

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

Petri Leskinen¹, Mikko Koho¹, Erkki Heino^{1,2}, Minna Tamper^{1,2}, Esko Ikkala¹, Jouni Tuominen^{1,2}, Eetu Mäkelä^{1,2}, and Eero Hyvönen^{1,2}

¹ Semantic Computing Research Group (SeCo), Aalto University, Finland and
² HELDIG – Helsinki Centre for Digital Humanities, University of Helsinki, Finland
<http://seco.cs.aalto.fi>, <http://heldig.fi>
`firstname.lastname@aalto.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., ULAN³), and actor ontologies (e.g. FOAF⁴, REL⁵, BIO⁶, schema.org [5]) are used for 1) identifying people, groups, 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,

³ <http://www.getty.edu/research/tools/vocabularies/ulan/about.html>

⁴ <http://xmlns.com/foaf/spec/>

⁵ <http://vocab.org/relationship/>

⁶ <http://vocab.org/bio/>

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 [6,2]) 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 [10].

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 WarSampo⁷, 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 service⁸ that can and has been used in digital humanities research and as well for developing online portals. For example, the community portal Sotapolku⁹, provided by a commercial company, makes use of the WarSampo actor data.

The work is done as part of the WarSampo project¹⁰, and builds upon our previous publications [12,7,8,10], which focus on the architecture, named entity 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

⁷ Sotasampo in Finnish; available with an English GUI at <http://www.sotasampo.fi/en>, but the content is in Finnish.

⁸ <http://www.ldf.fi/dataset/warsa>

⁹ <http://sotapolku.fi>

¹⁰ <http://seco.cs.aalto.fi/projects/sotasampo/en/>

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 portal¹¹ [10]. It provides the end user with richly interlinked data about the WW2 in Finland via application perspectives in the Sampo model [9]. An illustration of the WarSampo datasets is represented in Figure 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.

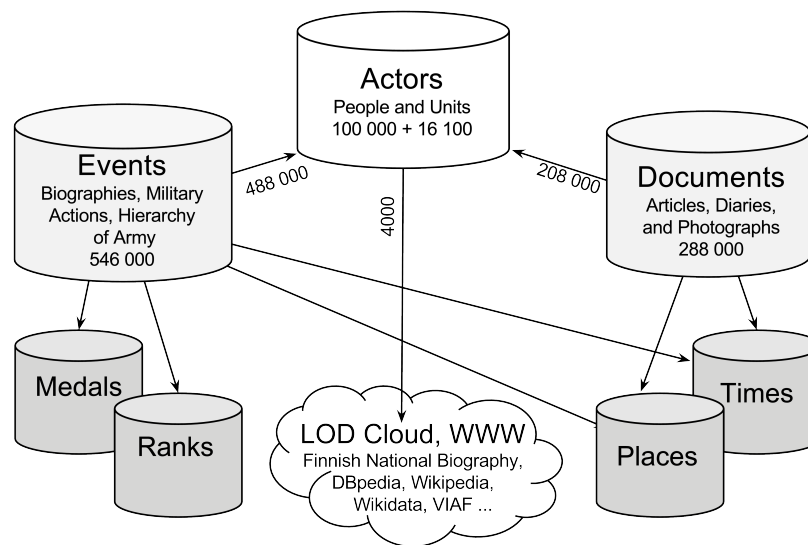


Fig. 1. Linkage in the actor-event based dataset

3 Actor Ontology Model

The ontology of actors is based on the CIDOC CRM¹² [4] model, where the resources of actors are essentially described in terms of the spatio-temporal events they participate

¹¹ <http://sotasampo.fi/en/>

¹² <http://cidoc-crm.org/>

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 to the database.

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:

Schema of the ontology is illustrated in Figure 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**¹³ 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**¹⁴.

