

Aalto-yliopisto
Perustieteiden korkeakoulu
Tietotekniikan koulutusohjelma

Petri Leskinen

Sotilashenkilöiden ja joukko-osastojen mallin- taminen ja käyttö toimijaontologiana

Diplomityö
Espoo, 28. marraskuuta 2016

Valvoja: Professori Eero Hyvönen, Aalto-yliopisto
Ohjaajat: TkT Eetu Mäkelä
FM Jouni Tuominen

Tekijä:	Petri Leskinen		
Työn nimi:	Sotilashenkilöiden ja joukko-osastojen mallintaminen ja käyttö toimijaontologiaa		
Päiväys:	28. marraskuuta 2016	Sivumäärä:	6 + 61
Pääaine:	Mediatekniikka	Koodi:	T-75
Valvoja:	Professori Eero Hyvönen		
Ohjaajat:	TkT Eetu Mäkelä FM Jouni Tuominen		
<p>Toimijaontologia mallintaa henkilöitä ja henkilöryhmiä linkitetystä avoimesta datasta. Toimijaontologiamallin tarkoitus on mahdollistaa eri lähteiden aineistojen kokoaminen yhteen ja sen julkaisu yhdenmukaisessa formaatissa, jotta tietoa voidaan hyödyntää niin digitaalisten ihmistieteiden tutkimuksessa kuin tarjoamalla käyttöliittymiä aineiston selaamiseen visuaalisessa muodossa.</p> <p>Laadittu ontologia noudattaa toimija-tapahtuma-mallia. Siinä toimija mallinnetaan häneen liittyvien elämäkerrallisten tapahtumien summana. Ratkaisujen perustana käytettiin CIDOC CRM -standardia, millä haluttiin taata mallin helppo laajennettavuus sekä noudattaa kulttuurihistorialliselle datalle yhdenmukaista julkaisukäytäntöä.</p> <p>Työ on tehty osana laajempaa Sotasampo-projektia, johon kerättiin kattava tietokanta toisen maailmansodan aikaista aineistoa Suomen osalta. Oma osuuteni tässä työssä oli toimijaontologiamallin laatiminen sekä sen populointi sotilashenkilöillä ja -osastoilla. Aineisto on julkaistu avoimena datana (http://www.ldf.fi/dataset/warsa) ja on selattavissa Sotasampo-portaalissa (http://www.sotasampo.fi).</p>			
Asiasanat:	toimijaontologia, semanttinen web, linkitetty data, RDF, SPARQL		
Kieli:	suomi		

Author:	Petri Leskinen		
Title:	Modelling and Using Military Persons and Units as an Actor Ontology		
Date:	November 28, 2016	Pages:	6 + 61
Major:	Professorship	Code:	T-75
Supervisor:	Professor Eero Hyvönen		
Advisors:	Eetu Mäkelä D.Sc. Jouni Tuominen M.Sc.		
<p>An actor ontology is used for modeling persons and groups in linked open data. The semantic ontology model enables combining material from various sources and publishing it in a harmonized format so that the data can be used in digital humanities research and as well for developing online portals.</p> <p>The actor ontology model represented in this work applies an actor-event model approach: the lifestory of an actor is modeled as a sum of biographical events. The solution is based on the CIDOC CRM standard providing an extensible semantic framework and an information standard for publishing culture historical data.</p> <p>This work has been done as a part of WarSampo project, where large collections of Finnish Second World War material were harmonized and provided as Linked Open Data. My part in the project was designing an actor ontology model and populating it with military persons and units collected from various sources. The WarSampo datasets are published as a Linked Open Data service (http://www.ldf.fi/dataset/warsa) and demonstrated in WarSampo portal (http://www.sotasampo.fi/en/).</p>			
Keywords:	actor ontology, semantic web, linked data, RDF, SPARQL		
Language:	Finnish		

Tekijän kiitokset

Haluan kiittää professori Eero Hyvöstä ja ohjaajiani Eetu Mäkelää ja Jouni Tuomista diplomityöni ohjauksesta.

Kiitos kaikille Semanttisen laskennan tutkimusryhmässä työskenneille, etenkin niille, joiden kanssa Sotasampo-projektia rakennettiin.

Kiitokset myös kaikille Sotasampoon aineistoa toimittaneille kuten Maanpuolustuskorkeakoulun professori ja historiantutkija Ohto Manniselle, Kansallisarkiston Tomi Ahorannalle ja tuotantoyhtiö Kinocompanylle.

Espoo, 28. marraskuuta 2016

Petri Leskinen

Käytetyt lyhenteet

BIO	A Vocabulary for Biographical Information
CIDOC CRM	International Committee for Documentation – Conceptual Reference Model
CSV	Comma Separated Values
DC	Dublin Core
FOAF	Friend of a friend
JSON	JavaScript Object Notation
LDF	Linked Data Finland
LOD	Linked Open Data
MPKK	Maanpuolustuskorkeakoulu
MRR	Mannerheim-ristin ritari
OCR	Optical Character Recognition
OWL	Web Ontology Language
PDF	Portable Document Format
PNG	Portable Network Graphics
RDF	Resource Description Framework
RDFS	Resource Description Framework Schema
SeCo	Semanttisen laskennan tutkimusryhmä
SKB	Semanttinen Kansallisbiografia
SKOS	Simple Knowledge Organization System
SPARQL	SPARQL Protocol and RDF Query Language
Turtle	Terse RDF Triple Language
ULAN	Union List of Artist Names
URI	Uniform Resource Identifier
URL	Universal Resource Locator
VIAF	Virtual International Authority File

Sisältö

1	Johdanto	1
1.1	Henkilöt ja ryhmät ontologiana	1
1.2	Tutkimuskysymykset	2
2	Taustaa	4
2.1	Semanttinen web ja linkitetty data	4
2.2	Toteutettuja projekteja	7
3	Toimijaontologia	11
3.1	Toimijaontologian lähtökohtia	11
3.2	Laadittu toimijaontologiamalli	17
4	Sotasammon toimijat	25
4.1	Sotasampo-projekti	25
4.2	Toimijoiden taustaa	26
4.3	Toimijatietokanta	27
5	Sovellukset	43
5.1	Warsa-datajulkaisu	43
5.2	Sotasampo-portaali	44
6	Yhteenveto	51
6.1	Tulosten arviointi	51
6.2	Kehitettävää	52

6.3 Kontribuutio	54
Lähteet	55
Liitteet	1
Liite 1: Toimijaontologian skeema	1
Liite 2: Henkilötoimija	7
Liite 3: Joukko-osasto	9
Liite 4: Sotilarvo	12
Liite 5: Sotapäiväkirja	13
Liite 6: Kunniamerkki	14
Liite 7: SPARQL-esimerkkihaku	15

Luku 1

Johdanto

Toimijaontologia kuvaa ihmisryhmää, sen jäseniä yksilöinä tai ryhminä. Se tutkii keskinäisten suhteiden, vuorovaikutusten ja toiminnan tapahtumaketjua ja rakentaa niistä elämänkaaria [38]. Toimijaontologia toimii referenssitietokantana, joka mahdollistaa tietoaineistojen yhteentoimivuuden. Historiaa kuvaavan toimijaontologian käyttötarkoitus voi olla digitaalinen ihmistutkimus, joka tietoteknisillä menetelmillä analysoi dataa ja pyrkii päättämään uutta tietoa. Aineistoa voidaan havainnollistaa web-portaalissa, jolloin tieto on tarjolla kaikille historiallisesta tiedosta kiinnostuneille.

1.1 Henkilöt ja ryhmät ontologiana

Linkitettyssä datassa todellisuuden osia mallinnetaan ontologioina. Ne kehitetään analysoimalla aihepiirin entiteettejä ja jaotteleamalla ne käsitteisiin ja niiden välisiin suhteisiin [23]. Toimijaontologian pienin yksikkö on yksittäinen henkilö. Hänellä on perustietoja kuten nimi, syntymäaika ja -paikka. Häneen liitetään yleisiä ominaisuuksia kuten esimerkiksi sukupuoli, ammatti ja kansallisuus, usein myös dokumentoivaa materiaalia kuten valokuvia tai kirjallisia artikkeleja. Verkosto henkilöiden välille voi rakentua esimerkiksi sukulais-, työpaikka- ja ystävyysuhteista.

Toimijaontologian toinen peruselementti on ryhmä, joka voi koostua henkilöistä tai pienemmistä alaryhmistä. Ryhmän perustietoja voivat olla sen nimi tai muu nimeävä tunniste, olemassaolon aikatieto tai maantieteellinen vaikutusalue. Henkilöiden ja ryhmien välisiä interaktioita ovat esimerkiksi ryhmään liittyminen ja siitä eroaminen. Ryhmien kuulumisesta toistensa alaisuuteen saattaa muodostua organisaatiorakennetta kuvaava hierar-

kiapuu. Ryhmää määrittää usein jokin jäseniä yhdistävä tekijä: aate, toiminta tai tavoite. Toiminta voi tuottaa abstraktin tai konkreettisen tuloksen; toiminnalla voi myös olla vaikutuksia toisiin toimijoihin. Henkilöillä on ryhmässä erilaisia rooleja, jotka saattavat vaihtua ajan funktiona ja joilla voi olla keskinäinen arvojärjestys. Rooli voi liittää henkilöön ammatin tai arvonimen, joka appositioattribuuttina [50] antaa henkilölle uuden tunnusteen.

Työssä kehitetty ontologia rakennettiin osana Sotasampo-projektia [25], jossa toimijoina ovat Suomen armeijan sotilashenkilöt ja joukko-osastot vuosina 1939–1945. Tämä aihepiiri pitkälti sanelee mitä piirteitä toimijoista halutaan kuvata. Henkilön mallintamisessa biografisten perustietojen lisäksi tärkeää on hänen sotilasarvonsa, rooli joukko-osastoissa, osallistuminen sotatoimiin sekä mahdolliset kunniamitalit. Tietoa osastosta ovat sen perustamisaika ja -paikka, siitä käytetyt nimet, sen asema armeijan hierarkiassa sekä sen keskeytykset ja taistelut.

1.2 Tutkimuskysymykset

Toimijaontologia on tietokokoelma, jossa henkilö tai henkilöryhmä esitetään biografisten tapahtumien summana. Mallinnuksesta tulisi pystyä rekonstruoimaan toimijan status tietyllä ajanhetkellä. Tässä työssä pyritään vastaamaan tutkimuskysymyksiin, joita ovat:

1. Miten semanttisessa webissä kuvataan toimijaa eli yksittäistä henkilöä?
2. Miten kuvataan henkilöryhmää ja ryhmien välistä hierarkiaa?
3. Millainen malli soveltuu mahdollisimman joustavasti eri lähteiden tiedon yhdistämiseen, tietokenttien harmonisointiin ja linkittämiseen?
4. Miten malli palvelee loppukäyttötarkoitusta kuten visualisointia verkkosivustolla tai biografista esittämistä historian tutkimusta varten?

Diplomityön sisältö

Tämän diplomityön luvussa 2 esitellään semanttisen webin periaatteita ja tekniikoita sekä esimerkkejä toimijaontologiamalleista ja niihin liittyvistä projekteista tulevan työn perustaksi. Luvussa 3 perehdytään tarkemmin laadittuun ontologiamalliin. Luvussa 4 esitellään Sotasampo-projekti ja sen hen-

kilö- ja joukko-osastoja mallinnuksen kannalta. Luvussa 5 esitellään Sotasammon datajulkaisu LDF.fi¹-palvelussa ja Sotasampo-portaalin² aineistonäkymiä. Luvussa 6 pohditaan tehdyn työn ratkaisuja ja esitetään suuntaa jatkokehitykselle. Liitteinä on toimijaontologian skeema ja teknisiä esimerkkejä mallin määrittelyistä ja tietokantakyselyistä.

¹<http://www.ldf.fi/dataset/warsa>

²<http://www.sotasampo.fi>

Luku 2

Taustaa

Tämän luvun ensimmäisessä aliluvussa esitellään semanttisen webin periaatteita, tekniikoita ja standardeja. Toisessa aliluvussa käsitellään toimijaontologian aihepiiriin liittyviä projekteja.

2.1 Semanttinen web ja linkitetty data

Semanttisen webin periaatteena on tiedon julkaisu linkitetyn datan muodossa. Linkitetyn datan pienin alkio on resurssi, joka on kuvaus olemassa olevassa entiteetistä tai abstraktista käsitteestä. Globaalissa tietoverkossa jokaisella resurssilla on yksilöivä verkko-osoite koneluettavine kuvauksineen. Resurssien välille rakentuu verkko, jossa niitä yhdistävillä kaarilla on semanttisesti määritelty sisältö, yhteyden luonteen kuvaus. Erillisten tietokantojen resursseja voidaan edelleen yhdistää uusiin tietokantoihin, joka rikastaa aineistoa. Standardimuotoisten ja koneluettavien aineistojen yhdistäminen voidaan automatisoida, samoin kuin uuden tiedon päättely tätä rakenteellista kokonaisuutta analysoimalla. Visio on luoda pienistä yksittäisistä tietokannoista maailmankattava tietokokonaisuus, semanttinen web. [5]

Tim Berners-Leen esittämiä periaatteita linkitetyn datan julkaisemiseksi [5] ovat:

1. Nimeä resurssit URI-tunnisteina,
2. jotka ovat HTTP-osoitteita,
3. joista voidaan lukea standardimuotoista tietoa,
4. joka ohjaa uusien resurssien URI-tunnisteisiin.

2.1.1 Standardit RDF, RDFS, OWL ja SPARQL

Tässä aliluvussa käsitellään linkitetyn datan mallintamisen ja julkaisemisen keskeisiä standardeja, joita ovat tietomalli RDF, skeema RDFS, ontologiakieli OWL ja kyselykieli SPARQL. Resource Description Framework (RDF) on perusta semanttisen tiedon esittämiseen. RDF:n perusrakenne on subjekti-predikaatti-objekti-kolmikko (triple), jolla määritellään kahta solmua (subjekti, objekti) yhdistävä kaari (predikaatti) RDF-graafissa. Solmut voivat olla URI-tunnisteisia resursseja, literaaleja arvoja tai tyhjiä solmuja (blank node). Literaaleilla arvoilla on niiden tietosisältö, tietotyyppi ja mahdollinen kielitunniste (language tag). Ne voivat olla esimerkiksi `xsd:date`-tyyppisiä päivämääriä tai kielitunnisteisia merkkijonoja. [12]

Resource Description Framework Schema (RDFS) tarjoaa työkaluja sanastojen määrittelyyn. Siinä keskeisiä komponentteja ovat luokka (`rdfs:Class`) ja ominaisuus (`rdf:Property`). RDFS:llä määritellään kolmikkojen subjektit ja objektit haluttujen luokkien instansseiksi (`rdf:type`). Luokille voidaan rakentaa hierarkia määrittelemällä niitä toistensa aliluokiksi (`rdfs:subClassOf`). Ominaisuuksilla tarkennetaan kolmikkojen predikaattien ilmaisemista. Ominaisuuksien hierarkia lisätään `rdfs:subPropertyOf`-määrittelyillä. Ominaisuudet voivat saada määrittely- ja arvojoukkoja; ts. minkä luokan instanssit voivat käyttää ominaisuuksia (`rdfs:domain`) ja minkä luokan arvoja ne voivat saada (`rdfs:range`). [8]

Ontologiakieli OWL on RDFS:n laajennus entistä ilmaisuvoimaisempien ontologioiden esittämiseen. Sillä voidaan asettaa sääntöjä ja rajoituksia tietokannan validoinnille ja konepäätelylle. OWL:n sanastolla voidaan merkitä kaksi resurssia samaa kuvaaviksi (`owl:sameAs`) tai eriyttää toisistaan (`owl:differentFrom`, `owl:AllDifferent`). OWL:lla voidaan rajata ominaisuuksien kardinaliteettiä, ts. määritellä kuinka monta eri arvoa yksittäinen ominaisuus voi saada. Ominaisuuksia voidaan tarkentaa määrittelemällä niitä toisilleen vastakohtaisiksi, symmetrisiksi tai transitiivisiksi. [45]

SPARQL on standardoitu kyselykieli, jota yleisimmät RDF-tietovarastot tukevat. Syntaksiltaan SQL:ää muistuttava kysely kohdistetaan SPARQL-palvelupisteeseen (endpoint). Kyselyillä voidaan hakea kuvauksia (describe) ja taulukoituja vastauksia (select), testata tiedon olemassaolo (ask) tai rakentaa kyselyn perusteella uusi RDF-graafi (construct). Kyselyt voivat yhdistää useista palvelupisteistä saatuja tuloksia (federated query). SPARQL Update Language -laajennus (SPARUL) sallii tietokannan muokkaamisen (delete, insert). [57]

2.1.2 Sanastot FOAF, REL ja BIO

Sanastot FOAF, REL ja BIO tarjoavat yhdistettyinä tavan rakentaa biografinen RDF-kuvaus henkilöstä. Friend of a Friend -projekti (FOAF) pyrkii tarjoamaan käyttäjille mahdollisuuden julkaista ja hallita omia henkilötietojaan verkossa. Näin julkaistuna tieto on myös koneluettavassa muodossa. Käyttäjä voi esittää omia perustietojaan kuten etu- ja sukunimensä, sukupuolensa, kiinnostuksen kohteitaan, asuinpaikan, kotisivunsa tai omakuvansa. FOAF:in kehittäjä Ian Davisin verkkosivulla [13] on esimerkki sanaston käytöstä.

FOAF:in ontologia on yksinkertainen ja sen toimijoiden yläluokka on **Agent**, josta on periytetty keskeisimmät luokat toimijoille kuten **Person**, **Group** tai **Organization**. Ryhmillä **Group** ja **Organization** ei ole teknistä eroa, **Organizationilla** vain tarkoitetaan pysyvämpää, vakiintuneempaa ja pitkäaikaisempaa toimijajoukkoa. Yksittäisen toimijan kuuluminen laajempaan ryhmään määritellään ominaisuudella `foaf:member` ja henkilöiden välisiä suhteita ominaisuudella `foaf:knows`. [9]

Relationship-sanasto (REL) on kehitetty kuvaamaan FOAF:in instanssien välisiä suhteita. Sanasto laajentaa `foaf:knows`-omaisuutta ja lisää sille 36 aliominaisuutta, joilla kuvataan perhe-, ystävyys-, työpaikka- tai aviosuhteita. REL määrittelee vain henkilöiden välisiä suhteita, ei esimerkiksi heidän roolejaan laajemmissa ryhmissä eikä ryhmien hierarkkisia suhteita. Sanasto sisältää myös yleisen luokan `rel:Relationship` käyttäjän omia määritelmiä varten. Lukuun ottamatta yksittäisiä ominaisuuksia kuten `rel:hasMet` tai `rel:lostContactWith` sanasto keskittyy määrittelemään nykytilannetta eikä pyri kuvaamaan ajallisia muutoksia. [15]

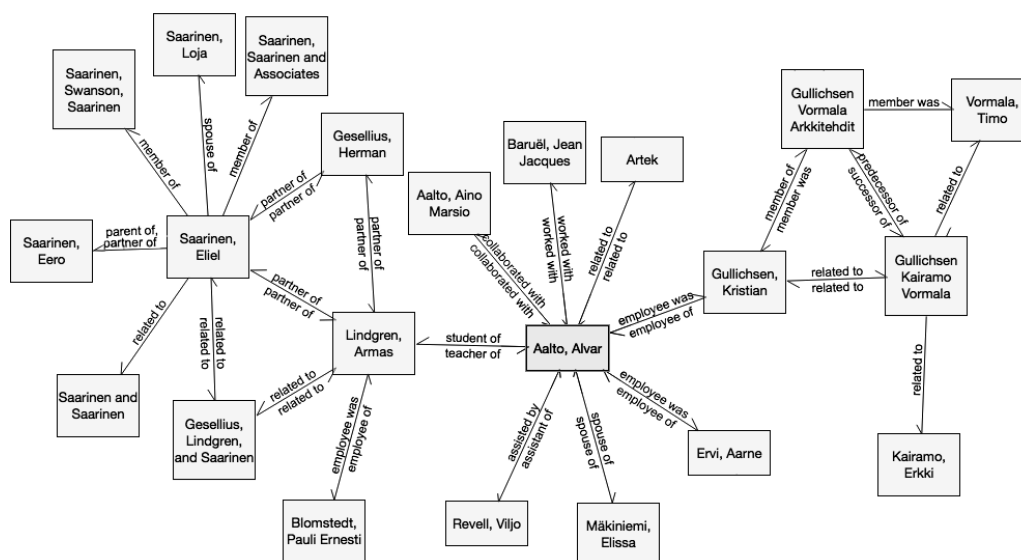
Biographical Vocabulary (BIO) on tapahtumaontologia, joka mahdollistaa biografisten tietojen lisäämiseen FOAF-määrittelyyn. BIO:n tapahtumilla esitetään elämäkerrallisia tietoja kuten synnyin- ja kuolinajat ja -paikat, tapahtumia kuten kaste tai avioliiton solmiminen sekä koulutus- ja työhistoriaa. Tietokannassa tällainen tapahtuma on itsenäinen resurssi, joka sisältää viittauksen henkilöön. Tapahtuman attribuutteina annetaan tapahtumapaikka ja -aika sekä mahdollista lisätietoa kuten toimenkuva työsuhteessa. Tapahtumilla voi lisäksi olla keskinäisiä suhteita, kuten niiden samanaikaisuus tai kausaalinen suhde. Esimerkkinä jälkimmäisestä `bio:Burial` on tapahtuma, joka on seuraus tapahtumasta `bio:Death`. [14]

2.2 Toteutettuja projekteja

Seuraaviin alilukuihin on poimittu tapausesimerkkejä aihepiiriin liittyvistä projekteista. Projektien kohdalla tarkastellaan tapoja mallintaa henkilöitä ja ryhmiä. Osa projekteista on ollut mallina tässä työssä toteutetulle ontologialle.

