warren, R.: Creating specialized ontologies using Wikipedia: the Muninn experience. In: Proceedings of Wikipedia Academy: Research and Free Knowledge (WPAC2012), Berlin, DE (2012). http://hangingtogether.org