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 resource¹⁵ (**: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

¹³ <http://www.cidoc-crm.org/cidoc-entities/e39-actor>

¹⁴ <http://www.cidoc-crm.org/cidoc-entities/e5-event>

¹⁵ http://ldf.fi/warsa/actors/person_294.ttl

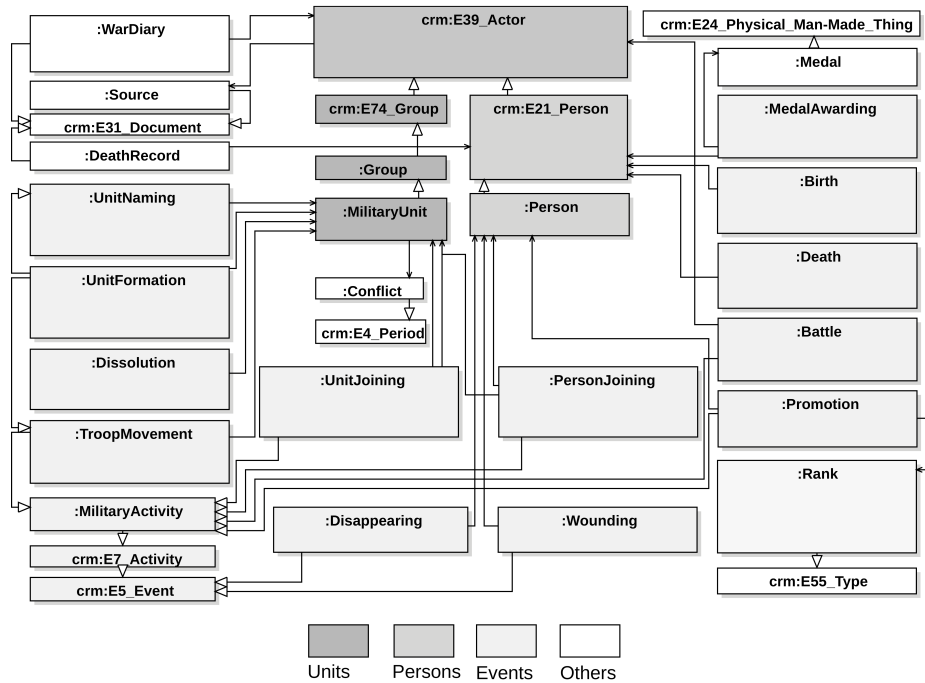


Fig. 2. Ontology schema of actors and events

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 often referred to 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.

An example of RDF resource of a military unit is shown in Table 4, the resource is also available in Turtle format¹⁶. Just like in the case of a person, the properties describe only constant information like unit's preferable name and abbreviation, descrip-

¹⁶ http://ldf.fi/warsa/actors/actor_459

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

tion, 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 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 database¹⁷, Finnish National Biography¹⁸, photographers mentioned in Finnish Wartime Photograph Archive¹⁹, Wikidata²⁰, 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 Air-warning, Finnish White Guard, and Foreign Volunteer Corps. At this stage Soviet and

¹⁷ kronos.narc.fi/menehtyneet/

¹⁸ <http://www.kansallisbiografia.fi/english/>

¹⁹ <http://sa-kuva.fi/neo?tem=webneoeng>

²⁰ <https://www.wikidata.org/>

Table 3. Examples of events describing pilot Jorma Karhunen

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

German troops were excluded. The main sources of information have been the War Diaries, Army Postal Code list²¹, 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 (Figure 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 Figure 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.

²¹ [http://www.arkisto.fi/uploads/Aineistot/kopsa\[1\].pdf](http://www.arkisto.fi/uploads/Aineistot/kopsa[1].pdf)

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

a) Jalkaväkirykmentti 7
JR 7. 3. Dsaan. **b)**

c) 1. Tampereella 14.10.39 JR 17

d) 2. Tam.sp., paitsi taistelujen aikana saatu täydennyshenkilöstö

e) 3. Rykmentin nimi muutettu 31.12.39 klo 22,00 alkaen JR 7:ksi (3.DEn kirj.№ 52/III/34/39 sal./28.12.39)

f) 4. Hyökkäys Summajoella

Summan puolustustaistelut	23.12.39	Eversti K.A.Heiskanen
Viivytystaistelut Summa-Säiniö (I/JR 7)	4.1 - 13.2.40	Eversti K.A.Heiskanen
Viipurin puolustustaistelut	13.2 - 17.2.40	Ratsumest.K.O.Alfthan
	20.2 - 13.3.40	Eversti K.A.Heiskanen.