2.2.1 Getty-säätiön ULAN-rekisteri

Union List of Artist Names (ULAN)¹ on yksi Getty-säätiön² ylläpitämistä rekistereistä ja siihen on koottu taidehistorian toimijoita. Tietokannassa on tunnettuja taiteilijoita ja taideryhmiä, pääpaino on kuvataitelijoilla ja arkkitehteillä. Ryhmiä voivat olla taiteilijaryhmät, taidekoulut ja -museot tai niiden osastot. Sanastossa on 120 000 tietuetta ja niissä 293 000 taiteilijanimeä. Tietokanta ei sisällä taideteoksia.



Kuva 2.1: Alvar Aaltoon liittyvää verkostoa ULAN:issa.

ULAN-rekisterin malli kuvaa yksityiskohtaisesti toimijan roolit, elinvuodet, kansallisuuden tai maantieteellisen sijainnin, nimien eri kirjoitusasuja sekä keskeisiä sosiaalisia kontakteja. Kullekin taiteilijalle on koottu nimen erikielisiä kirjoitusasuja ja mahdollisesti käytettyjä pseudonyymejä. Esimerkiksi

¹<http://www.getty.edu/research/tools/vocabularies/ulan/>

²<http://www.getty.edu/foundation/>

saksalaiselle Albrecht Dürerille on lähes 40 erilaista kirjoitusasua kuten *Albertus Durerus*, *Albrekht Diurer*, *Albrecht Dürer*, *Alverto Dureño* tai *Alberduer*³.

Toimijoilla on kaksi luokkaa `gvp:PersonConcept` ja `gvp:GroupConcept`. Toimijoiden välillä voi olla 124 erilaista suhdetta⁴. Niillä voidaan kuvata sukulaisuutta, yhteistyötä toimijoiden välillä, opettaja–oppilas-suhdetta, taiteilija–mesenaatti-suhdetta tai henkilön osuutta ryhmän toimintaan. [56]

Kuvassa 2.1 havainnollistetaan arkkitehti Alvar Aaltoon kolmen linkin etäisyydellä liittyviä toimijoita. SPARQL-kyselyn⁵ tuloksena saadaan verkko tunnetuista 1900-luvun suomalaisista arkkitehteistä ja -toimistoista. Verkon solmussa on toimijan nimi ja niitä yhdistävissä kaarissa kuvaus suhteesta.

2.2.2 Muninn Military Ontology -projekti

Linkitettyä avointa dataa hyödynnetään digitaalisessa ihmistutkimuksessa [54], sotahistoriallinen tieto yhtenä osa-alueena. Ensimmäinen maailmansodan osalta on julkaistu hankkeita kuten *Europeana Collections 1914–1918* [28], *1914–1918 Online* [19], *WW1 Discovery* [52], *Out of the Trenches* [10], *CENDARRI* [20] ja *WW1LOD* [48]. Toinen maailmansota tarjoaa vielä laajemman tutkimuskentän; siitä tähän mennessä toteutettuja hankkeita ovat italialainen *Open Memory Project* sekä Suomen osalta *Sotasampo*-hanke.

Muninn Military Ontology [59] on koonnut yhteen sotilasalan terminologiaa. Taustalla on Robert Warrenin projekti, jossa on tutkittu ontologioiden generointia *Wikipedian* ja *DBpedian* [16] tietokannoista [58]. Ontologiaan on koottu eri maiden armeijoissa käytössä olleita aselajeja, joukko-osastotyyppisiä, sotilasarvoja ja muuta tietoa. Mukana on niin historiallista tietoa antiikin Kreikan armeijoista kuin nykyisiä *Nato*-standardeja.

Sotilasarvoresurssia on tarkennettu sen sotilasarvoluokalla⁶ sekä puolustushaaralla ja maalla tai armeijalla, jossa se on ollut käytössä. Sotilasarvoja on linkitetty arvojärjestyksessä ylempiin, alempiin ja samanarvoisiin sotilasarvoihin, mistä rakentuu hierarkkinen kokonaisuus. Sotilasyksiköt käsittävät armeijoiden puolustushaaroja, aselajeja ja joukko-osastoja. Kokonaisuuden hierarkia on rakennettu linkityksellä osastotyyppien välillä, esimerkiksi `:Regiment :hasPart :Battalion`. *Muninnin* mallia käytettiin *Sotasampo*-projektin sotilasarvo-ontologian perustana.

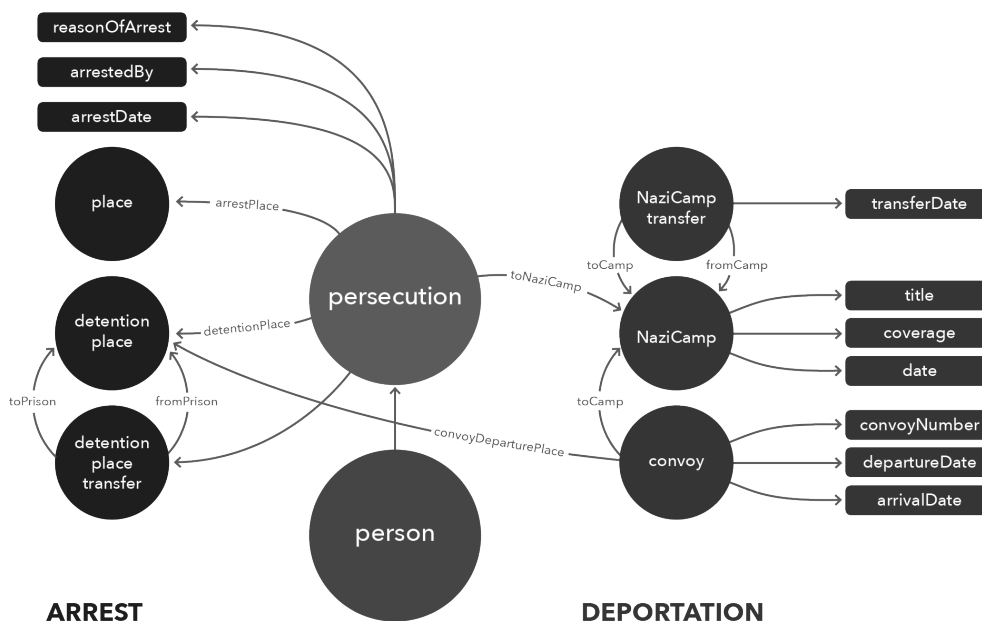
³<http://www.getty.edu/vow/ULANFullDisplay?find=&role=&nation=&subjectid=500115493>

⁴SPARQL-kysely ULAN:in suhteista: <http://yasgui.org/short/H1cXEoxzg>

⁵<http://yasgui.org/short/rk2UFFbze>

⁶Esimerkiksi upseerit tai miehistö

2.2.3 Open Memory Project -tietokanta



Kuva 2.2: Open Memory Project -tietomalli [11].

Open Memory Project⁷ on CDEC:n⁸ projekti, johon on koottu tietoa 19 000 holokaustin uhrista toisen maailmansodan aikana Italiassa. Tietokantaan voi tutustua esimerkkihenkilö Levi Primon⁹ sivulla. Tietoa on rikastettu linkityksellä vastaavaan henkilöresurssiin muissa linkitetyn datan tietokannossa, kuten Wikidatassa [60], DBpediassa ja VIAF:issa [4].

Havainnollistus tietomallista on esitetty kuvassa 2.2. Henkilön mallinnus on eroteltu biografiatietoihin (**person**) ja sotavainotietoihin (**persecution**), eli yhtä henkilöä kohti on kaksi resurssia, eikä vaihtelevaa määrää tapahtumia kuten BIO:ssa. Henkilön biografiaesitys perustuu FOAF-malliin. Henkilö on **foaf:Person**-luokan instanssi, johon on liitetty perustiedot, kuten nimet ja sukupuoli. Biografiamallia on täydennetty 30 ominaisuudella, joilla voidaan tarkentaa monia tietoja kuten vaikka sisaruussuhteet, koulutus, asuinpaikka, äidin tyttönimi tai isovanhempien nimet.

Henkilöresurssi on linkitetty sitä vastaavaan sotavainoresurssiin, joka kuvaa tapahtumia kuten pidättämistä, vankeutta tai vankikuljetusta. Tietomalliin kuuluu myös täydentäviä sanastoja tapahtumapaikoista, vankileireistä

⁷<http://digital-library.cdec.it/cdec-web/>

⁸<http://www.cdec.it/>

⁹<http://digital-library.cdec.it/cdec-web/persone/detail/person-5002/levi-primo.html>

ja kuljetuksista. [7]

2.2.4 Semanttinen kansallisbiografia

Semanttinen kansallisbiografia (SKB) on Aalto-yliopiston Semanttisen laskennan tutkimusryhmässä toteutettu toimija-tapahtuma-ontologia. Sen aiheistona on käytetty Suomalaisen Kirjallisuuden Seuran Biografiakeskuksen biografioita [6]. Aiemmin toteutettu Suomen historian ontologia (HISTO) kattaa 1 173 merkittävää tapahtumaa ja 80 085 muuta historiallista mainintaa. SKB:ta varten tähän on linkitetty 7 302 henkilöä ja pari sataa toimijaryhmää. Toimijaryhmiä ovat esimerkiksi jääkärit ja eduskunta¹⁰. [24]

SKB on julkaistu avoimena datana¹¹, ja sen pohjalta on toteutettu Elämäkerturi-käyttöliittymä¹². Monet SKB:n resursseista kuuluvat Sotasampo-projektin aihepiiriin, kuten Karl Lennart Oesch^{13 14} ja hänen ylentäminen kenraaliluutnantiksi¹⁵.

¹⁰<http://ldf.fi/history/histo/p1983>

¹¹<http://www.ldf.fi/dataset/history/>

¹²<http://www.ldf.fi/dataset/history/map.html>

¹³<http://ldf.fi/history/kb/p161>

¹⁴<http://ldf.fi/history/histo/p4092>

¹⁵<http://ldf.fi/history/kb/p61099>

Luku 3

Toimijaontologia

Tämän luvun ensimmäisessä aliluvussa perehdytään toimijaontologian taustalla käytettävään CIDOC CRM -standardiin ja käsitellään taustatekijöitä laaditulle toimijaontologialle. Toisessa aliluvussa esitellään laadittu toimija-tapahtuma-malli.

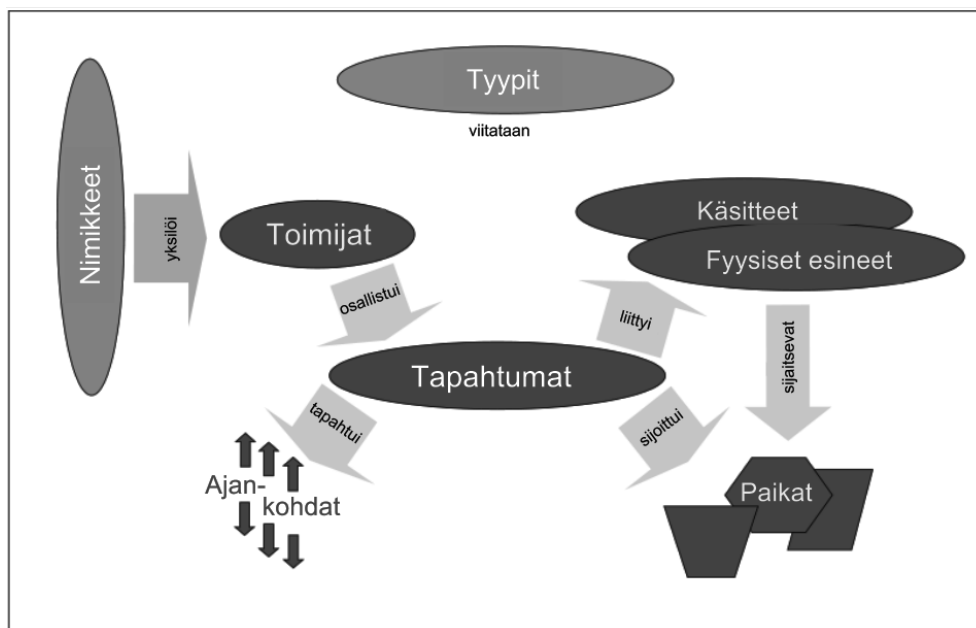
3.1 Toimijaontologian lähtökohtia

Tässä aliluvussa esitellään CIDOC CRM -standardi ja pohditaan miten sen periaatteita sovelletaan sotahistoriallisen tiedon mallintamiseen. Lisäksi eritellään ominaisuuksia, joita henkilöistä ja ryhmistä halutaan mallintaa sekä millaisia tapahtumia näihin tulisi liittää. Lopuksi käydään läpi millaista problematiikkaa aihepiiri asettaa esimerkiksi toimijoiden yksilöllisen tunnistamisen kannalta.

3.1.1 CIDOC CRM

CIDOC eli International Committee for Documentation (CIDOC) on yksi International Council of Museums -organisaation (ICOM) komiteoista. Se pyrkii museoiden dokumentaatiotapojen kehittämiseen ja parhaiden käytäntöjen löytämiseen. CIDOC CRM (CIDOC Conceptual Reference Model) on johdettava standardi kulttuuriperinnön tallennuksessa ja luokittelussa. Standardin päämäärinä on helpottaa sisällön tunnistamista eri lähteistä, varmistaa tiedon säilyminen muunnoksissa sekä yksinkertaistaa tiedon hakua yhdistetyistä aineistoista. [18]

CIDOC CRM:n (kuva 3.1) keskeisiä perusluokkia ovat [44]:



Kuva 3.1: CIDOC CRM:n keskeisiä perusluokkia. Suomennettu lähteestä [18].

1. **Toimija** (Actor) on henkilö tai ryhmä.
2. **Tapahtuma** (Event) kiinnittää toimijoita, ajankohtia, paikkoja ja esineitä toisiinsa. Tietomallintamisessa tapahtuman käsite on laajempi kuin arkipuheessa. Sillä tarkoitetaan kaikkia asiantilojen muutoksia, jotka jakavat aikajatkumon tapahtumaa edeltäneeseen ja sitä seuraavaan vaiheeseen.
3. **Ajankohta** (Timespan) tarkoittaa tapahtuman ajankohtaa. Ajankohdan alulle, lopulle ja kestolle sallitaan epätarkkuutta. Ajankohtien välinen sisäkkäisyys, päällekkäisyys tai erillisuus voivat muodostaa hierarkian. Tapahtuman ajankohtaan voidaan liittää tarkasti tunnettu ajankohde. Näistä päivämäärä- ja kellonaikatiedoista käytetään nimitystä aikaprimitiivi.
4. **Paikka** (Place, Site) kertoo tapahtuman sijainnin. Sijainti voi viitata suoraan todelliseen, kartalle nimettyyn alueeseen tai merkittyyn koordinaattiin. Sijainti voidaan myös ilmaista suhteessa muihin entiteetteihin.
5. **Käsitteet ja esineet** (Concept, Thing) ovat abstrakteja käsitteitä tai fyysisiä esineitä, joiden kanssa toimijat ovat vuorovaikutuksessa.

6. **Nimike** (Appellation) on nimi, koodi tai muu tunniste toimijoille, tapahtumille, paikoille, ajankohdille ja käsitteille.
7. **Tyyppi** (Type) on luonteeltaan adjektiivinen tai kategoria ja se kuvaa esimerkiksi objektinsa kokoa, arvoa, sukupuolta tai kansallisuutta.

CIDOC CRM -mallia voidaan kritisoida sen kompleksisuudesta ja siitä, miten yksinkertaisenkin asian ilmaiseminen tuottaa runsaasti RDF-kolmikkoja. Esimerkiksi henkilön nimi- ja syntymätiedon *Petri Leskinen, syntynyt 26. tammiukuuta Helsingissä* esittämiseksi joudutaan generoimaan henkilöresurssi, synnyintapahtuma, tapahtuman aika- ja paikkaresurssi sekä nimiresurssit henkilölle, ajalle ja paikalle. Mallinnus tuottaa yhteensä seitsemän resurssia sekä niitä sitomaan vähintään kuusi linkkiä. Yksinkertaisimmalla mallilla kolmikkoja olisi vain neljä, yksi resurssin luokalle (`rdf:type`) ja tietokentät nimelle, syntymäajalle ja -paikalle.

Biografisessa lähtöaineistossa voi olla tietoja kuten asuinpaikka, sukulaisuhteita tai lasten lukumäärä, joiden mallintaminen CIDOC CRM:lla voi olla hankalaa. Tiettyyn rajaan saakka sukulaisuhteita voidaan päätellä synnyintapahtumien lapsi-äiti-isä-kentistä. Käyttötarkoituksen ja tiedon luonteen mukaan tulisi päättää mallinnetaanko taulukkoinformaatio tapahtumina vai yksinkertaisimmalla kolmikkotietokenttänä. Tietokentän käyttö on perusteltua, jos informaation voi olettaa staattiseksi. Esimerkiksi jo edesmenneen henkilön kohdalla lasten lukumäärä on vakioarvo, jota tietokannan myöhemmät päivitykset eivät enää voi muuttaa. Kun kolmikkotietokenttiä käytetään, voidaan CIDOC-mallin rinnalla käyttää tarkoitukseen soveltuvaa sanastoa kuten esimerkiksi FOAF:ia.

3.1.2 Sotahistorian mallintaminen

Monet toimijaontologian ratkaisut on tehty aihepiiriin eli sotahistoriallisen tiedon perusteella. Mitkä sitten ovat sotahistoriallisen tiedon tärkeimpiä mallinnettavia entiteettejä ja miten ne jäsenellään CIDOC CRM:n (kuva 3.1) mukaisiin perusluokkiin? Mitkä entiteeteistä voivat toimia itsenäisinä, ja mitkä taas ovat niin läheisesti kytköksissä toimijoihin, että ne joudutaan integroimaan osaksi ontologian suunnittelua? Millaisia muita sanastoja tarvitaan esimerkiksi alan terminologian tueksi? Seuraavassa esitetty jako on tarkoitettu suuntaa-antavaksi ja esimerkkimäiseksi entiteettiluokitteluksi eikä pyri olemaan kaikenkattava sotaontologiamalli.

Henkilötoimijoiden runkona ovat Suomen armeijassa palvelleet sotilaat ja lotat. Sotatoimet toki ovat koskettaneet valtaosaa silloisesta siviiliväestöstäkin.

Ulkomaisista henkilöistä voidaan mukaan lukea valtionpäämiehiä ja merkittävimpiä sotavoimien komentajia. Toimijaryhmät kattavat koko sotavoimien hierarkian maiden armeijoista ja niiden puolustushaaroista aina pienimpiin joukko-osastoihin, alusten miehistöihin, ilmavalvonta-asemiin ja väestönsuojelukeskuksiin. Suomen yhteydessä mukaan voidaan liittää suojeluskunnat ja Lotta Svärd -järjestö. Muita ryhmiä olisivat esimerkiksi eri maiden hallitukset.

Tapahtumilla muodostetaan kuvaus toimijoiden elämänvaiheista. Henkilön elinkaari rakennetaan synnyin- ja kuolintapahtumien väliin ja sitä täydennetään yksityiskohdilla, jotka tässä aihepiirissä esittävät sotilasuraa ja osuutta sotatoimiin. Tavoite on muodostaa ylennysten, ansiomitaleiden ja osastoihin liittymisien ketju; unohtamatta kohtalokkaampiakaan tapahtumia kuten haavoittumista, katoamista tai menehtymistä.

Ryhmän toiminta-ajan voi rajata sen perustamisen ja lopettamisen väliin. Perustamisten, siirtojen ja taistelujen tapahtumaketjusta rakentuu kokonaiskuva osaston liikkeistä. Koko armeijan hierarkia voidaan esittää osastojen liittämisinä toistensa alaisuuteen.

CIDOC CRM -luokittelun tyyppejä ovat henkilöiden perustiedot kuten kansalaisuus tai äidinkieli ja sotilasyhteydessä sotilasarvo tai toimenkuva joukko-osastossa. Ryhmien tyyppejä ovat niiden toimintaa ja laajuutta lajittelevat kategoriat. Sotilasosaston kuvaukseen saataisiin lisää semanttista sisältöä ilmaisemalla sen joukko-osastotyyppi; tieto osastosta armeijakuntana, rykmenttinä tai komppaniana antaisi arvion sen miesvahvuudesta, alaosastojen rakenteesta ja asemasta asevoimien kokonaishierarkiassa.

Sotahistoriallisessa kontekstissä tärkeimpiä esineitä ja käsitteitä ovat käytössä ollut varustus ja laitteisto sekä aihehistoriaa dokumentoiva informaatio. Suurin osa armeijan käyttämästä laitteistosta rajattiin tässä vaiheessa työn ulkopuolelle, lukuun ottamatta laivaston aluksia ja tykistön linnakkeita, jotka voidaan yhdistää niillä palvelleeseen miehistöön. Historiallinen informaatio käsittää käytössä olevan lähdemateriaalin: valokuvat, matrikkelit, kortistomateriaalin ja muistelma-artikkelit.

Paikat voidaan jakaa kahteen kategoriaan: paikannimirekisterimäisiin nimettyihin paikkoihin ja aihepiirin tapahtumapaikkoihin. Jälkimmäisillä tarkoitetaan esimerkiksi varuskuntia, taistelukenttiä, kenttäsaaraloita, korsuja ja linnakkeita. Historiallisen tiedon yhteydessä viitataan usein vanhoihin, jo käytöstä poistuneisiin paikannimiin; muistetaan etenkin, että talvi- ja jatkosodan taisteluja käytiin myöhemmin luovutetulla alueella. Ajat voidaan ilmaista tarkkoina päiväyksinä ja kellonaikoina tai viittaamalla toiseen tapahtumaan tai ajanjaksoon kuten *välirauhan aikana*.

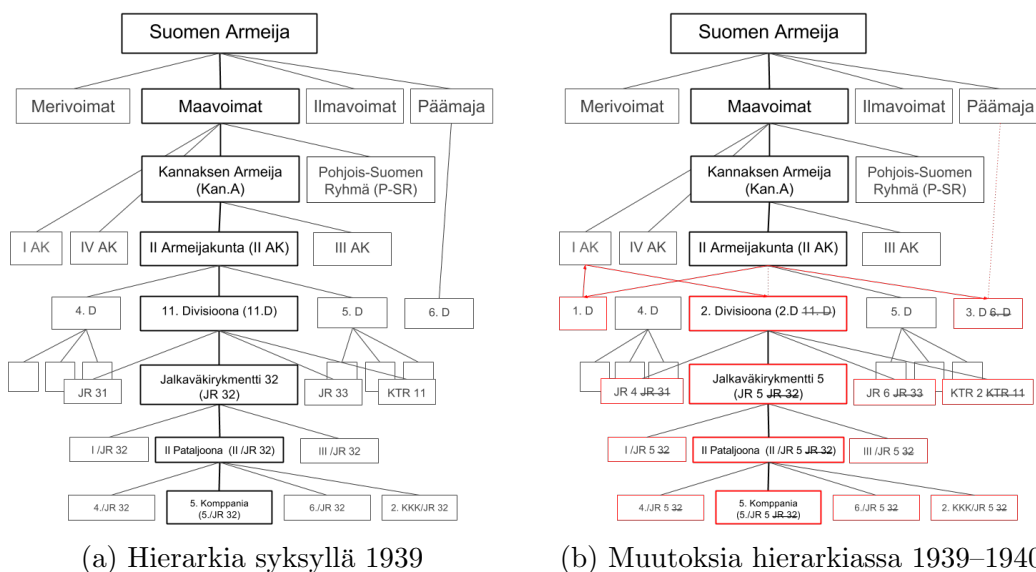
3.1.3 Tunnisteiden problematiikka

#	Kirjoitusasu	lyhenne (peiteluku)
1	Päämaja	PM
2	Laatokan meripuolustus	LMe.P
3	Panssarilaiva Ilmarinen	Psl I, P/L Ilmarinen (8671)
4	Viipurin suojeluskuntapiiri	Vii.SKP
5	II armeijakunta	II AK
6	11. divisioona → 2. divisioona	11. D → 2. D
7	Jalkaväkirykmentti 32 → Jalkaväkirykmentti 5	JR 32 → JR 5
8	Lentolaivue 22 → Lentolaivue 32 → Hävittäjälentolaivue 32	LLv 22 → LLv 32 → LeLv 32 → HLeLv 32 (8531)
9	Jalkaväkirykmentti 11, II pataljoona	II/JR 11
10	Jalkaväkirykmentti 11, 3. komppania	3./JR 11, 3./I/JR 11
11	Lentolaivue 10, 3. lentue	3./LLv 10

Taulukko 3.1: Joukko-osastojen nimiä, lyhenteitä ja niihin tehtyjä muutoksia.

Nimikkeet ovat henkilöiden ja ryhmien nimiä, lyhenteitä tai muita tunnisteita. Historiallinen aineisto voi tarjota henkilöstä kirkonkirjamaisen kuvauksen etu- ja sukunimiseen, syntymäpaikkoineen ja vanhempien nimiseen ja ammatteineen, eikä käytössä ole sosiaaliturvatunnuksen kaltaista varmuudella yksilöivää tunnistetta. Henkilöön voidaan viitata koko nimellä kuten *Jorma Karhunen*, lempinimellä ”*Joppe*” *Karhunen* tai etunimen initiaaleilla *J. Karhunen*. Sotilasaihepiiri tuo mukaan lisäksi tavat viitata henkilöön sotilasarvolla ja sukunimellä *kapteeni Karhunen*, *majuri Karhunen* tai osana osastoaan *Lentolaivue 24:n komentajana*. Merkille pantavaa on tässä se, että sekä sotilasarvo että joukko-osasto ovat usein ajan saatossa muuttuvia tarkenteita.

Taulukossa 3.1 on esimerkkejä joukko-osastojen tunnistekäytännöistä. Osastoilla on kirjoitusasuinen nimi (Jalkaväkirykmentti 11) ja lyhenne (JR 11), joiden lisäksi on voitu käyttää lempinimiä (Ässärykmentti, S-rykmentti). Myös lyhennekäytännöt saattoivat muuttua sodan aikana. Esimerkiksi talvisodassa ja jatkosodan alussa lentolaivue lyhennettiin LLv, 3.5.1942 lyhen-



(a) Hierarkia syksyllä 1939

(b) Muutoksia hierarkiassa 1939–1940

Kuva 3.2: Suomen armeijan hierarkia.

teeksi vahvistettiin LeLv ja 14.2.1944 lähtien käytettiin pommituslentolaivueesta lyhennettä PLeLv ja hävittäjälentolaivueesta lyhennettä HLeLv [42].

Joukko-osastojen nimien lisäksi myös koko armeijan hierarkiassa tehtiin useita suuria muutoksia. Kuva 3.2a havainnollistaa armeijan rakennetta talvisodan aikana. Ylimpänä hierarkiassa on Suomen armeija, jonka alla ovat aselajit kuten maa-, meri- ja ilmavoimat. Maavoimat koostui neljästä armeijakunnasta, joista toinen (II AK) ja kolmas (III AK) yhdessä muodostivat Kannaksen armeijan (Kan.A). Sodan alkaessa II armeijakuntaan kuului 4. D, 5. D ja 11. divisioona (11. D) ja pienempiä suojajoukkojen osastoja. Divisioonaa pienemmät osastot noudattivat kolmijakoa ja ne numeroitiin järjestelmällisesti niin että 1. divisioona (1. D) koostuu jalkaväkirykmenteistä JR 1, JR 2 ja JR 3. Vastaavasti 2. divisioonaan (2. D) kuuluvat Jalkaväkirykmentit JR 4, JR 5 ja JR 6. Lyhenteet noudattavat käytäntöä, jossa ennen nimeä tai kirjainlyhennettä käytetään joko roomalaista numeroa (IV AK) tai pisteellistä järjestyslukua (2. D), nimen perässä sen sijaan kardinaalilukua (JR 5).

Vihollisen tiedustelun ja sodanjohdon häämäämiseksi tehtiin nimenvaihtoja 1.1.1940. Tällöin esimerkiksi 11. divisioona (11. D) nimettiin 2. divisioonaksi (2. D) (kuva 3.2b), osasto myös siirtyi II armeijakunnan (II AK) alaisuudesta I armeijakuntaan (I AK). Hierarkkisesti periytyvässä järjestelmässä tämä vaikutti myös alaosastojen nimiin: aiemmista Jalkaväkirykmenteistä 31, 32 ja 33 tehtiin näin Jalkaväkirykmentit 4, 5 ja 6. Muutos heijastui hierarkiassa alaspäin edelleen pataljoonien ja komppanioiden tunnuksiin, esimerkiksi

JR 5:n kolmas pataljoona (III/JR 5). Esimerkkinä uudelleenjärjestelyistä voidaan tarkastella alun perin Päämajan alaista 6. divisioonaa, joka 3. divisioonaksi muutettuna liitettiin II armeijakuntaan. Kaikki edellä esitetyt muutokset tehtiin talvisodan 1939–1940 aikana. [39] Jatkosodan aikana perustettiin samannimisiä osastoja kuin talvisodassa, tosin eri paikkakuntien ja sotilaspiirien miehistöstä, joten kyse on täysin uudesta osastosta.

3.2 Laadittu toimijaontologiamalli

Lyhenne	Nimiavaruus
crm:	http://www.cidoc-crm.org/cidoc-crm/
dc:	http://purl.org/dc/elements/1.1/
dct:	http://purl.org/dc/terms/
foaf:	http://xmlns.com/foaf/0.1/
mil:	http://rdf.muninn-project.org/ontologies/organization#
owl:	http://www.w3.org/2002/07/owl#
rdf:	http://www.w3.org/1999/02/22-rdf-syntax-ns#
schema:	http://schema.org/
skos:	http://www.w3.org/2004/02/skos/core#
xsd:	http://www.w3.org/2001/XMLSchema

Taulukko 3.2: Nimiavaruuksia ja niistä käytettyjä lyhenteitä.

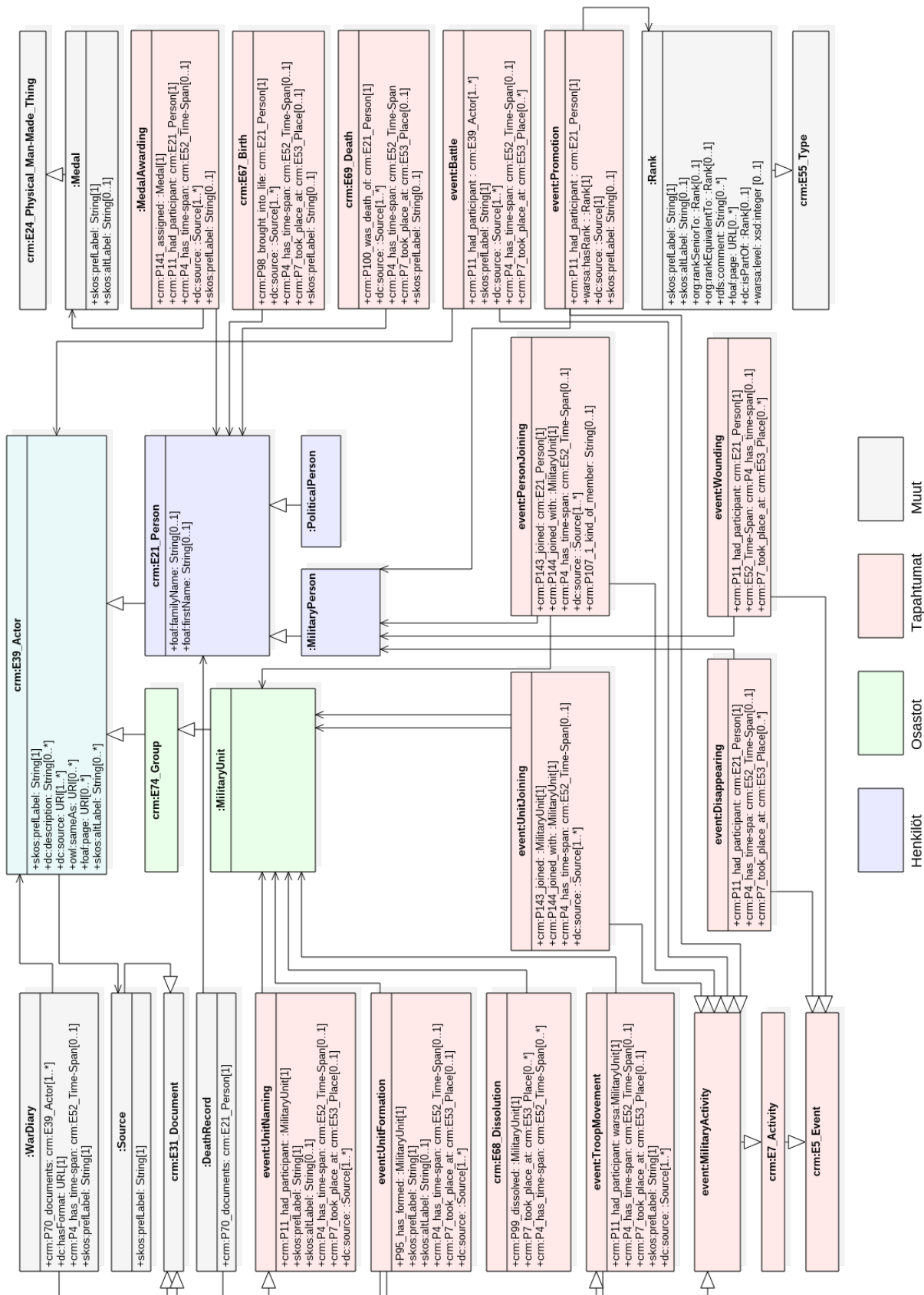
Työssä laadittu toimijaontologiamalli on CIDOC CRM -perustainen ja sitä on laajennettu keskeisillä sanastoilla kuten Dublin Core (DC) [17], FOAF [9], OWL [45] ja SKOS [46]. Käytetyt nimiavaruudet ja niiden lyhenteet on koottu taulukkoon 3.2. Ontologian keskeiset luokat on esitetty kaaviossa 3.3 ja skeema liitteessä 1. Toimijoiden yhteinen yläluokka `crm:E39_Actor` on kaaviossa ylhäällä keskellä.

Henkilötoimijoille on kolme luokkaa ja ryhmille kaksi. Toimijoihin liittyvien tapahtumien kuvaamiseen on 19 luokkaa, joiden yhteinen yläluokka on `crm:E5_Event`. Kokonaisuuteen kuuluu myös yksittäisiä luokkia, joilla ilmaistaan sotilasarvoa, sotapäiväkirjaviittausta, kunniamitalin tyyppiä ja tietolähdettä.

Toimijoiden yläluokkaan `crm:E39_Actor` on sisällytetty toimijoiden yhteiset ominaisuudet. Näitä ovat resurssin ensisijainen (`skos:prefLabel`) ja toissijainen (`skos:altLabel`) nimi. Näiden arvot ovat merkkijonotyyppisiä. Yleisenä käytäntönä on käyttää ensisijaisena nimenä lyhyttä kuvausta resurssin sisällöstä eli useimmiten kirjoitusasuista nimeä. Toimijalla saa olla vain yksi `skos:prefLabel`-ominaisuuden arvo, paitsi jos resurssille halutaan ilmaista erikielisiä nimiä, tällöin kuitenkin vain yksi kutakin kielileimaa kohti. Resurssin mahdollisia vaihtoehtoisia nimiä, lempinimiä tai lyhenteitä ilmaistaan `skos:altLabel`-ominaisuuden arvoina.

Ominaisuuden `dc:description` arvona on kirjallinen kuvaus resurssista tai mahdollisesti useampi erikielisinä. Lähdetieto annetaan `dc:source`-ominaisuuden arvona. Sen arvo voi olla viittaus suoraan `www`-osoitteeseen tai merkkijonomuotoisena kuvaus käytetystä kirjallisesta lähteestä. Yhdelle toimijalle voi kertyä viittauksia useisiin lähteisiin. Tiedon jäljitettävyyden takia tulisi `dc:source`:lla olla aina vähintään yksi arvo. Toimijaresurssin `owl:sameAs`-ominaisuudella viitataan muihin vastaavaan resursseihin eli samaa ihmistä tai ryhmää kuvaaviin tietoihin toisissa tietokannoissa, kuten Wikidatassa [60] tai DBpediassa [16]. Ominaisuuden `foaf:page` arvot ovat resurssiin liittyviä `www`-osoitteita.

Myös tapahtumaluokilla on useita yhteisiä ominaisuuksia. Niiden `skos:prefLabel`-, `dc:description`- ja `dc:source`-ominaisuuksia käytetään samoin kuin toimijaluokkienkin. Tapahtumat noudattavat yhteistä CIDOC CRM -käytäntöä toimija-, aika- tai paikkayhteyden ilmaisemiseen. Toimija liitetään tapahtumaan `crm:P11_had_participant`-ominaisuuden tai jonkin siitä periytetyn aliominaisuuden arvona. Tapahtuma-aika annetaan liittämällä `crm:E52_Time-Span`-instanssi ominaisuuden `crm:P4_has_time-span` arvoksi. Tapahtumapaikka ilmaistaan liittämällä ominaisuuden `crm:P7_took_place_at` tai `crm:P8_took_place_on_or_within` arvoksi paikkaluokan `crm:E53.Place` instanssi. Paikka voi saada useita arvoja, esimerkiksi kunnan ja kylän tarkkuudella. Joillekin tapahtumille, kuten henkilöiden ylennyksille, ei voida olettaa paikkamäärettä. Sitä vastoin taistelut voivat sijoittua laajemmalle, useinkin kunnan alueelle.



Kuva 3.3: Toimijoiden ja tapahtumien RDF-malli.