Fig. 3. Information on an Organization Card

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

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,

Table 5. Examples of events describing 24th Fighter Squadron

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

DBpedia²², and VIAF²³. 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 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 API²⁴, and Wikipedia API for Python²⁵. Entity linking was done with ARPA service[13].

The datasets of conflicts, war diaries, medals, and ranks are in separate graphs. Conflicts²⁶ contain four main periods of WW2 in Finland. The War Diary graph²⁷ has

²² <http://wiki.dbpedia.org/>

²³ <https://viaf.org/>

²⁴ <https://en.wikipedia.org/w/api.php>

²⁵ <https://pypi.python.org/pypi/wikipedia>

²⁶ See, e.g., <http://ldf.fi/warsa/conflicts/LaplandWar>

²⁷ See, e.g., http://digi.narc.fi/digi/hae_ay.ka?sartun=319.SARK

26 400 entries. There are 200 medal types²⁸ and 200 rank entries²⁹. 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*). [10]

5 The WarSampo portal

The perspectives at WarSampo portal³⁰ visualize the linkage between the various datasets (e.g. military unit, personnel, casualties, events, places) etc [10,8]. 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 application³¹ 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 people³² 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.

WarSampo military unit perspective application³³ is illustrated in Figure 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).

²⁸ See, e.g., http://ldf.fi/warsa/medals/medal_83

²⁹ See, e.g., <http://ldf.fi/warsa/actors/ranks/Majuri>

³⁰ <http://www.sotasampo.fi/en/>

³¹ <http://www.sotasampo.fi/en/persons>

³² <http://yasgui.org/short/Blw2071gb>

³³ <http://sotasampo.fi/en/units>

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. **If you cannot find the person you are looking for, you can take a look at the unit's timeline.**

Search by person name

Aage/Aake Salmela
Aake Järmo
Aake Viljam Kataja
Aalto Heino
Aamos Benjamin Mällinen
Aapeli Kinnunen
Aapeli Ruotsalainen
Aapo Edward Rautilainen
Aapo Laitinen
Aapo Similä
Aappo Väinö Hyödynen
Aaprio Viikari
Aarne Aalton Valtameri Ahola
Aarne Abraham Mäkinen
Aarne Ahola
Aarne Aimo Ermi Nieminen
Aarne Akseli Hämäläinen
Aarne Albert Lehtinen
Aarne Aalton Lindholm
Aarne Akseli Kärgäläinen
Aarne Alexander Simes
Aarne Alfred Salo
Aarne Anselm Mikkoita
Aarne Anselm Salmisen
Aarne Anselmi Leppänen
Aarne Antero Johannes Salmisen
Aarne Antero Salmi
Aarne Antton Hironen
Aarne Armas Aarnio
Aarne Armas Jokinen
Aarne Armas Mäkinen

a)

b)