3.2.1 Henkilön mallintaminen

Henkilöiden mallintamiseen käytetään kolmea eri luokkaa: `crm:E21_Person`, `:MilitaryPerson` ja `:PoliticalPerson`. Luokilla ei ole toiminnallista eroa, vaan jako on tehty lajittelemaan henkilöitä rooliensa perusteella. Armeijassa palvelleet henkilöt mallinnetaan luokan `:MilitaryPerson` instansseina ja aikakauden politiikan vaikuttajat luokan `:PoliticalPerson`. Henkilöitä, joita ei voida luokitella kumpaankaan kategoriaan, kuten kulttuurin, tiedemaailman, urheilun tai viihteen vaikuttajat, mallinnetaan luokalla `crm:E21_Person`. Roolit ovat usein päällekkäisiä, ja tällainen rajanveto voi olla hankalaa: sotilashenkilöt ovat voineet olla siviilissä aktiivisia poliittisia ja kulttuurialojen toimijoita. Tällaiset henkilöt voidaan mallintaa useamman luokan instansseina.

Aliluvussa 3.2 on esitelty henkilöresurssien pääluokasta `crm:E39_Actor` periytyvät ominaisuudet `skos:prefLabel`, `skos:altLabel`, `dc:description`, `dc:source`, `owl:sameAs` ja `foaf:page`. CIDOC CRM -skeemasta poiketen henkilön nimeä ei ole esitetty erillisenä `crm:E82_Actor_Appellation`-nimikkeenä. Tätä yksinkertaistusta voidaan perustella sillä, että suurimmalla osalla henkilöistä ei nimessä tapahdu muutosta tai lähdemateriaali käyttää viimeisintä tunnettua nimeä. Sen sijaan nimi jaetaan suku- ja etunimiin käyttäen `foaf:familyName`- ja `foaf:firstName`-kenttiä. Näistä yhdistetty koko nimi esitetään `skos:prefLabel`-kentän arvona. Mahdollinen lempinimi kirjataan `foaf:firstName`-kenttään suluissa eroteltuna kuten *Jorma (Joppe) Karhunen*. Lähdeaineistossa henkilöön saatetaan viitata tittelillä ja sukunimellä, esimerkiksi *kapteeni Karhusena*, jolloin `foaf:familyName` on ainoa tunnettu tietokenttä ja etunimikenttä jätetään määrittelemättä. Kenttä `foaf:firstName` saattaa niinkään koostua vain henkilön nimikirjaimista kuten *J. Karhunen*. Sellaisiin synonyymeihin, joiden ei haluta näkyvän `skos:prefLabel`-kentässä, voidaan käyttää `skos:altLabel`-kenttää. Tällaisia tapauksia ovat esimerkiksi *Marski* tai *Ylipäällikkö* viitattaessa Carl G.E. Mannerheimiin.

Ominaisuutta `owl:sameAs` on käytetty myös työnaikaisena linkkinä, kun kahden resurssin on tunnistettu viittaavan samaan henkilöön. Vastakkaisessa tapauksessa voidaan käyttää `owl:differentFrom`-ominaisuutta: kaksi samannimistä henkilöä, jotka ovat eroteltavissa jonkin muun heitä yksilöivän ominaisuuden perusteella. Ontologiamalli ei tarjoa valmista ratkaisua esimerkiksi henkilön valokuvalle, ammatille tai sukupuolelle, mutta ne voidaan ilmaista laajentamalla mallia esimerkiksi FOAF-standardin mukaan tai käyttötarkoituksen mukaan hakea federoituna SPARQL-kyselynä lähteenä toimivasta tietokannasta. Poiminta henkilötoimijan RDF-resurssista tapahtumaesimerkkeineen on esitetty liitteessä 2.

Henkilöön liittyviä tapahtumia ovat elinaikaa kuvaavat synnyin- ja kuolemistapahtuma, sotilasuraa kuvaavat ylennykset ja kunniamitalit sekä sota-toimitapahtumat kuten taistelu, haavoittuminen tai katoaminen. Malli sallii uusien tapahtumaluokkien kuten kotiuttamisen, sotavankeuden tai eläkkeelle siirtymisen lisäämisen.

Henkilön elinkaari esitetään mallissa synnyin- ja kuolintapahtumina, jotka ovat luokkien `crm:E67_Birth` ja `crm:E69_Death` instansseja. Synnyintapahtuma linkittyy henkilöön `crm:P98_brought_into_life`- ja kuolintapahtuma `crm:P100_was_death_of`-ominaisuudella. Tapahtumaluokat sallivat myös lapsi-vanhempi-suhteiden mallintamisen kenttien `crm:P96_by_mother` ja `crm:P97_from_father` avulla. Jotta synnyintapahtumainstanssi olisi mielekäs, tulisi sillä olla vähintään joko aika- tai paikkatieto tai tieto ainakin toisesta vanhemmasta.

Sotilasarvo on tärkeä toimijaa määrittelevä attribuutti, ja sen kuvaamiseen haluttiin helposti laajennettava malli, joka sallii pitkän sotilasuran esittämisen tapahtumien avulla. Henkilön sotilasarvo ilmaistaan liittämällä henkilö tapahtumaan `:Promotion`. Linkki sotilasarvoon (aliluku 3.2.3) annetaan ominaisuuden `:hasRank` arvona. Jokaisella `:Promotion`-instanssilla tulee olla sekä henkilö- että sotilasarvokenttä.

Tapahtuma luodaan aina kertomaan henkilön arvo, vaikkei päivämäärää tai aiempaa arvoa tiedettäisikään, joten tapahtuma ei itse asiassa kuvaa ylennystä vaan tietyn sotilasarvon saamista. Tapahtuman ajalla `crm:P4_has_time-span` kentässä tarkoitetaan sotilasarvon saamishetkeä eikä ajanjaksoa, jona henkilölle on ollut ilmoitettu sotilasarvo. Lähteissä tieto ilmoitetaan useimmiten juuri tässä muodossa. Jos tietoa on riittävästi, voidaan ajanjakso päätellä peräkkäisten ylennysten aikaleimoista, koska jakson loppu on seuraavan ylennystapahtuman alkuhetki. Saman ajanhetken lisääminen kahteen eri resurssiin alku- ja loppupisteiksi tuottaisi tietomalliin turhaa päällekkäisyyttä. Toisaalta saman instanssin käyttö toisen tapahtuman loppu- ja toisen alkupisteinä sitoisi tapahtumat toisiinsa, osoittaisi niiden peräkkäisyyden ja antaisi näin malliinnukseen lisää semanttista syvyyttä.

Henkilön kuuluminen joukko-osastoon ilmaistaan liittämällä henkilö tapahtuman `:PersonJoining` ominaisuuden `crm:P143_joined` arvoksi. Joukko-osastoresurssi (aliluku 3.2.2) linkitetään ominaisuudella `crm:P144_joined_with`. Tapahtuma luodaan aina ilmaisemaan henkilön joukko-osastoa, vaikka ajankohta ei olisikaan tiedossa. Jos nimellä tunnettua osastoa ei ole graafissa, voidaan nimi tai lyhenne lisätä tekstimuotoisena `:hasUnit`-kenttään. Kentän perusteella on tarkoitus linkittää tapahtuma mahdolliseen myöhemmin lisättävään joukko-osastoresurssiin.

Tapahtumaan voidaan liittää henkilön rooli joukko-osastossa, kuten ”*komentaja*” tai ”*esikuntapäällikkö*”. Tieto annetaan tekstimuotoisena `crm:P107_1_kind_of_member`-kentän arvona. Toisin kuin ylennyksissä, on lähteissä useimmiten ilmoitettu jakso, jonka henkilö on palvellut kyseisessä osastossa.

Niin sotilasarvon saamis- kuin osastoonkuulumistapahtumiin olisi mahdollista liittää yhteen resurssiin useampi toimija. Näin voitaisiin tehdä, kun kaikki muut tietokentät ovat samoja, esimerkiksi jos henkilöistä tiedetään vain sotilasarvo ilman tarkempaa päivämäärätietoa. Tällöin tietokannan koko pienenee, mutta näin ei kuitenkaan kannata tehdä, koska mallin ylläpito olisi hankalampaa, kun aikaleimoja lisätään tai päivitetään.

Kunniamitalin myöntäminen esitetään luokan `:MedalAwarding` instanssina. Siihen on linkitetty henkilön lisäksi kunniamitalityyppi (aliluku 3.2.5) ominaisuudella `crm:P141_assigned`. Tapahtumalla edellytetään olevan sekä henkilö- että mitaliviittaus. Jos sama henkilö on palkittu samalla kunniamitalilla useamman kerran, mallinnetaan tapahtumat itsenäisinä omilla aikaleimoillaan.

Haavoittumis- ja katoamistapahtumat ovat keskenään hyvin samankaltaisten luokkien `:Disappearing` ja `:Wounding` instansseja. Luokissa on viittaukset henkilöön, ajankohtaan ja tapahtumapaikkaan. Uutis-, poliittiset ja taistelutapahtumat ovat osa Erkki Heinon työtä [25]. Niissä tapahtuman kenttä `crm:P11_had_participant` voi saada arvokseen niin henkilön kuin joukko-osastonkin. Taistelutapahtumaan `:Battle` liitetty henkilö on yleensä osaston komentaja tai taistelukoneen pilotti. Ei tietenkään ole mielekästä liittää kaikkia osastoon kuuluvia taistelutapahtumaan, koska tämä tieto on pääteltävissä osastoon kuulumisen kautta.

3.2.2 Joukko-osaston mallintaminen

Kaikki joukko-osastot ovat luokan `:MilitaryUnit` instansseja. Aliluvussa 3.2 on esitelty osastoresurssien pääluokasta `crm:E39_Actor` periytyvät ominaisuudet `skos:prefLabel`, `skos:altLabel`, `dc:description`, `dc:source`, `owl:sameAs` ja `foaf:page`. Joukko-osaston tärkeimmät tunnisteet ovat sen kirjoitusasuinen nimi kentässä `skos:prefLabel` ja siitä käytettävä lyhenne kentässä `skos:altLabel`. Usein osastosta on eri vaiheissa käytetty eri nimiä ja lyhennejä. Tällöin osastoinstanssiin kirjataan ne, joita halutaan käyttää sen ensisijaisina tunnisteita, esimerkiksi ajallisesti viimeisimmät.

Osastoon liitettävät tapahtumat käyttävät samoja aika-, paikka- ja lähdetieto-ominaisuuksia kuin tapahtumien yhteinen yläluokka (aliluku 3.2). Osaston

perustaminen ilmaistaan liittämällä tapahtuman `crm:E66_Formation` ominaisuuden `crm:P95_has_formed` arvoksi linkki osastoresurssiin. Tapahtuman kentät `skos:prefLabel` ja `skos:altLabel` sisältävät yksikön alkuperäisen nimen, joka voi olla eri kuin myöhemmin sotatoimissa käytetty. Osaston lopettaminen ilmaistaan tapahtumana `crm:E68_Dissolution`, jossa osastoon viitataan ominaisuudella `crm:P99_dissolved`.

Mahdolliset myöhäisemmät nimivaihdokset mallinnetaan `:UnitNaming`-tapahtumina, joissa nimi ja lyhenne annetaan kentissä `skos:prefLabel` ja `skos:altLabel`. Etsittäessä osastoa tietyllä nimellä on siis tutkittava `:MilitaryUnit`-instanssin ja siihen liittyvien tapahtumien `crm:E66_Formation` ja `:UnitNaming` nimikentät. Osaston liikkeitä ja keskityksiä mahdollisine aika- ja paikakaleimoinen mallinnetaan `:TroopMovement`-tapahtumina. Osastolle käytetään samaa taistelutapahtumaa `:Battle` kuin henkilötoimijallekin.

Osastot kuuluvat toistensa alaisuuteen ja niistä muodostuu koko armeijan rakennetta kuvaava hierarkkinen puu. Ontologiassa tämä puurakenne ja sen ajalliset muutokset esitetään `:UnitJoining`-tapahtumien avulla. Luokan instanssit toimivat graafin kaarina liittäen osaston `crm:P143_joined` ylemmän osaston `crm:P144_joined_with` alaisuuteen. Liitteessä 3 on esimerkki osaston RDF-määrittelystä.

3.2.3 Sotilasarvot

Sotilasarvojen termiontologia sisältää yhden ainoan luokan `:Rank`, jolla kuvataan sotilasarvoja ja sotilasarvoluokkia. Luokan instansseja käytetään henkilön ylennystapahtuman `:Promotion` sotilasarvoa kuvaavan ominaisuuden `:hasRank` arvona (aliluku 3.2.1). Luokka `:Rank` perii `crm:E55_Type`-luokan. Sotilasarvon erikieliset nimet kootaan `skos:prefLabel`-kenttään ja käytettyjä lyhenteitä varten on `skos:altLabel`-kenttä. Liitteessä 4 esitetään RDF-esimerkki sotilasarvon määrittelystä.

Sotilasarvojen sanaston lisäksi niille voidaan rakentaa hierarkia. Se luodaan muodostamalla linkki ylemmästä arvosta alempaan ominaisuudella `mil:rankSeniorTo`. Eri aselajien samantasoiset arvot rinnastetaan ominaisuudella `mil:rankEqualTo`. Yksittäiset sotilasarvot on ryhmitelty sotilasarvoluokkiin, kuten miehistö, aliupseerit ja upseerit. Sotilasarvo voidaan jakaa tyyppeihin myös maidensa mukaan kuten puna-armeijan tai saksalaiset sotilasarvot. Sotilasarvo yhdistetään sotilasarvotyyppiinsä ominaisuudella `dct:isPartOf`, joka voi saada useita arvoja kuten *puna-armeijan arvo* ja *upseeri*.

Sotilasarvoon on lisäksi liitetty `:level`-kokonaislukuarvo, joka on sotamie-

helle 1, korpraalille 2, jatkuen aina Suomen marsalkan arvoon 21. Ominaisuus on tarkoitettu sotilasarvojen vertailuun; kokonaislukujen järjestely on helpompaa kuin hierarkian hakeminen erikseen tietokannasta. Ominaisuus myös helpottaa visualisointien tuottamista, koska sen avulla voidaan esimerkiksi esittää ylennyksiä aikajanalla.

3.2.4 Sotapäiväkirjat

Sotapäiväkirjaluokan instansseja käytetään linkittämään toimija Kansallisarkiston sotapäiväkirjasivustolle [34]. Instanssien luokka `:WarDiary` on periydytetty `crm:E31_Document`-luokasta. Linkki päiväkirjan verkkosivulle annetaan `dct:hasFormat`-kentässä. Aika, jona päiväkirjaa on pidetty, annetaan `crm:P4_has_time-span`-kentässä. Linkki päiväkirjaa laatineeseen henkilöön tai osastoon on `crm:P70_documents`-kentässä. Otsikkomaisesta `skos:prefLabel`-kentästä on luettavissa päiväkirjan laatija ja ajanjakso. Liitteessä 5 on kaksi RDF-muotoista esimerkkiä sotapäiväkirjainstansseista.

3.2.5 Mitalityypit

Mitalityypillä esitetään myönnettyjen kunniamerkkien eri tyyppisiä kuten *Vapaudenristin suurristi* tai *3. lk:n Vapaudenristi*. Instanssit ovat `crm:E24_Physical_Man-Made_Thing`-luokasta periytyvän `:Medal`-luokan resursseja ja ne kuvaavat mitalityyppejä (aliluku 4.3.7). Resurssissa on `skos:prefLabel`-tyypin attribuutti mitalin kirjoitusasulliselle nimelle ja mahdollinen `skos:altLabel`-attribuutti siitä käytetylle lyhenteelle. Mitalisanastolla ei ole hierarkiaa, ts. luokittelua, jossa koottaisiin yhteen esimerkiksi erityyppiset vapaudenristit¹. Mitalityyppejä käytetään `:MedalAwarding`-tapahtumien `crm:P141_assigned`-kentässä. Esimerkki mitalityyppien määrittelystä on liitteessä 6.

3.2.6 Tietolähteet

Tietolähde `:Source` on `crm:E31_Document`:in aliluokka. Luokan instanssien `skos:prefLabel`-kenttä kertoo käytetyn tietolähteen kirjoitusasullisen nimen. Tietolähteitä käytetään toimija- ja tapahtumaresurssien `dc:source`-kentissä tallentamaan viittaus tietolähteeseen.

¹https://fi.wikipedia.org/wiki/Vapaudenristin_ritarikunta#Ritarikunnan_nykyiset_kunniamerkit

Luku 4

Sotasammon toimijat

Tässä luvussa esitellään Sotasampo-projekti ja perehdytään sen toimijatietokannan kokoamiseen, työssä käytettyihin tietolähteisiin ja prosessointiin yhtenäiseksi tietokannaksi.

4.1 Sotasampo-projekti

Sotasampo-projekti aloitettiin vuonna 2015 Aalto-yliopiston semanttisen laskennan tutkimusryhmässä. Projektia on tehnyt yhteistyössä Kansallisarkiston, Puolustusvoimien, Maanmittauslaitoksen, Suomen Sotahistoriallisen Seuran ja muun yhteistyöverkoston kanssa. Sotasampo-tietopalvelun on tarkoitus kattaa talvi-, jatko- ja Lapin sodan aineistoa vuosilta 1939–1945. Aineistoa on kerätty sotilas- ja siviilihenkilöistä, sota-ajan tapahtumista, armeijan rakenteesta ja toiminnasta sekä silloisista paikannimistä. Tavoite on ollut yhdistää tietoa eri lähteistä kuten verkkoon jo luoduista tietokannoista, luettelo- tai matrikkeliuotoisista paperi- tai digitaalisista julkaisuista tai muistelmakirjoituksista. [27]

Aineistoa on kerätty seuraavista lähteistä [25]:

1. Menehtyneet-tietokanta sisältää tietoa sota-aikana kuolleista, haavoittuneista ja kadonneista. Tietokanta on aiemmin julkaistu sekä Kansallisarkiston portaalina¹ että avoimena datana² [36].
2. Sotapäiväkirjat-kokoelma on digitoitu alkuperäisistä joukko-osastojen päiväkirjoista [34].

¹<http://kronos.narc.fi/menehtyneet/>

²<https://www.avoindata.fi/data/fi/dataset/suomen-sodissa-1939-1945-menehtyneet>

3. SA-kuva-arkistoon on koottu sodan ajan tiedotuskomppanioiden tallentamaa valokuvamateriaalia [49].
4. Kansa Taisteli -lehti ilmestyi vuosina 1957–1986 ja julkaisi etupäässä rintamalla olleiden kertomuksia heidän kokemuksistaan [53].
5. Paikka- ja kuntatietoa sota-ajan Suomesta, etenkin Neuvostoliitolle luovutetuilta alueilta [29, 32, 41].
6. Organisaatiokortistoon on koottu tietoa armeijan yksiköistä [31] .
7. Kansallisbiografia ja semanttinen kansallisbiografia sisältävät tietoa yli 6 300 merkkihenkilöstä [6, 24].
8. Sodan tapahtumia eri lähteistä koottuna, esimerkiksi tapahtumalista teoksessa Talvisodan pikkujättiläinen [40].
9. Tietoa henkilöistä ja joukko-osastoista koottuna eri lähteistä kuten sotahistoriallisesta kirjallisuudesta tai Wikipediasta.

Sotasampo-aineiston kokoaminen aloitettiin kesäkuussa 2015. Huhtikuussa 2016 Sotasammon tietokanta koostui yli 7,4 miljoonasta RDF-kolmikosta. Oma osuuteni tästä työstä on ollut tiedon kokoaminen henkilöistä (1, 2, 3, 7 ja 9) ja joukko-osastoista (2, 6 ja 9) sekä tämän tiedon muokkaaminen ja siltaaminen muihin tietojoukkoihin (2, 5, 7 ja 9).

4.2 Toimijoiden taustaa

Sotasammon tietokanta sisältää lähes 100 000 henkilöä. Pääpaino on Suomessa sotatoimiin osallistuneilla, mutta mukaan on luettu myös aikakauden politiikan, tieteen, urheilun ja kulttuurielämän vaikuttajia sekä Suomesta että ulkomailta. Joukko-osastot käsittävät yli 3 000 Suomen armeijan yksikköä. Niihin kuuluu maa-, ilma- ja merivoimien osastoja, lääkintämuodostelmia, kotiseudun ilmatorjunta- ja ilmavalvonta-asemia, suojeluskuntia sekä ulkomaalaisten vapaaehtoisten osastoja. Työn alkaessa vuonna 2015 rajattiin osastot talvisodan aikaina toimineisiin suomalaisjoukkoihin. Osa niistä on toiminut jatkosodassakin, muutoin jatkosodan [37] ja puna-armeijan osastot [43] jätettiin myöhemmin täydennettäviksi. Henkilöiden suhteen ei tehty ajallista rajausta.

4.2.1 Joukko-osastojen nimet ja lyhenteet

Joissakin lähteissä osastoon viitataan sen lyhenteellä, osassa taas koko kirjoitusasulla, joten oli selvää, että näitä jouduttaisiin yhdistämään toisiinsa. Osastojen nimiä ja niitä vastaavia lyhenteitä kerättiin taulukkoon. Taulukon pohjana käytettiin soveltuvilta osin Arkistolaitoksen sivuilla olevaa listaa puolustusvoimien lyhenteistä [35]. Kääntäminen automatisoitiin kirjoittamalla sitä varten lyhyt Python-kielinen ohjelma, joka lukee sanastonsa käännohjelmalistasta ja etsii sieltä vastinetta käännettävälle merkkijonolla. Ohjelma käyttää säännöllisiä lausekkeita erotellessaan tulkittavasta merkkijonosta numerokoodeja tai erikoismerkeillä eroteltuja kenttiä.

Lyhennekäytännöt voidaan jakaa kolmeen periaatteelliseen tapaukseen (taulukko 3.1). Ensimmäisessä tapauksessa pelkkä kirjainyhdistelmä yksilöi osaston, esimerkkinä Päämaja (PM) tai Laatokan meripuolustus (LMe.P). Toisessa tapauksissa riittää kirjainlyhenteen korvaaminen kirjoitusasuisella nimellä. Yksikkö nimeltään II armeijakunta lyhennetään II AK. On vakiintunut käytäntö, että edessä käytetään roomalaista (II AK) tai arabialaista järjestysnumeroa (3. D) ja lopussa arabialaista kardinaalilukua (JR 11).

Monimutkaisemmassa tapauksessa käy nimestä ilmi minkä ylemmän osaston osa on kyseessä. Esimerkkejä tällaisesta ovat Jalkaväkirykmentti 11, 3. komppania (3./JR 11) ja Lentolaivue 10, 3. lentue (3./LLv 10). Lyhenteen alussa on ensimmäisenä varsinaisen osaston numero ja sen perässä kautta- viivalla eroteltuna yksi tai kaksi hierarkiassa ylempää osastoa. Esimerkeissä numero 3. käännetään joko komppaniaksi tai lentueeksi, joten kääntäminen edellyttää tietoa osastojen rakenteista.

Nimikkeiden kääntämisellä standardoitiin kirjoitusasut. Osastojen moniosaiset nimet kirjoitetaan pilkulla eroteltuina perusmuodossaan. Lyhenteet saatettiin formaattiin, jossa kauttaviivojen ja välilyöntien käyttö on yhdenmu- kaista.

4.3 Toimijatietokanta

Tähän alilukuun on poimittu esimerkkejä RDF-datan tuottamisesta eri tietolähteistä. Taulukossa 4.1 eritellään lähteitä, joista henkilöiden ja osastojen tiedot ovat peräisin. Aineisto jakautuu kirjalliseen ja sähköiseen materiaaliin. Koska yksittäisen toimijan tiedot voivat olla peräisin useammasta lähteestä, on alimmalla rivillä ilmoitettu kokonaismäärä pienempi kuin erilisistä lähteistä laskettujen lukumäärien summa.

Kirjallinen aineisto	Henkilöitä	Osastoja
Talvisodan organisaatiokortisto [31]	278	132
Suomen rintamamiehet 1939–1945 -kirjasarja [1]	311	69
Suomi Sodassa [2]	217	
Talvisodan historia 1-4 [39]		n. 800
Sähköinen aineisto		
Menehtyneet-tietokanta [36]	94 683	
Wikipedia	2 540	106
SA-kuva-arkisto. Puolustusvoimat [49]	1 718	
Semanttinen kansallisbiografia [6]	705	
Mannerheim-ristin ritarit [42]	191	14
Sotapäiväkirjat-kokoelma [34]	12	2 207
Luettelo komentajista. Ohto Manninen, MPKK	56	1
Muita lähteitä	13	
Kaikkiaan tietokannassa	99 464	3 044

Taulukko 4.1: Henkilöiden ja osastojen lukumääriä tietolähteittäin.

4.3.1 Organisaatiokortit

Organisaatiokortistossa on puolustusvoimien sodan jälkeen keräämää aineistoa maavoimien joukko-osastoista. Aineisto talvisodan osalta saatiin Arkistolaitokselta. Skannatut PDF-tiedostot sisälsivät kaikkiaan noin 1 600 korttia.

Esimerkki organisaatiokortista on esitetty kuvassa 4.1. Kortin vasemmassa ylänurkassa (a) on osaston nimi ja lyhenne, esimerkkitapauksessa Jalkaväkirykmentti 7 (JR 7). Oikeassa ylänurkassa (b) on osasto, jonka alaisuuteen JR 7 on kuulunut eli 4. divisioona (4. D). Kohdassa (c) on perustamispaikka ja -päivämäärä. Lisätietona voi olla osaston silloinen nimi, joka tässä tapauksessa on Jalkaväkirykmentti 17 (JR 17). Kohdassa (d) lukee mistä miehistö on koottu, tässä tapauksessa Tampereen sotilaspiiristä (Tam.sp.). Kohdassa (e) on lisätietoa mahdollisista muutoksista. Tässä tapauksessa nimeksi on 31.12.1939 muutettu JR 7. Taistelutapahtumia luetellaan kohdassa (f). Kullakin rivillä mainitaan taistelupaikka, päivämäärät ja osastoa silloin johtanut henkilö.

Muunnos PDF-tiedostoista RDF-muotoon eteni useissa vaiheissa. Aluksi PDF-tiedostojen sivut muunnettiin yksittäisiksi PNG-kuvatiedostoiksi. Näiden kont-

Jalkaväkirykmentti 7		
a)	JR 7.	3. Dsaan.
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.DE:n kirj.№ 52/III/34/39 sal./28.12.39)	
f)	4. Hyökkäys Summa-joella	23.12.39 Eversti K.A.Heiskanen
	Summan puolustustaistelut	4.1 - 13.2.40 Eversti K.A.Heiskanen
	Viivytystaistelut Summa-Säiniö (I/JR 7)	13.2 - 17.2.40 Ratsimest.K.O.Alfthan
	Viipurin puolustustaistelut	20.2 - 13.3.40 Eversti K.A.Heiskanen.

Kuva 4.1: Organisaatiokortin tietokentät.

rastia kasvatettiin, niistä poistettiin kohinaa ja sivun asemointia oikaistiin tarvittaessa. Muunnos ja käsittely tehtiin komentoriviltä käytettävällä työkalulla ImageMagick [30]. Seuraavaksi kuvatiedostot OCR-muunnettiin tekstimuotoon, mihin käytettiin Tesseract-työkalua [51]. Tekstin muuntaminen ei ollut ongelmaton, ja tuotettua tekstiä jouduttiin editoimaan runsaasti. Tyypillisiä ongelmia olivat kirjainten kääntyminen numeroiksi ($l \rightarrow 1$), suomenkielisiin aakkosiin kuulumattomiksi merkeiksi ($ö \rightarrow ü$) tai kahdentuminen ($m \rightarrow rn$). Suurin osa editoinnista tehtiin automaattisesti säännöllisillä lausekkeilla, esimerkiksi korjaamalla kahden kirjaimen välissä olevan numero 1 pieneksi l-kirjaimeksi. Sotilassanaston termien runsas toistuminen aineistossa vähensi manuaalista editointia. Päivämäärien automaattisessa korjaamisessa auttoi tieto tapahtumien sijoittumisesta vuosille 1939–1940. Jotta työmäärä säilyisi kohtuullisena, valittiin korteista ne, jotka sisälsivät maavoimien tärkeimmät osastot kuten divisioonat, jalkaväki- ja kenttätykistörykmentit.

Tuotetusta tekstistä eroteltiin alkuperäisen kortin otsikkokenttien nimitiedot (a, b), perustamistiedot (c), lisätiedot (d), muutokset (e) ja taistelut (f kuvassa 4.1). Data jaettiin tyypeittäin viiteen CSV-tiedostoon, jotka muunnettiin RDF-muotoon. Datan perusteella luotiin instansseja joukko-osastoille :MilitaryUnit ja niissä toimineille henkilöille :MilitaryPerson ja heidän sotilasarvoilleen :Promotion, sekä laaditun toimijamallin mukaiset tapahtumat osastojen perustamiselle crm:E66_Formation, niiden hierarkialle :UnitJoining, osaston siirroille :TroopMovement, uudelleennimeämisille :UnitNaming ja taisteluille :Battle. Esimerkki osaston RDF-kuvauksesta on liitteessä 3.

4.3.2 Muu kirjallinen aineisto

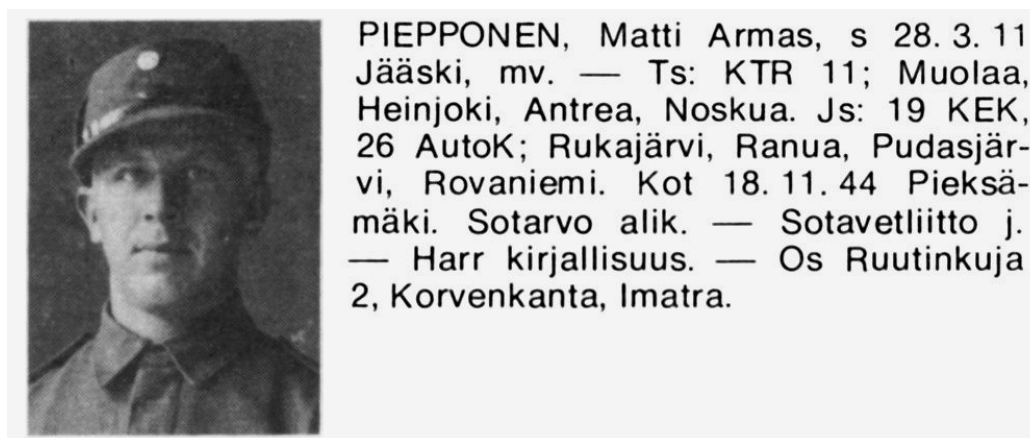
Organisaatiokortiston lisäksi aineistoa on koottu kirjoista kuten Talvisodan historia 1 [39], Talvisodan pikkujättiläinen [40], Suomi Sodassa [2] ja Suomen rintamamiehet 1939–1945 -kirjasarja [1]. Näistä kerättiin listoja joukko-osastoista ja kenraalikunnasta sekä matrikelimaisia henkilökuvauksia. Kuvassa 4.2 on taulukko ilmavoimien osastoista. Taulukossa on sarakkeittain nimi, yläosasto, perustamis- ja keskittämipaikkakunnat.

Ilmavoimien perustaminen ja keskittäminen		
Joukko (vastaava)	Perustamispaikka	Keskittäminen (alistaminen)
Ilmapuolustuksen		
Esikunta	Helsinki	4. 12. Mikkeli (IlmaVar.Os. Järvenpää)
Ilmasotakoulu	Kauhava	Kauhava
Mekanikkokoulu	Santahamina	21. 10. Tampere
Ilmailuvarikko	Tampere	Tampere
Lentorykmentti 1		
Lentoryhmän esik.	Suur-Merijoki	15. 10. Imatra (Kan.A)
Lentolaivue 10	Suur-Merijoki	11. 10. Lappeenranta
Lentolaivue 12	Suur-Merijoki	8. 10. Suur-Merijoki (II AK)
Lentolaivue 14	Suur-Merijoki	11. 10. Ilmee (III AK)
Lentolaivue 16(-)	Pirttipohja	21. 10. Värtsilä (IV AK)
3./Lentolaivue 16	Pirttipohja	10. 10. Jaakkima (LMe.P)
Lentorykmentti 2		
Esikunta	Utti	12. 10. Immola
Lentolaivue 24	Utti	12. 10. Immola
Lentolaivue 26	Utti	8. 10. Utti
Lentorykmentti 4		
Esikunta	Immola	12. 10. Joroinen (25. 11. Luonetjärvi)
Lentolaivue 44	Immola	11. 10. Joroinen (27. 11. Luonetjärvi)
Lentolaivue 46	Immola	11. 10. Luonetjärvi
Lentolaivue 36	Santahamina	14. 10. Kallvik (Me.VE)
Täydennyslentorykmentti 1	Suur-Merijoki	14. 10. Jämijärvi–Karvia
Täydennyslentorykmentti 2	Utti	20. 10. Parola
Täydennyslentorykmentti 4	Immola	14. 10. Luonetjärvi
Täydennyslentolaivue 39	Helsinki–Turku	13. 10. Turku–Naantali
Ilmailukenttävarikko 1	Suur-Merijoki	14. 10. Lappeenranta
Ilmailukenttävarikko 2	Utti	10. 10. Utti
Ilmailukenttävarikko 3	Immola	15. 10. Joroinen (25. 11. Luonetjärvi)
Ilmailukenttävarikko 4	Pirttipohja	12. 10. Värtsilä
Kentänhuoltojoukkueet 10–26, 30–33, 40–43	Lentorykmentit ja -laivueet	lokakuun kuluessa eri lentokentille

Kuva 4.2: Sotilasosastoja kuvaava taulukko [39].

Suomen rintamamiehet 1939–1945 -kirjasarja [1] on julkaissut useita osia, joihin on koottu sodassa palvelleita henkilöitä joukko-osastoittain. Mate-

riaali saatiin digitoituna Kinocompany-tuotantoyhtiöltä³, joka oli tuottanut sen Sotapolku-projektiaan⁴ varten. Aineisto oli taulukkomuotoista ja siitä korjattiin joitakin OCR-prosessin virheitä, kuten numeroja nimikentissä tai väärin tulkittuja skandinaavisia aakkosia. Kuvassa 4.3 on esitetty yhden henkilön kuvauskenttä. Aineiston perusteella luotiin henkilöinstansseja ja niihin liittyviä synnyin-, joukko-osasto- ja sotilasarvotapahtumia. Kuvan kenttä *Ts: KTR 11; Muolaa, Heinjoki, Antrea, Noskua* kertoo talvisodan osaston (Kenttätykistörykmentti 11) sekä listan paikkakuntia. Paikkakuntia ei kuitenkaan muunnettu osaston liikkeiksi, koska tietoon ei sisällynyt aikamääreitä. Kuvausteksti on tallennettu henkilöinstanssin kenttään *crm: P3_has_note* ja sen avulla voidaan henkilöitä myöhemmin liittää mahdollisiin jatkosodan osastoihin.



Kuva 4.3: Esimerkki matrikkelimaisesta henkilökuvauksesta [1].

4.3.3 Menehtyneet-tietokanta

Kansallisarkiston keräämä Menehtyneet-tietokanta on Sotasampo-kokonaisuuden laajin henkilölähde. Siinä on tietoa 94 696 Suomen puolella taistelleesta, jotka ovat kuolleet talvi-, jatko- tai Lapin sodan aikana tai näissä sodissa saamiinsa vammoihin. Tietokantaan voi Sotasampo-sivuston lisäksi perehtyä Kansallisarkiston ylläpitämässä portaalissa⁵.

Tietokanta saatiin käyttöön CSV-muotoisena taulukkona, jossa henkilöresurssia kohti on 30 tietokenttää, joilla esitetään perustiedot ja kuvaus hen-

³<http://www.kinocompany.fi/>

⁴<http://www.sotapolku.fi/>

⁵<http://kronos.narc.fi/menehtyneet/>

kilöön liittyvistä sota-ajan tapahtumista. Sotasampo-tietokantaan menehtymistiedot on tuotu luokan `:DeathRecord` instansseina. Näistä on joko luotu uusia `crm:E21_Person`-henkilöresursseja tai ne on linkitetty tietokannassa jo olleisiin. `DeathRecord` on `crm:E31_Document`-luokan aliluokka eli semanttisesti tulkittuna toimijaresurssia tarkentava dokumentti. Yhteys dokumentin ja henkilön välille on luotu ominaisuudella `crm:P70_documents`. Muunnostyötä on Sotasampo-työryhmässä tehnyt etupäässä Mikko Koho. [36]

Lisäksi Menehtyneet-tietokannasta tuotettiin henkilöitä tarkentavia tapahtumia. Tietokenttien synnyin- ja kuolinajat ja -paikat toimivat lähteinä tapahtumille `crm:E67_Birth` ja `crm:E69_Death` sekä sotilasarvot ja joukko-osastot tapahtumille `:Promotion` ja `:PersonJoining`. Yhdistettäessä tietokantaa Sotasampoon otettiin käyttöön kaksi uutta tapahtumaluokkaa, `:Disappearing` ja `:Wounding`.

Joukko-osastojen tietokenttiä lisättiin vastaamaan menehtyneissä esiintulleita lyhenteitä, esimerkiksi jos on käytetty rinnakkain kirjoitusasultaan eroavia lyhenteitä kuten panssarilaivan *P/L* tai *PsL*. Dataa on tarkastettu hakemalla SPARQL-kyselyllä ne tekstimuotoiset henkilöiden osastokentät, jotka eivät olleet automaattisesti linkittyneet yhteenkään tietokannassa olleeseen osastoon.

4.3.4 Sotapäiväkirjat-kokoelma

Sotapäiväkirja on esitys osaston ja sen miehistön tapahtumista tiettyinä aikana. Päiväkirjaa on pitänyt osastosta tähän määrätty upseeri ja tyypilliset sivut ovat hänen käsin kirjoittamiaan kuvauksia yksikön liikkeistä, oleskelupaikoista ja sotatoimista [33]. Kokoelmaa ylläpitää Kansallisarkisto ja siihen kuuluu 3536 talvisodan sekä 23339 jatko- ja Lapin sodan aikaista päiväkirjaa. Digitoitu aineisto on tutkittavissa Kansallisarkiston portaalisissa⁶, jossa yksittäinen päiväkirja⁷ esitetään sivu kerrallaan kuvatiedostoina⁸. Aineistoa ei ole OCR-prosessoitu, vaikka se tarjoaisikin runsaasti tapahtumainformaatiota.

Runko Sotasampo-projektin osastoille kerättiin päiväkirja-aineistosta. Ote aineistosta on esitetty taulukossa 4.4. Osa hierarkiaa voidaan lukea osastojen nimistä: osastoon *Päämaja* kuuluu alaosastoja kuten *Tiedusteluosasto* tai *Työvoimatoimisto*. Tätä tietoa käytettiin myös luotaessa `:MilitaryUnit`-instansseja ja niiden välisiä suhteita. Yksittäisellä osastolla voi olla useita

⁶http://digi.narc.fi/digi/hae_ay.ka?sartun=319.SARK

⁷<http://digi.narc.fi/digi/slistaus.ka?ay=37285>

⁸<http://digi.narc.fi/digi/view.ka?kuid=1522281>

päiväkirjoja eri ajankohdilta, ks. esim. Rakennuspataljoona 4. Joskus yksi päiväkirja voi käsitellä useampaa osastoa, yleensä pieniä komppanioita tai pataljoonia.