c) Paavo Juho Talvela
19.01.1897 Helsingin msk - 30.09.1973 Helsinki

d) Information

e) During the Winter War (1939 - 1940), Talvela commanded "Group Talvela" which took part in the Battle of Toivajärvi. For this success he was promoted to Major General in December 1939, the first promotion to general's rank during the war. In February 1940 Talvela took the command of the Finnish II Corps in the Karelian Isthmus. When the war ended on 13 March 1940, Talvela returned to civilian life. However, once the Finnish-German relations warmed, he was used in semi-official missions to Germany in late 1940. During the early Continuation War Talvela commanded the Finnish IV Corps in Karelia. From January 1942, when he was promoted to Lieutenant General, until February 1944 Talvela was the Finnish representative at the German High Command. Once back in Finland, Talvela commanded first the Finnish II Corps in northern Karelia until June 1944 when he took over the command of the Aunus Group. In July 1944 Talvela was sent back to Germany where he remained until Finland made peace with the Soviet Union in early September 1944. When he was about to depart for Finland, Himmler reportedly asked Talvela to become the head of a pro-German faction in Finland. Talvela refused out of hand. After the war Talvela spent some years in South America as a representative of Finnish paper industry, until returning to Finland. He was promoted to General of Infantry in retirement in 1966. Talvela was very able and aggressive commander in offense, but he was less well suited to defensive warfare. He was prone to envy and temper tantrums and his stubbornness made Talvela a very difficult subordinate. He performed best when given independent commands. Talvela was awarded the Mannerheim Cross in 1941. Source: Wikipedia, Suomalaisen Kirjallisuuden Seura, Biografiateksti, Suomi Solossa, ISBN 9519078940, Valtio Paat 1983, Mannerheim-ristin risti: http://www.mannerheim-ristinrivi.fi/rasti URL: http://www.finnsarsiaadrapaperi_50 Show information page

f) 193 related photos for this resource | Show all photos

g) Events (7) Ylipäällikkö muodostoi eventit P. Talvelan johtoon Toivajärven ja Iomanin suunnissa... Eventit Talvelan joukot aloittivat menestyksekkään vastahyökkäyksen. Aloitte siirtyi Toivajärven Äglajärven suunnassa... Kenraaliluutnantti Talvelan ja kenraalimajuri Böcklin jouduksia ennelttyykin ylipäällikkö määräs VI...

h) Units (2) Aunuksen ryhmä (Jatkosota) VI armeijajoukko (Jatkosota)

i) Battles (5)

j) Ranks (8)


k) Medals (17)

l) Persons (4377)

m) Places (2) Helsingin msk Helsinki

n) Wikipedia (1) Paavo Juho Talvela

o) Kansa Taisteli articles (3) Tulhoinen alla Varsijärven Venäläisten maahinnoisuus Tuulokissa 1944 2. osa Viitasaarelta maahinnoisuus

p) Talvela, Paavo Born on 1897 in Helsinki. Died on 1973 in Helsinki.  Kenraali Paavo Talvelaa on sanottu Suomen korkeaa arvoisimmaksi reservinupseeriksi. Tälle leikkimieliselle sanomalle on talletettu etä, että Talvela erosi näitä kertaa armeijan vakinaisesta palveluksesta joko osallisuuteen vapauttuneena heimosotain tai komissaan liike-elämän palveluksessa. Talvelalla oli kuitenkin keuhkoihin tehtävä talvi- ja jalkosoten aikaisena sotatörmityksen komentajana sekä ylipäällikkön edustajana Saksan sotavoinimen pääesikunnassa. Hän liittyi monin tavoin myös Suomen itsenäistymisen ajan historiaan: hän oli lupautuneeseen organisaattori, Alkoholiliikkeen johtokunnassa ja Pelhamon liike-elämän johtajana. Source: Semantinen Kansallistietografi 1923:n Kansallistietografi

SeCo Open Science and Research Programme Centenary of Finland's Independence Semantic Web Linked Data

Fig. 4. Information on Person Perspective

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

The screenshot displays the WarSampo website interface. On the left, there is a search bar and a list of military units. The main content area features a map of Finland with several red location markers. Below the map, a detailed view of a specific unit is shown, including its name, description, and a list of related resources. The resources include 'Super units', 'Subdivisions', 'Other units', 'Battles', 'Kansa Taisteli articles', 'Wikipedia', and 'War diaries'. The unit details also include a photo and a list of related resources.

Labels a) through p) are placed around the screenshot to indicate specific areas of interest:

- a) Search bar
- b) List of units
- c) Map of Finland
- d) Unit details (Osasto Hanell)
- e) Unit details (Osasto Hanell)
- f) Unit description
- g) Unit description
- h) Unit description
- i) Unit description
- j) Unit description
- k) Unit description
- l) Unit description
- m) Unit description
- n) Unit description
- o) Unit description
- p) Unit description

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 perspective³⁴ 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.