Nimeke	Ajanjakso	digi.narc.fi/digi/
Päämaja. Tiedusteluosasto	30.11.1939–15.3.1940	dosearch.ka?ay=12481.SARK
Päämaja. Työvoimatoimisto	22.1.1940–25.4.1940	dosearch.ka?ay=12482.SARK
Päämaja. Viestikomentaja	6.10.1939–31.3.1940	dosearch.ka?ay=12483.SARK
Päämaja. Viesti 2 (salakirjoitustoimisto)	8.10.1939–30.9.1940	dosearch.ka?ay=12484.SARK
Päämaja. Lääkintäosasto I. Lääkintämajuri Esko Helske	1.9.1943–21.6.1944	dosearch.ka?ay=41435.SARK
5. Radankorjauskomppania	30.11.1939–18.5.1940	dosearch.ka?ay=12485.SARK
6. Radankorjauskomppania	30.11.1939–13.3.1940	dosearch.ka?ay=12486.SARK
...
5. Rajakomppania	30.11.1939–6.12.1939	dosearch.ka?ay=12495.SARK
Rakennuspataljoona 1. 3. komppania	14.1.1940–6.4.1940	dosearch.ka?ay=12496.SARK
Rakennuspataljoona 2. 6. komppania	15.2.1940–18.4.1940	dosearch.ka?ay=12497.SARK
Rakennuspataljoona 2. 8. komppania	2.3.1940–16.4.1940	dosearch.ka?ay=12498.SARK
Rakennuspataljoona 4	21.10.1939–30.11.1939	dosearch.ka?ay=12499.SARK
Rakennuspataljoona 4	30.11.1939–24.1.1940	dosearch.ka?ay=12500.SARK
Rakennuspataljoona 4. Koirinojan työmaa	29.10.1939–10.12.1939	dosearch.ka?ay=12501.SARK
Rakennuspataljoona 4. Komentaja	24.1.1940–3.4.1940	dosearch.ka?ay=12502.SARK
Rakennuspataljoona 4. Komentaja	4.4.1940–14.5.1940	dosearch.ka?ay=12503.SARK
...

Kuva 4.4: Sotapäiväkirjat-kokoelman CSV-taulukkoa.

Työn ensimmäisessä vaiheessa koottiin talvisodan osastot sekä kerättiin jatkosodan osastoista samannimiset, etupäässä laivaston alukset ja rannikkotyökistön asemat. Muutamissa tapauksissa päiväkirja voitiin yhdistää nimettyyn henkilöön. Esimerkkejä tällaisista otsikoista olivat *eversti Simeliuksen valtuuskunta*⁹ tai *Päämajan lääkintämajuri Esko Helske*¹⁰.

Kolmas sarake kertoo päiväkirjan www-osoitteen Arkistolaitoksen sivustolla. Jokaisen taulukkorivin perusteella lisättiin tietokantaan :WarDiary-instanssi, joka luo linkin osastosta sen päiväkirjasivustolle. Kukin instanssi sisältää päiväkirjan nimen, sen ajanjakson ja www-linkin. RDF-muotoinen esimerkki :WarDiary-instanssista on liitteessä 5.

4.3.5 Semanttinen kansallisbiografia

Semanttisen kansallisbiografian (SKB) alustava esittely on aliluvussa 2.2.4. Sotasampo-projektia varten SKB:stä haettiin henkilöitä ja heihin liittyviä tapahtumia. Linkityksiä tietokantojen välillä tehtiin kolmesti. Alussa, kun Sotasammon tietokannassa oli lähinnä vain tärkeimpiä komentajia ja kenraa-

⁹http://www.sotasampo.fi/page?uri=http://ldf.fi/warsa/actors/person_110

¹⁰http://www.sotasampo.fi/page?uri=http://ldf.fi/warsa/actors/person_168

leja, etsittiin SKB:stä henkilöitä vertaamalla nimikenttiä ja suodattamalla tulos elinvuosien perusteella.

Toisella kerralla etsittiin Sotasampoon uusia toimijoita. SPARQL-kyselyllä haettiin henkilöitä, joilla on vähintään yksi vuosille 1930–1945 ajoittuva tapahtuma ja jonka kuvauksessa esiintyy tiettyjä avainsanoja kuten *kenraali*, *amiraali* tai *talvisota*. Haku tuotti tietokantaan runsaasti sotaan liittyviä henkilöitä: Väinö Tanner, Väinö Linna, aikakauden poliitikkoja ja puna-armeijan henkilöitä. Tällainen aineistohaku tuotti myös runsaasti linkkejä henkilöihin, jotka olivat osallistuneet esimerkiksi vapaaehtoisina ja joiden toimintaa sota-aikana ei välttämättä muissa lähteissä mainittaisi kuten poliitikko Erkki Kivimäki¹¹. Aineistosta valittiin ennen vuotta 1929 syntyneet ja vuonna 1939 eläneet. Joukosta karsittiin myös avainsanoista johtuneita virheellisiä tuloksia, kuten Pelastusarmeijan komentajana tai siviilialuksen kapteenina toimineita.

Kolmannella haulla rikastettiin henkilöiden tapahtumatietoja. Henkilöihin liittyviä tapahtumia oli 1 544, joista 730 oli ylennyksiä ja 244 kunniamerkin osoituksia. Muut tapahtumat olivat synnyin- tai kuolintapahtumia sekä liittymisiä tehtäviin jossakin joukko-osastossa. Sekä Sotasampo että SKB ovat CIDOC CRM -perustaisia ja käyttävät aika- ja paikkamääreille samoja `crm:E53.Place`- ja `crm:E52.Time-Span`-luokkia. SKB:n määreet harmonisoitiin vastaamaan Sotasammossa käytettyjä.

4.3.6 Wikipedia

Wikipedia tarjoaa mahdollisuuden hakea helposti uusia toimijoita. Aineistoa sotatoimiin liittyvistä henkilöistä on runsaasti, mutta joukko-osastoista huomattavasti vähemmän. Vaikka Wikipediaa ei pidetä kaikkein luotettavimpana tietolähteenä, saadaan sen kautta mahdollisuus sillata henkilöresurssia ja linkitetyn avoimen datan tietokantoihin kuten Wikidataan [60], DBpediaan [16] ja VIAF:iin [4].

Taulukossa 4.2 on henkilöiden lähteinä käytettyjä Wikipedia-sivuja, joista Sotasammon tietokantaan kerättiin noin 2500 henkilöä. Taulukon kohta 1. sisältää noin sata tärkeintä valtionpäämiestä tai sotatoimien komentajaa, kuten Churchill tai Stalin¹², joita linkitettiin poliittisiin tapahtumiin Euroopassa ja maailmalla. Kun toimijoihin poimittiin henkilöitä kohdan 2. luettelosta, suodatettiin mukaan vain vuosien 1939–1945 aikana eläneet henkilöt. Useat kohdan 3. tai 4. sotilaista olivat talvi- ja jatkosodan aikana Suomen

¹¹http://www.sotasampo.fi/page?uri=http://ldf.fi/warsa/actors/person_2163

¹²http://www.sotasampo.fi/page?uri=http://ldf.fi/warsa/actors/person_2354

1. Luokka: Toisen maailmansodan henkilöt
2. Luettelo suomalaisista upseereista
 - (a) Luokka: Suomalaiset kenraalit
 - (b) Luokka: Suomalaiset amiraalit
 - (c) Luokka: Suomalaiset everstit
3. Luettelo jääkäreistä
4. Sisällissodan valkoiset
5. Suomalaiset toisessa maailmansodassa kaatuneet
6. Tiedotuskomppaniat
7. Kaukopartiomiehet
8. Luokka: Ruotsalaiset vapaaehtoisotilaat
9. Yksiköiden sivut, kuten lentäjät sivulla Lentolaivueen 32 ilmataisteluvuodot

Taulukko 4.2: Esimerkkejä henkilöiden lähteinä käytetyistä Wikipedia-sivuista.

armeijan merkittäviä upseereita ja osastojen komentajia. Saksassa koulutettiin vuosina 1915–18 kaikkiaan 1 895 suomalaista jääkäriä, joista 774 osallistui talvisotaan [62]. Kohdasta 5. saatiin täydentäviä tekstikuvauksia myös Menehtyneet-tietokannassa olleille henkilöille, kuten esimerkiksi taidemaalari ja graafikko Jorma Gallen-Kallelalle¹³. Useat tiedotuskomppanioiden TK-kuvaajat (kohta 6.) olivat siviilissäkin tunnettuja kulttuurialan ihmisiä, kuten Olavi Paavolainen¹⁴. Lisäksi monia listan henkilöitä on mainittu myös SA-kuva-arkiston kuvaajina (ks. myös 4.3.8), esimerkiksi Björn Soldan¹⁵ ja Heikki Aho.

Ruotsalaisia vapaaehtoisia (kohta 8.) on Sotasammon tietokannassa 15 henkilöä¹⁶, joista monet olivat toimineet Lentorykmentti 19:ssä¹⁷. Wikipedia-sivut ilmavoimien taistelulentäjistä (kohta 9.) olivat formaatiltaan hyvin erilaisia kuin yksittäinen henkilösivu. Sivulla on taulukko, jossa on sarakkeittain ilmataistelun päivämäärä, kellonaika, paikka, koneen ohjaaja sekä käytetyn koneen ja alasammutun koneen malli. Koneen ohjaajan kentässä on mainittu lentäjän sen hetkinen sotilasarvo sekä etu- ja sukunimi.

Henkilöresurssien generointiin käytettiin Pythonille kirjoitettua Wikipedia-API:a [21]. Ohjelmallisesti sivua käsitellään luokan `WikipediaPage` instanssina. Instanssin metodi `url` palauttaa Wikipedia-sivun täyden osoitteen ja

¹³http://www.sotasampo.fi/page?uri=http://ldf.fi/warsa/actors/person_p500751

¹⁴http://www.sotasampo.fi/page?uri=http://ldf.fi/warsa/actors/person_2271

¹⁵http://www.sotasampo.fi/page?uri=http://ldf.fi/warsa/actors/person_2351

¹⁶<http://yasgui.org/short/SkWkAS96>

¹⁷http://www.sotasampo.fi/page?uri=http://ldf.fi/warsa/actors/actor_457

metodi `title` sivun otsikon, henkilön tapauksessa koko nimen. Metodilla `summary` saadaan sivun lyhyt kuvaus, joka on sama kuin sivun ensimmäinen tai muutama ensimmäisistä kappaleista. Tämä kuvaus on suoraan otettu RDF-graafin `dc:description`-tekstikentän arvoksi. Kielitunniste on sama kuin Wikipedia-sivun kieli eli useimmissa tapauksissa suomi. Työn edetessä haluttiin lisätä henkilöille englanninkielisen vastinsivun kuvauksia.

Henkilökuvauksen ensimmäinen lause noudattaa hyvin säännönmukaisesti samaa muotoilua. Esimerkiksi kuvauksessa ”*Aarne Nopsanen (17. helmikuuta 1907 Lahti – 23. syyskuuta 1990 Lahti) oli suomalainen taidemaalari, taidegraafikko ...*” on henkilön koko nimen perässä suluissa mahdollinen aiempi sukunimi, synnyinaika ja -paikka sekä kuolinaika ja -paikka. Päivämäärän ja paikan tietokentät noudattavat aina samaa muotoilua, jossa päivän lopussa on piste, kuukauden nimi on tekstimuodossa, vuosi neljän numeron tarkkuudella ja paikan nimi on perusmuodossaan. Säännönmukainen tietokenttä on helppo tunnistaa ja erotella säännöllisellä lausekkeella.



Kuva 4.5: Wikipedia-sivun ylennykset ja kunniamerkit -kenttä. Lähde: https://fi.wikipedia.org/wiki/Pehr-Axel_Wahren

Henkilöä täydentävää tapahtumatietoa haettiin ylennykset ja kunniamerkit -kentästä (ks. kuva 4.5). Ylennyksiä varten poimittiin säännöllisellä lausekkeella tekstiosuus, jossa sotilasarvoa seuraa aikamääre päivän tai vuoden tarkkuudella.

Sotasammon tietokannassa on hiukan yli sata Wikipediaan linkittyvää joukko-osastoa. Taulukossa 4.3 on esitetty esimerkkejä osastoihin liittyvistä Wikipedia-sivuista. Osastojen Wikipedia-sivujen rakenne ole yhtä standardoitu kuin henkilöiden eikä perustietoja voi lukea yhtä helposti. Wikipedia-sivuilta ei luotu ainuttakaan uutta yksikköä, vaan kaikista oli jo jonkin muun lähteen kautta saatu maininta. Osastoille haettiin `dc:description`-tekstejä suomeksi ja englanniksi.

1. Talvisodan suomalaiset maavoimat
2. Talvisodan suomalaiset sota-alukset
3. Suomen ilmavoimien joukko-osastot ja yksiköt toisessa maailmansodassa
4. Talvisodan vapaaehtoiset
5. Talvisodan suomalaiset ilmatorjuntayksiköt

Taulukko 4.3: Esimerkkejä osastojen lähteinä käytetyistä Wikipedia-sivuista

4.3.7 Mannerheim-ristin ritarit

Mannerheim-ristin ritareiksi nimitettiin 191 henkilöä vuosina 1941–1945. Kunnianosoitus sai alkunsa, kun ylipäällikkö Mannerheim halusi talvisodan jälkeen uudistaa aiempia, sotilasarvosta riippuneita mitalikäytäntöjä ja palkita tasarvoisemmin taisteluissa taitojaan ja urhoollisuuttaan osoittaneita sotilasarvosta riippumatta. Mannerheim-ristin ritariksi voitiin lyödä niin sotamiehiä kuin kenraalejakin, eikä vastaavaa järjestelmää ollut käytössä missään muualla kuin Suomessa. [42]

Sotasampo-projektin rikastaminen Mannerheim-ristin ritareilla (MRR) tehtiin kahdessa vaiheessa. Ensimmäisessä vaiheessa luotiin henkilöinstanssit tai linkitettiin ne jo olemassa oleviin toimijoihin. Lähtökohtana käytettiin MRR-sivuston¹⁸ ja Wikipedian listoja¹⁹ ritareista. Listasta saadaan henkilön nimi, linkit MRR- ja Wikipedia-sivulle (ks. myös 4.3.6), myöntämispäivä, sotilasarvo myöntämishetkellä ja ritarin numero. Numero on välillä 1–191 ja se ilmaistiin henkilöresurssiin ominaisuuden `:knight` arvona. Henkilö on linkitetty ominaisuudella `foaf:page` häntä kuvaavalle MRR-sivuston sivulle (ks. kuva 4.6), josta on henkilön elinkaaritietojen lisäksi luettavissa osastot sota-vuosina, ylennykset päivämäärineen ja myönnetyt mitalit.

Henkilöresurssija haluttiin rikastaa vielä tapahtumatiedolla. Prosessi tietokenttien poimimiseen MRR-sivuilta oli hyvin samankaltainen kuin Wikipedian kohdalla kuvattu, paitsi että talvisota- ja jatkosotakenttien kuvaukset, ylennyskentän sotilasarvot ja kunniamitalit sekä niitä vastaavat päivämäärät haettiin Python-ohjelmalla sivun lähdekoodista. Ohjelma tuotti RDF-kolmikoita tilapäisillä `:hasEvent-` ja `:hasMedal-` predikaateilla, joiden pohjalta luotiin varsinaiset tapahtumatresurssit. Poiminta esimerkkihenkilö Oiva Tuomisen datasta on esitetty taulukossa 4.4. Huomaa, että joukossa on tietokenttiä kuten *mittarilentokurssi 1934*, joista ei Sotasammon toimijoiden nykyisen skeeman perusteella generoida niitä vastaavia tapahtumia.

¹⁸<http://www.mannerheim-ristinritarit.fi/ritarit>

¹⁹https://fi.wikipedia.org/wiki/Mannerheim-risti#Luettelo_ritareista

```

:person_218 :hasEvent "alikersantti 1929-04-01".
:person_218 :hasEvent "kuoli 1976-01-28 Nummelassa".
:person_218 :hasEvent "lentomestari 1941-07-23".
:person_218 :hasEvent "mittarilentokurssi 1934".
:person_218 :hasEvent "mekaanikkokurssi 1927".
:person_218 :hasEvent "syntyi 1908-03-05 Iitissä".
:person_218 :hasEvent "vääpeli 1940-04-25".

:person_218 :hasMedal "MR 2 1941-07-18".
:person_218 :hasMedal "2 VR 4 mk".
:person_218 :hasMedal "Js mm".
:person_218 :hasMedal "SaksRR 2".
:person_218 :hasMedal "SVR M 2".
:person_218 :hasMedal "Ts mm".
:person_218 :hasMedal "VR 3 mk".

```

Taulukko 4.4: Esimerkkihenkilö Oiva Tuomisen MRR-sivulta tuotettuja RDF-kolmikkoja.

Kaikki 191 Mannerheim-ristin ritaria ovat Sotasammon tietokannassa. Heidät voidaan tunnistaa merkityn tietolähteen²⁰ tai `:knight`-numerokentän perusteella. Koska ritarit ovat sotahistoriassa tunnettuja henkilöitä, tiedetään heidän sotilasarvonsa ja joukko-osastonsa varsin tarkasti, ja heihin on usein linkittyynyt runsaasti tietoa muista lähteistä, kuten Semanttisesta kansallisbiografiasta tai SA-kuva-arkistosta.

4.3.8 SA-kuva-arkisto

SA-kuva-arkisto on Puolustusvoimien Kuvakeskuksen tuottama ja Anygraaf Oy:n toteuttama palvelu. Aineisto kattaa talvi-, jatko- ja Lapin sotien ajalta noin 160 000 valokuvaa sekä alle sata elokuvakatkelmua. Aineisto on aiemmin julkaistu omassa portaalissaan (<http://sa-kuva.fi>). [49] Tässä alivussa käsitellään henkilöiden luomista kuvamateriaalissa mainittuista valokuvaajista ja kuvausteksteissä esitetyistä henkilöistä. Materiaalin muuntaminen osaksi Sotasampo-tietokantaa sekä kuvien siltaaminen tietokannan muihin osiin on ollut työryhmässä Erkki Heinon työtä [25].

Sotasammon tietokantaan kuva-aineisto on tuotu `:Photograph`-luokan instanssina. Tietokenttinä resursseissa on kuvan aihe, kuvaajan nimi, kuvauspaikka, päivämäärä sekä linkit kuvatiedostoihin sa-kuva.fi-palvelussa. Valokuvaajat on ilmoitettu `:photographer_string`-kentässä. Tietokannasta tuo-

²⁰<http://ldf.fi/warsa/sources/source5>

tettiin lista kuvaajien nimistä ja kullekin merkittyjen kuvien lukumäärästä. Kirjoitusasuissa oli runsaasti eri käytäntöjä viitata samaan henkilöön (ks. taulukko 4.5). Kuvien lukumäärää käytettiin havaitsemaan yksittäiset kirjoitusvirheet, ratkaisemaan ristiriidat esimerkiksi taulukossa nimien *P.E. Janson* vai *P.R. Jansson* kohdalla sekä yhdistämään nimikirjaimet kokonaiseen etunimeen. Jos datassa oli useita viittauksia tiettyyn, mutta vain muutama viittaus hiukan poikkeavaan kirjoitusasuun, tulkittiin tämä siten, että kyseessä on sama henkilö. Tietysti on mahdollista, että kuvaajina olisi ollut sekä *P.E. Janson* että *P.R. Jansson*, mutta koska käytettävissä ei ollut tarkempaa lisätietoa, on päätelty nimien viittaavan samaan henkilöön.

Valokuvaajia lisättiin 575 henkilöä, joiden joukossa on tunnettuja TK-kuvaajia kuten Nils Björn Soldan tai hänen velipuolensa Heikki Aho. SA-kuvamateriaalista tuotettiin henkilöitä myös kuvaustekstien perusteella. Niistä haettiin kaikki titteli-Etunimi-Sukunimi -muotoiset katkelmat, koska haluttiin poimia selvästi koko nimellä ja mieluiten sotilasarvollakin tunnistettavia henkilöitä.

Sotilasarvot	Etun.	Sukunimi	viittaukset	lukum.	esimerkkikuva
Vänrikki	Arvi V.	Janhunen	Vänrikki Arvi V. Janhunen	12	sakuva_50654
Sotilasvirkailija	Per Erik	Janson	Sot.virk. P.E.Janson; Sot.virk. P.E. Janson; P.E.Janson; P.E. Janson; Sot.virk. P. E. Janson; P. E. Janson	234	sakuva_102586
Sotilasvirkailija	Per Erik	Janson	Sot.virk. P.E.Jansson; Sot.virk. P. E. Jansson; Sot.virk. P.E. Jansson; P.E. Jansson	210	sakuva_11220
Sotilasvirkailija	Per Erik	Janson	Janson; Sot.virk. Janson; Sot.virk. Janson.	171	sakuva_101411
Sotilasvirkailija	Per Erik	Janson	Sot.virk. P.R.Jansson; Sot.virk. P.R. Jansson	33	sakuva_6280
Sotilasvirkailija	Per Erik	Janson	Per Erik Jansson	24	sakuva_57149
Sotilasvirkailija	Per Erik	Janson	TK -frontofotograf Jansson	10	sakuva_39009
Sotilasvirkailija	Per Erik	Janson	P.Janson	6	sakuva_102429
Everstiluutnantti	J. B.	Jauhiainen	Everstiluutnantti J.B.Jauhiainen	88	sakuva_151651
Sotilasvirkailija	Lars Börje	Johnsson	Sot.virk. L.Johnsson; Sot.virk. L. Johnsson; Sot.virk L. Johnsson; L. Johnsson; L.Johnsson; Sot.virk.L.Johnsson; Sot.virk L.Johnsson	874	sakuva_100418
Sotilasvirkailija	Lars Börje	Johnsson	Sot.virk. Johnsson; Johnsson	195	sakuva_151614
Sotilasvirkailija	Lars Börje	Johnsson	Sot.virk. Lars Johnsson	112	sakuva_10621
Sotilasvirkailija	Lars Börje	Johnsson	L. Johansen	94	sakuva_86360
Sotilasvirkailija	Lars Börje	Johnsson	Johnsson L.; Johnsson, L.	30	sakuva_78342
Sotilasvirkailija	Lars Börje	Johnsson	Johnsson, Lars	17	sakuva_75882
Sotilasvirkailija	Lars Börje	Johnsson	Sot.virk. L.Johansson	15	sakuva_99103
Sotilasvirkailija	Lars Börje	Johnsson	Johansson	11	sakuva_83127
Sotilasvirkailija	Lars Börje	Johnsson	Sot.virk. Lars Börje Johansson	10	sakuva_127166
Sotilasvirkailija	Lars Börje	Johnsson	Sot.virk. Lars B. Johnsson	10	sakuva_11891
Sotilasvirkailija	Lars Börje	Johnsson	L. Jonsson	8	sakuva_52702
Sotilasvirkailija	Lars Börje	Johnsson	Sot.virk. L.Johursin	6	sakuva_43957
Sotilasvirkailija	Lars Börje	Johnsson	Johnson. L.	3	sakuva_78352
Sotilasvirkailija	Lars Börje	Johnsson	Johnsson, L.B.	3	sakuva_87459
Sotilasvirkailija	Armas	Jokinen	Sot.virk. A.Jokinen;Sot.virk. A. Jokinen	23	sakuva_16427
Sotilasvirkailija	Armas	Jokinen	Armas Jokinen	6	sakuva_87917
Sotilasvirkailija	Armas	Jokinen	Sot.virk. Alma Jokinen	6	sakuva_156797
Sotilasvirkailija	Walter	Jokinen	Walter Jokinen; Sot.virk. Walter Jokinen	464	sakuva_10145
Sotilasvirkailija	Walter	Jokinen	Sot.virk. Waltter Jokinen; Waltter Jokinen	64	sakuva_145605
Sotilasvirkailija	Walter	Jokinen	Sot.virk. W.Jokinen; Sot.virk. W. Jokinen	7	sakuva_103358
Sotilasvirkailija	Walter	Jokinen	V. Jokinen	2	sakuva_11644
		Jokinen	Jokinen	10	sakuva_75821
		Jokinen	TK-Jokinen	1	sakuva_139527

Taulukko 4.5: Lista SA-kuva-arkiston kuvaajista.

4.3.9 Sotilasarvot

Sotilasarvograafiin²¹ kuuluu noin 200 sotilasarvoa tai käytettyä titteliä. Graafi sisältää myös hierarkian ja muodostaa puurakenteen alokkaasta Suomen marsalkkaan. Graafiin on koottu yksittäisten arvojen lisäksi sotilasarvoluokkia kuten upseerit, aliupseerit ja miehistö sekä erityisluokkia kuten lääkintä- tai kirkollinen henkilöstö.

²¹<http://ldf.fi/warsa/actors/ranks>

Alkuperäisenä tietolähteenä käytettiin Wikipedian luettelon [61] lisäksi menehtyneiden tietokannan sotilasarvoja²². Tämä sanasto sisälsi käytettyjä nimikkeitä ja arvoluokkia. Joissakin harvoissa tapauksissa on käytössä ollut useampia kirjoitusasuja kuten *Eläinlääkintämajuri* tai *Eläinlääketieteen majuri*. Lyhenteitä koottaessa käytettiin Puolustusvoimien lyhennelistaa [35].

Työn edetessä sanastoa täydennettiin esimerkiksi jääkäreiden, Lotta Svärdin, Saksan ja Neuvostoliiton yleisimmillä arvoilla. Suomenkielisessä tekstissä käytetään jääkäreiden tai Saksan armeijan arvoista niiden alkuperäiskielisiä nimiä kuten *Zugführer* tai *Hauptmann*. Koska nämä arvot kuitenkin halutaan näyttää dataa selaavalle tällaisina, ei niitä merkitty kielitunnisteella. Puna-armeija käytti toukokuuhun 1940 saakka järjestelmää, jossa sotilasarvo kuvaa henkilön komentaman yksikön kokoa kuten *Armeijakunnankomentaja* (Комкор) tai *Divisioonankomentaja* (Комдив)²³. Jatkosodan alkaessa oli puna-armeijassa siirrytty käyttämään Suomen armeijaa vastaavaa järjestelmää kenraalikunnan osalta.

Sotilasarvoja linkitettiin Wikipedia-sivuille, joista haettiin 157 suomen- ja 87 englanninkielistä kuvausta `dc:description`-kenttiin. Linkitys tehtiin hakemalla Python-kielen Wikipedia-API:lla²⁴ nimikettä vastaava suomenkielinen Wikipedia-sivu. Liitteessä 4 on esimerkki sotilasarvomääritelmästä RDF-muodossa.

Sotilasarvon menetystä tai sen palauttamista kuvaavia luokkia tai instansseja ei ole luotu, mikä olisi tarpeen joissakin yksittäistapauksissa kuten Lauri Allan Törnin²⁵ kohdalla. Ylennysten materiaalia ei ole validoitu. Käytännössä tämä tarkoittaisi yksittäisen henkilön sotilasarvoketjun aukottomuuden ja aikaleimojen peräkkäisyyden tarkistamista.

²²<http://demo.seco.tkk.fi/saha/project/index.shtml?type=http://ldf.fi/schema/narc-menehtyneet1939-45/Sotilasarvo&model=narc-menehtyneet1939-45>

²³https://fi.wikipedia.org/wiki/Kenraali#Talvisodan_aikaisia_puna-armeijan_arvoja

²⁴<https://pypi.python.org/pypi/wikipedia/>

²⁵http://www.sotasampo.fi/fi/persons/?uri=http://ldf.fi/warsa/actors/person_342

TUOMINEN Oiva Emil Kalervo

6



Lentomestari, taksiautoilija.

Syntyi 5.3.1908 Iitissä, kuoli 28.1.1976 Nummelassa, haudattu Helsinkiin Malmin hautausmaalle.

Ritari 6, 18.8.1941, lentomestari, ohjaaja 1./LLv 26. Vanhemmat Ida Maria Tuominen. Puoliso 1) Saimi Musto ero 1953, 2) Ingeborg Ström ero 1958, 3) Sanny Louise Hakulinen ero 1959, 4) Oili Lassila k 1996. Lapset Leo Kalervo (1932) mainospiirtäjä k 2012.

4/4 kansakoulu, varusmies IlmAK ja MLE 1926–27, reservimekaanikkokurssi 1927, MekK mekaanikkokurssi 1930, IlmK aliupseeriohjaajakurssi 3 1933–34, mittarilentokurssi 1934, laskuvarjokurssi 1935, TK jalkaväkikurssi 1936–37.

Lentokonemekaanikko ja ohjaaja-aliupseeri LAs 5 1933–38, LLv 26 1938–39 ja LLv 24 1939, ohjaaja-aliupseeri HLeLv 34 1944–45, evp 6.1.1945.

Talvisota: ohjaaja-aliupseeri LLv 24 1939–40 ja LLv 26 1940, Karjalan kannas, Laatokan Karjala. Jatkosota: ohjaaja-aliupseeri LLv 26 ja LeLv 26 1941–43, LeLv 34 ja HLeLv 34 1943–44, komennettu Saksaan 1943 (Messerschmitt-hankinta), Itä-Karjala, Karjalan kannas, Suomenlahti. Saavutti 44 ilmavoittoa, joista 8 talvisodassa.

Työnjohtaja Wärtsilä Oy Helsingin telakka, lentäjänä S P J Keinäsen, Kaleva Salmisen ja Imatran Voima Oy:n palveluksessa, taksiautoilija Helsingissä.

Mannerheim-ristin ritarien säätiön valtuuskunnan varajäsen 1960–76.

Alikersantti 1.4.1929, vääpeli 25.4.1940, lentomestari 23.7.1941. MR 2, VR 3 mk, 2 VR 4 mk, SVR M 2, Ts mm, Js mm, SaksRR 2.



"Ylipääällikkö on pvm:llä 18.8.41 nimittänyt vääpeli Oiva Emil Kalervo Tuomisen, joka erittäin menestyksellisenä hävittäjälentäjänä on ampunut alas kahdeksantoista viholliskonetta, Vapaudenristin 2. luokan Mannerheim-ristin ritariksi."

Kuva 4.6: Esimerkki Mannerheim-ristin ritarin sivusta [42]

Luku 5

Sovellukset

Tässä luvussa esitellään Warsa-datajulkaisu LDF.fi-palvelussa ja Sotasampo-portaalin¹ aineistonäkymiä.

5.1 Warsa-datajulkaisu

Sotasammon tietokanta on kuuden tähden linkitetyn avoimen datan julkaisu ja saatavilla Linked Data Finland -alustalla [26]. Tietojoukoista on esittely sivulla <http://www.ldf.fi/dataset/warsa> ja josta ne ovat ladattavissa RDF-muotoisina. Tietojoukkojen graafit on lueteltu taulukossa 5.1.

URI-tunnisteet ovat muotoa http://ldf.fi/warsa/actors/person_N tai http://ldf.fi/warsa/actors/actor_N (tunniste päättyy juoksevaan numeroon). Tunnisteita voidaan käyttää tietokantakyseluihin tai HTTP-osoitteina. Ohjelmallisen käytön tietokantakyselyjä varten on SPARQL-palvelupiste (<http://ldf.fi/warsa/sparql>). Tunnisteen käyttö HTTP-osoitteena ohjautuu SAHA-editoriin [47], jolloin tietokantaa voidaan tutkia verkkoselaimella. Jos tietokantaa käytetään ohjelmallisesti, voidaan HTTP Accept -otsakkeessa määritellä haluttu muoto kuten esimerkiksi "text/turtle", jolloin HTTP-osoite palauttaa pyydetyn URI:n RDF-kuvauksen.

¹<http://www.sotasampo.fi>

Tietojoukko	Graafi
Toimijat	http://ldf.fi/warsa/actors
Skeema	http://ldf.fi/warsa/actors/actor_types
Sotilasarvot	http://ldf.fi/warsa/actors/ranks
Mitalit	http://ldf.fi/warsa/medals
Päiväkirjat	http://ldf.fi/warsa/diaries

Taulukko 5.1: Tietojoukkojen graafit LDF.fi-palvelussa.

5.2 Sotasampo-portaali

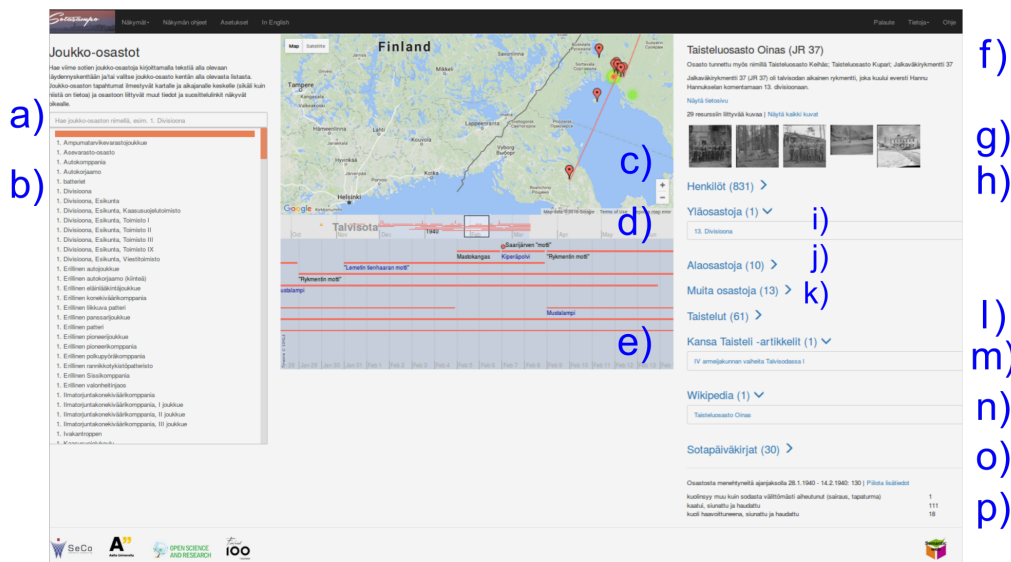
Sotasampo-portaali visualisoi Warsa-tietokannan materiaalin ja tuo sen selaimella tutkittavaan muotoon. Sivusto on tarkoitettu Suomen sotien 1939–1945 aineistoista kiinnostuneille. Oletuskäyttötapauksessa etsitään tietoa sodassa palvelleesta sukulaisesta ja hänen joukko-osastonsa vaiheista. Sotasampo-portaalin etusivulta voidaan valita aineistonäkymäksi sodan tapahtumat, henkilöt, joukko-osastot, paikat, Kansa Taisteli -artikkelit, Menehtyneet-tietokanta tai SA-kuva-aineisto. Aineistoja on linkitetty runsaasti toisiinsa, mikä tekee siirtymisen aineistonäkymien välillä helpoksi.

Oma osuuteni oli joukko-osasto- ja henkilönäkymien sekä näihin liittyvien sotilasarvo- ja kunniamitalisivujen rakentaminen. Näiden käyttöliittymät ovat HTML5- ja JavaScript-tekniikoille pohjautuvia Rich Internet Application -toteutuksia. Sivujen teknisenä perustana toimi Erkki Heinon tapahtumasivu- ja varten rakentama malli, jossa sisällön dynaamisesta päivittymisestä huolehtii AngularJS-sovelluskehys [22] ja responsiivisuudesta Bootstrap-alusta [55]. Oman työni keskeinen osuus oli laatia palvelupisteeseen kohdistettavia tietokantakyselyjä, toteuttaa saatujen vastausten käsittely ja rakentaa niiden perusteella visuaalinen esitys.

Sivut on toteutettu suomen- ja englanninkielisinä². Kielivalinta on sivun yläpalkissa ja sillä vaihdetaan sivun staattiset otsikot ja ohjetekstit halutulle kielelle. Tietokannan aineistossa on englanninkielisiä kuvaustekstejä henkilöille, osastoille ja sotilasarvoille niiltä osin kun lähteenä käytetylle Wikipedia-sivulle on englanninkielinen vastinsivu.

²<http://www.sotasampo.fi/en/>

5.2.1 Hakunäkymä joukko-osastoille



Kuva 5.1: Sotasampo-sivuston joukko-osastonäkymä.

Joukko-osastonäkymä³ on esitetty kuvassa 5.1. Sivulla vasemmalla on hakusanakenttä (a), josta osastoja voidaan hakea nimellä tai käytetyllä lyhenteellä. Hakutulokset näkyvät valikossa (b), josta valitaan tarkasteltava osasto. Informaation määrä voi vaihdella paljon eri osastojen sivuilla, koska joistakin saatetaan tuntea yksityiskohtainen tapahtumaketju ja tarkka miehistörakenne, toisista vain nimi ja lyhenne.

Kartta (c) toimii yhdessä aikajanan (d, e) kanssa. Kartalla näytetään osaston perustamiset, siirrot ja taistelut tarkasteltavana ajanjaksona. Lämpökartalla visualisoidaan osastosta surmansa saaneiden lukumäärää alueittain ja siinä kiivaimmat taistelupaikat erottuvat punaisina. Aikajana on vierityspalkki, jonka yläosassa (d) käytetään kuukauden intervalleja, alaosassa (e) päivän tarkkuutta. Osa tapahtumajanojen otsikoista on linkkejä erillisille tapahtumasivuille, jotka tarjoavat yksityiskohtaisempaa tietoa.

Ylimpänä oikealla (f) on osaston nimi ja lyhenne, muita käytettyjä nimiä ja tekstimuotoinen kuvaus. Osastoon linkittyneitä SA-kuva-arkiston materiaalia (g) voidaan myös selata modaalisena kuvagalleriana, johon päästään klikkaamalla yksittäistä kuvaa. Alapuolella on linkkejä osastossa palvelleisiin henkilöihin (h). Hierarkiasta esitetään ylempi (i), alempi (j) sekä samalla tasolla olevia (k) osastoja. Tämän alla on lista taisteluista (l), joihin osaston

³ http://sotasampo.fi/fi/units?uri=http://ldf.fi/warsa/actors/actor_940

tiedetään osallistuneen sekä osastoon liittyvät Kansa Taisteli -artikkelit (m), Wikipedia-sivu (n) ja sotapäiväkirjat (o). Alimpana (p) on lisätietoa osaston menehtyneistä. Näiden lukumäärä lasketaan aikajanan (d,e) valinnan mukaisena aikana.

Sivujen tietosisältö haetaan kohdistamalla SPARQL-kysely Warsa-palvelupisteeseen, jonne laadittu kysely lähetetään enkoodattuna URL-parametrinä. Palvelupisteen palauttama JSON-muotoinen vastaus muunnetaan JavaScript-tietotyypeiksi ennen esittämistä käyttöliittymän komponenteissa. Kysely saattaa hakea tietokannasta kaikki osastoon liittyvät tapahtumat, jotka lajitellaan ja esitetään tapahtumaluokkiensa mukaisissa tietokentissä.

Toimija-tapahtuma-rakennetta hyödyntävä SPARQL-kysely on liitteessä 7. Kysely hakee osaston ja kaikkien sen alaosastojen miehistöjen tietyn ajanjakson sisälle sijoittuvien kuolemisien, haavoittumisien ja katoamisten aika- ja paikkamääreet. Kysely on samankaltainen kuin se, jonka tulokset näytetään osastosivun lämpökartalla. Kysely hakee aluksi osastoon ja kaikkiin sen alaosastoihin kuuluneet henkilöt. Näistä suodatetaan ne, joilla on aikajanan jaksolle ajoittuva kuolin-, katoamis- tai haavoittumistapahtuma, josta lisäksi tunnetaan tapahtumapaikka. Lopuksi palautetaan kunkin tapahtuman aika, paikkakoordinaatit ja henkilöresurssi.

SPARQL-kyselyjä voidaan tiivistää käyttämällä property path -rakenteita⁴. Osastoon liittyvät henkilöt haetaan rakenteella `?unit (^crm:P144_joined_with/crm:P143_joined)+ ?person`. Aluksi `^crm:P144_joined_with` poimii kaikki tapahtumat, joista osastoon `?unit` viitataan liityttävänä osastona, kauttaviivan jälkeinen `crm:P143_joined` kaikki liittyneet henkilöt ja alaosastot. Muuttujista ei siis viitata varsinaiseen tapahtumaresurssiin. Lauseke on ympäröity suluilla, joiden perässä on plusmerkki, mikä tarkoittaa ehdon toistoa, eli sama haku kohdistetaan edelleen alaosastoihin, joista uusilla toistoilla voidaan iteroida läpi koko hierarkian.