³⁴ <http://sotasampo.fi/en/articles>



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 reading 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 itself³⁵ 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 1914–1918³⁶, 1914–1918 Online³⁷, WW1 Discovery³⁸, Out of the Trenches³⁹, Muninn [19], and WW1LOD [15]. There are few works that use the Linked Data approach to World War II, such as [3,1], Defence of Britain⁴⁰, and Open Memory Project⁴¹. The main focus on our work 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, CultureSampo⁴², BookSampo⁴³, and Norssit—High School Alumni [11] datasets. Bio CRM model⁴⁴ 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

³⁵ <http://www.sotasampo.fi/en/photographs>

³⁶ <http://www.europeana-collections-1914-1918.eu>

³⁷ <http://www.1914-1918-online.net>

³⁸ <http://ww1.discovery.ac.uk>

³⁹ <http://www.canadiana.ca/en/pcdhn-lod/>

⁴⁰ http://thesaurus.historicengland.org.uk/thesaurus.asp?thes_no=365&thes_name=Defence%20of%20Britain%20Thesaurus

⁴¹ http://www.bygle.net/wp-content/uploads/2015/04/Open-Memory-Project_3-1.pdf

⁴² <http://seco.cs.aalto.fi/applications/kulttuurisampo/>

⁴³ <http://seco.cs.aalto.fi/applications/kirjasampo/>

⁴⁴ <http://seco.cs.aalto.fi/projects/biographies/>

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 percent of users come from Finland and 25 percent 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 spatio-temporal 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.

Acknowledgements Our work is funded by the Open Science and Research Initiative⁴⁵ of the Finnish Ministry of Education and Culture, the Finnish Cultural Foundation, and the Academy of Finland.

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: Proceedings of the 4th International Semantic Web Conference (ISWC 2005). pp. 127–141. Springer-Verlag (November 2005)
4. Doerr, M.: The CIDOC CRM – an ontological approach to semantic interoperability of meta-data. *AI Magazine* 24(3), 75–92 (2003)

⁴⁵ <http://openscience.fi/>

5. Guha, R.V., Brickley, D., Macbeth, S.: Schema.org: Evolution of structured data on the web. *Communications of the ACM* 59(2), 44–51 (2016)
6. Hachey, B., Radford, W., Nothman, J., Honnibal, M., Curran, J.R.: Evaluating entity linking with Wikipedia. *Artificial Intelligence* 194, 130–150 (Jan 2013), <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: *Language, Technology and Knowledge 2017*. June 19-20, Galway, Ireland. Springer-Verlag (2017), in press
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)*. pp. 39–52. *CEUR Workshop Proceedings* (May 2016), <http://ceur-ws.org/Vol-1608/#paper-06>, vol 1608
9. 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*. Forthcoming
10. 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)*. pp. 758–773. Springer-Verlag (2016)
11. 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: *Proceedings, Language, Technology and Knowledge 2017*. June 19-20, Galway, Ireland. Springer-Verlag (February 2017), <http://ldk2017.org/>, accepted
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: Sack, H., Rizzo, G., Steinmetz, N., Mladenić, D., Auer, S., Lange, C. (eds.) *The Semantic Web: ESWC 2016 Satellite Events*. Springer-Verlag (June 2016)
13. Mäkelä, E.: Combining a rest lexical analysis web service with sparql for mashup semantic annotation from text. In: *European Semantic Web Conference*. pp. 424–428. Springer (2014)
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. *International Journal on Digital Libraries* (2016), in press.
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: *Extended Semantic Web Conference*. pp. 593–607. Springer (2013)
17. Tamper, M., Leskinen, P., Ikkala, E., Oksanen, A., Mäkelä, E., Heino, E., Tuominen, J., Koho, M., Hyvönen, E.: AATOS – a Configurable Tool for Automatic Annotation. In: *Proceedings, Language, Technology and Knowledge 2017*. June 19-20, Galway, Ireland. Springer-Verlag (February 2017), accepted
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*. Berlin, DE: *Proceedings of Wikipedia Academy: Research and Free Knowledge (WPAC2012)*. URL: <http://hangingtogether.org> (2012)