5.2.2 Henkilöhakunäkymä

Kuvassa 5.2 on Sotasampo-sivuston henkilöhakunäkymä⁵. Hakukenttä (a) ja tulosvalikko (b) ovat samankaltaiset kuin osastosivulla. Henkilöitä haetaan suku- tai etunimen perusteella. Ylimpänä esitetään henkilön nimi ja mahdollinen muutaman lauseen mittainen tekstikuvaus (c). Alapuolella on lista (d), jossa kerrotaan synnyin- ja kuolinajat ja -paikat sekä tunnettu sotila-

⁴<https://www.w3.org/TR/sparql11-property-paths/>

⁵http://www.sotasampo.fi/fi/persons/?uri=http://ldf.fi/warsa/actors/person_50

sarvo tai mahdollinen lista ylennyksistä. Esimerkkihenkilöstä on kuvamateriaalia (e) SA-kuva-arkistossa. Kohdassa (f) on Kansallisbiografian leijuke. Avattavista listakentistä voi siirtyä henkilöön liittyviin tapahtumiin (g), joukko-osastoihin (h), taisteluihin (i), sotilasarvoihin (j), kunniamerkkeihin (k), Wikipedia-sivulle (l) tai Kansa Taisteli -artikkeleihin (m).

Valokuvien yläpuolelle sijoitetusta Näytä tietosivu -linkistä siirrytään henkilötietosivulle⁶, joka on hakuvalikko lukuunottamatta samanlainen kuin henkilöhakunäkymä. Henkilötietosivua on käytetty leijukkeena Kansa Taisteli -aineistonäkymässä. Sivun yläpalkissa (kohdan b yläpuolella) on painike käyttäjäpalautteelle. Suuri osa saadusta palautteesta on koskenut sodassa olleita henkilöitä ja niissä on annettu tarkennuksia tai ehdotettu korjauksia tietokannan henkilötietoihin.

⁶http://www.sotasampo.fi/page?uri=http://ldf.fi/warsa/actors/person_50

The screenshot shows the Sotasampo website profile for Paavo Juho Talvela. On the left, there is a search bar and a list of names (a) through (m). The main profile area contains the following information:

- Henkilöt:** Paavo Juho Talvela (a) (13. tammikuuta 1897 Helsinki – 30. syyskuuta 1973 Helsinki) oli suomalainen upseeri, joka oli Saksassa jääkärijoukkoon. Hän osallistui sisällissotaan, heimosotiin sekä talvi- ja jatkosotaan. Hänestä tuli armeijakunnan komentaja sekä monakkia Mannerheimin luottamus, joka nautui Saksan korkeimman johdon kanssa. Hän toimi liikemiehenä 1930-luvulla sekä jatkosodan jälkeen muun muassa paperiteollisuuden palveluksessa.
- 13.1.1897** Helsinki msk. – **30.9.1973** Helsinki
- Eversti** 15.6.1928
- Kennaalajuri** 18.12.1939
- Kennaalhuhmari** 16.1.1942
- Jätkäaikakansalli** 6.12.1966
- Lähde:** Suomi Sodassa, ISBN 9519078943, Valtut Palaut 1963.
- Näytä tiedosto**
- 167** kuvaa | **Näytä kaikki kuvat**
- Kansallisbiografia**
- Talvela, Paavo**
- Syntynyt 1897 paikkassa Helsinki. Kuollut 1973 paikkassa Helsinki.
- armeijakunnankomentaja**
- Kennali Paavo Talvelaa on sanottu Suomen korkein arvoisimmaksi reservinupseeriksi. Tälle leikkimieliselle sanonnalle on katetta sikäli, että Talvela erosi meitä kertaa armeijan vakinaisesta palveluksesta joko osallistukseen vapaaehtoisena heimosotiin tai toimikseen liike-elämän palveluksessa. Talvelalla oli kuitenkin keskeinen tehtävä talvi- ja jatkosodan aikaisena sotatoimintayhdyksen komentajana sekä ylipäätään edustajana Saksan sotavotimen pääesikunnassa. Hän löytyi meidän tavoin myös Suomen itsenäisyyden ajan historian näin oli luopuutakseen organisaattori, Akhonorilleen joutonentävä ja Pehtamon tilanteen (jopa)ää.**
- Lähde:** Suomenlinen Kansallisbiografia | SHS:n Kansallisbiografia
- Tapahtumat (30)** >
- Joukko-osastot (2)** v
- Osasto Hovet, Ahenanmaan Ryhmä, Ryhmä Talvela, 14. Divisiona, Osasto Talvela
- B.Armeijakunta
- Taistelut (5)** v
- Talvejäni
- Äijäjävi
- Vietoksi
- Ronanti
- Altoppi
- Sotilasarvot (8)** >
- Kunniamerkit ja -mitalit (17)** >
- Wikipedia (1)** v
- Paavo Juho Talvela
- Kansa Taisteli -artikkelit (3)** v
- Tuhtoinen aita Vanginveits
- Vihollisen Nihpaukkoakelu Altoppela
- Vilvoitahalla maholisuutta

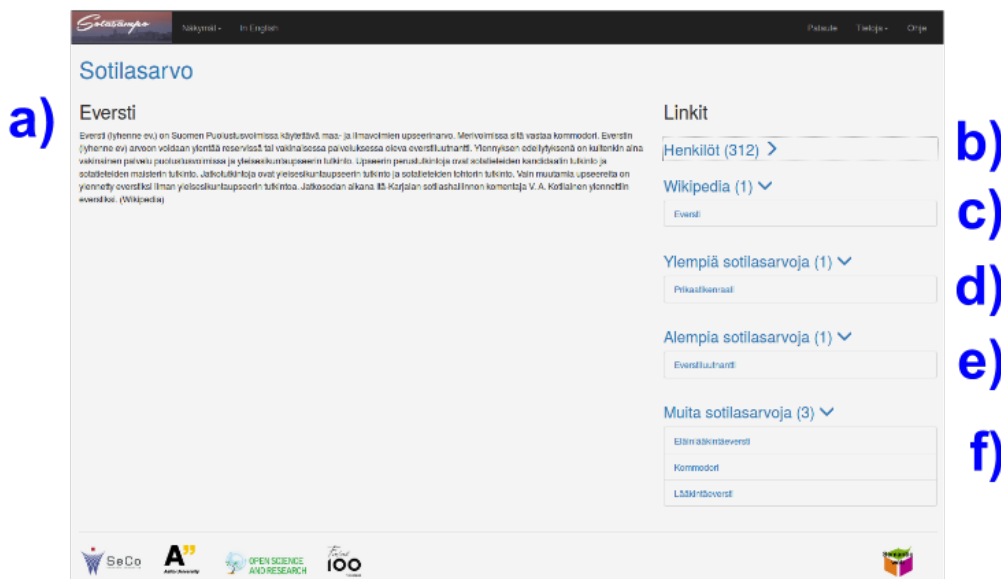
Kuva 5.2: Sotasammon henkilöhakunäkymä.

5.2.3 Sotilasarvonäkymä

Sotilasarvonäkymä⁷ on esitetty kuvassa 5.3. Sotilasarvosivuille ei ole toteutettu hakutoimintoa, vaan sivulle saavutaan henkilönäkymän kautta tai toi-

⁷Esimerkiksi <http://www.sotasampo.fi/page?uri=http://ldf.fi/warsa/actors/ranks/Eversti>

selta sotilasarvosivulta. Tarkasteltava sotilasarvo ja sen tekstikuvaus näkyvät sivun vasemmassa yläosassa (a). Oikealla esitetään lista sotilasarvoon liittyvistä henkilöistä (b) ja sen alapuolella linkki mahdolliselle Wikipedia-sivulle (c). Sivulta voi siirtyä hierarkiassa ylemmän (d) tai alemman (e) sotilasarvon sivuille. Sivulla viimeisenä (f) on linkkejä muiden aselajien, toisissa maissa käytettyihin tai erikoisupseereiden vastaaviin sotilasarvoihin.



Kuva 5.3: Sotilasarvonäkymä.

5.2.4 Mitali- ja kunniamerkkinäkymä

Mitali- ja kunniamerkkinäkymä⁸ on esitetty kuvassa 5.4. Hakutoimintoa kunniamerkeille ei toteutettu, vaan sivulle saavutaan henkilönäkymän kunniamerkkivalikosta tai toisen kunniamerkin sivulta. Kunniamerkin nimi lukee sivun vasemmassa yläosassa (a), oikealla on lista kunniamerkin saaneista henkilöistä (b) sekä sen alapuolella linkkilista muiden kunniamerkkien sivuille (c). Lista generoidaan SPARQL-kyselyllä, joka hakee tietokannasta kaikki henkilöt, joille kyseinen mitali on myönnetty, sekä edelleen muut mitalit, joita nämä henkilöt ovat saaneet.

⁸Esimerkiksi http://www.sotasampo.fi/page?uri=http://ldf.fi/warsa/medals/medal_83

a) **Kunniamerkki**
Mannerheim-ristin ritari 2. lk.

Linkit

Henkilöt (191) ▾

Martti Johannes Aho
Arvo Ahola
Johan Aarne Einar Ahola
Aksel Fredrik Airo
Olavi Eelis Alakulppi

1 2 3 4 5

Muita kunniamerkkejä (35) ▾

Vapaudenristi IV miekkojen kera
Talvisodan muistomitali
Jatkosodan muistomitali
Vapaudenristi 3. lk. miekkojen kera
Suomen Reservupseeriliiton kultainen ansiomitali

1 2 3 4 5

b)

c)

Kuva 5.4: Mitali- ja kunniamerkkinäkymä.

Luku 6

Yhteenveto

Tässä luvussa vastataan johdannossa esitettyihin tutkimuskysymyksiin, sekä esitetään jatkokehitysideoita Sotasampo-projektia varten.

6.1 Tulosten arviointi

Ensimmäinen tutkimuskysymys on ”Miten semanttisessa webissä kuvataan yksittäistä henkilöä?”. Semanttisessa webissä henkilöresurssi yksilöidään sen URI-tunnisteella ja laaditussa ratkaisussa resurssiin kytketään vain pysyväksi määritelty tieto. Henkilöresurssissa esitetään tietoa, johon ei liity ajallista muutosta tai joka voidaan olettaa mallin käyttötarkoituksen kannalta pysyväksi, kuten sukupuoli tai kansallisuus. Semanttisen webin kannalta pysyviä kenttiä ovat tiivistetyt kuvaustekstit ja linkit vastaaviin resursseihin muissa tietokannoissa.

Henkilöresurssia rikastetaan tapahtumilla. Ontologiassa tapahtumiksi tulkitaan kaikki sellaiset muutokset, jotka jakavat aikajatkumon vaiheeseen ennen tapahtumaa ja sen jälkeiseen. Tapahtuma on semanttisesti määritelty yhdysilta, joka yhdistää eri luokkien resursseja. Esimerkiksi synnyintapahtuma on skeemassa määritelty ’olemassaolon alkamiseksi’ ja graafissa se yhdistää henkilöresursseja synnyinhetken aikamääreisiin ja paikkaresursseihin. Henkilöön liitetään tarkentavaa dokumentaatiota kuten valokuvia ja lehtiartikkeleja.

Toinen tutkimuskysymys on ”Miten kuvataan henkilöryhmää ja ryhmien välistä hierarkiaa?”. Ryhmän elinkaari sijoittuu sen perustamisen ja lopettamisen välille. Ryhmän liikettä kuvataan pistemäisinä tapahtumina aika- ja paikkamääreineen. Ryhmän mallin on kuvattava dynaamisempia muutoksia kuin henkilön, koska kaikkein tärkeimmätkin tunnisteet, kuten nimi ja

lyhenne, saattavat muuttua ajan funktiona. Ryhmien kokonaishierarkiaa kuvaa graafi, jossa liittymistapahtumat kiinnittävät ryhmiä toisiinsa.

Kolmas tutkimuskysymys on ”Millainen malli soveltuu mahdollisimman joustavasti eri lähteiden tiedon yhdistämiseen, tietokenttien harmonisointiin ja linkittämiseen?”. Toimija-tapahtuma-malli sallii eri lähteistä kerätyn tiedon yhdistämisen ja joustaa tapauksissa, joissa datan määrä vaihtelee paljon toimijaresurssia kohti. Dataa on yksiselitteistetty aina liitettäessä uudesta tietolähteestä tuotettua informaatiota. Toimija-tapahtuma-mallissa samoiksi tunnistettujen toimijoiden yhdistäminen vaatii myös tapahtumadatan käsittelyä, joka johtaa tietomäärän kasvuun ja tietokenttien tarkentumiseen.

Toimijoihin liittyy runsaasti ajan funktiona muuttuvaa tietoa kuten henkilöiden sotilasarvot tai osastojen nimet. Toimijoihin on linkitetty runsaasti valokuva- ja artikkelimateriaalia. Jotta materiaali on voitu kohdistaa oikeaan toimijaan, on mallista pitänyt pystyä esittämään toimijan status tietyllä ajanhetkellä.

Esimerkkinä tiedon tarkentumisesta iteratiivisen prosessin tuloksena voidaan tarkastella organisaatiokortissa (kuva 4.1) mainittua komentajaa. Eversti K.A. Heiskanen on tarkentunut Kaarlo Aleksanteri Heiskaseksi¹, johon on yhdistynyt muista lähteistä Kansallisbiografian ja Wikipedian kuvaukset, Kansas Taisteli -artikkeli, synnyin- ja kuolintapahtumat, 10 ylennystä päivämäärineen, 16 kunniamerkkiä ja 81 valokuvaa.

Neljäs tutkimuskysymys on ”Miten malli palvelee loppukäyttötarkoitusta kuten visualisointia verkkosivustolla tai biografista esittämistä historiantutkimusta varten?”. Toimija-tapahtuma-malli soveltuu hyvin kartta- ja aikajana-tyyppisen visualisoinnin perustaksi. Linkitetystä materiaalista muodostuvaa verkostoa voidaan analysoida etsimällä yhteisten tapahtumien ja dokumenttien välityksellä toisiinsa kytkeytyviä toimijoita.

6.2 Kehitettävää

Tässä aliluvussa tutkitaan laaditun toimijaontologian puutteita sekä esitetään ehdotuksia kootun toimijatietokannan ja Sotasampo-portaalin jatkokehitykselle.

¹http://www.sotasampo.fi/page?uri=http://ldf.fi/warsa/actors/person_72

6.2.1 Toimijaontologia

Toimijaontologiaa voidaan laajentaa uusilla tapahtumatyypeillä tai lisäämällä malliin semanttista sisältöä. Henkilöresurssille on kolme luokkaa `crm:E21_Person`, `:MilitaryPerson` ja `:PoliticalPerson`. Kuitenkaan henkilöiden jakaminen toimintansa perusteella ei ole yksiselitteistä, joten henkilömallia voisi yksinkertaistaa käyttämällä vain `crm:E21_Person`-luokkaa ja lisäämällä luokittelevia tarkenteita `crm:P2_has_type`-aliominaisuuksina.

Joukko-osastojen nimet toteutettiin `skos:prefLabel`- ja `skos:altLabel`-kenttinä osasto- tai tapahtumaresurssissa. CIDOC CRM -skeema käyttää tähän `crm:E41_Appellation`-nimikkeitä, joiden käyttöönotto yksinkertaisesti mallia ja tietokantakyselyjä tiettyä osastoa etsittäessä. Semanttista sisältöä voi lisätä tyypittämällä nykyisiä joukko-osastoja, toisin sanoen luokitella ne `crm:P2_has_type`-ominaisuudella esimerkiksi divisiooniksi, rykmenteiksi tai pataljooniksi. Attribuutti kertoisi osaston aseman hierarkiassa ja sisältäisi oletuksen miesvahvuudesta², sekä edesauttaisi osastohierarkian validointia. Vastaavalla periaatteella voitaisiin osastoja tyypittää talvi-, jatko- tai Lapin sodassa toimineisiin.

6.2.2 Toimijatietokanta

Tietokantaan täydennetään myöhemmin jatko- ja Lapin sodan aikaiset osastot. Lähteinä voidaan käyttää sotapäiväkirjojen aineistoa [34] ja Pentti Kopsan kokoamaa peitenumerolistaa [37]. Aineistoa voidaan niin ikään täydentää puna-armeijan osastoilla ja keskeisimmillä henkilöillä. Tiedon perustana voidaan käyttää Ohto Mannisen dokumenttia puna-armeijan joukoista [43]. Jotta tietokanta palvelisi osana kansainvälistä semanttista tietoverkkoa, tulisi uudet toimijat linkittää suuriin LOD-tietokantoihin kuten DBpediaan, Wikidataan ja VIAF:iin.

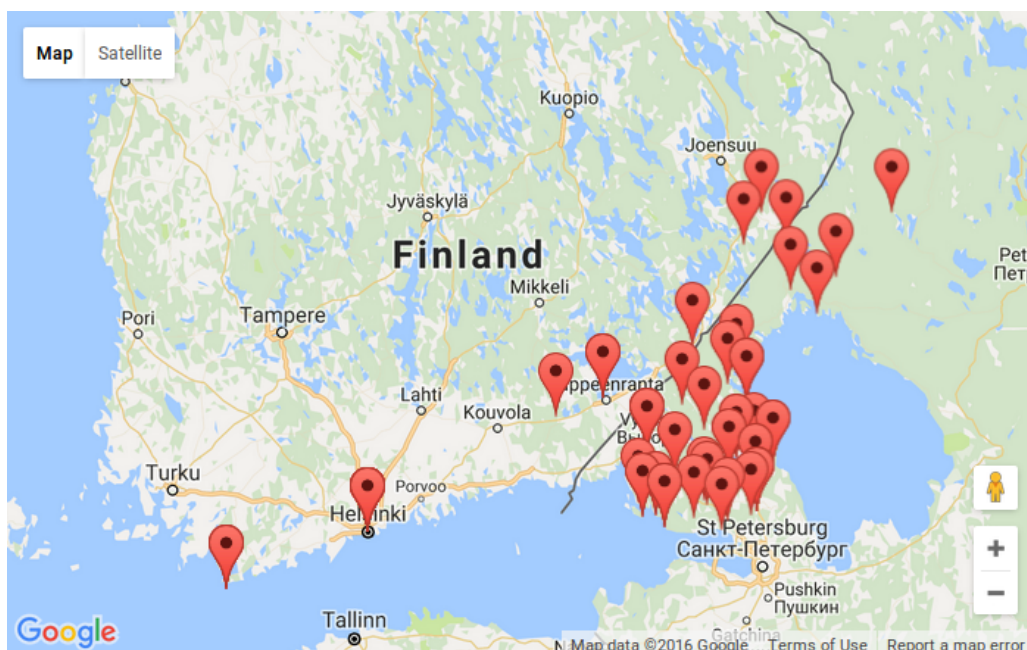
6.2.3 Sotasampo-portaalin jatkokehitys

Portaalia voidaan täydentää henkilön aikajananäkymällä (aliluku 5.2.2). Tapahtumien lisäksi henkilön liikkeitä voidaan seurata häneen linkittyneen kuvamateriaalin päivämäärien ja paikkakuntien avulla. SPARQL-haku³ marsalkka Mannerheimin olinpaikoista tuottaa 72 tulosta, jotka on esitetty kuvassa 6.1. Vastaava lisäys voidaan tehdä myös osastonäkymään (aliluku 5.2.1).

²Esimerkiksi https://fi.wikipedia.org/wiki/Jalkaväkipataljoonan_määrävahvuus

³<https://goo.gl/n5h5aa>

Tämänhetkisessä versiossa henkilöhaaku toimii pelkästään nimen perusteella, vaikka olisi käyttökelpoisempaa, jos hakuun voisi käyttää termejä kuten esimerkiksi *jääkäri* tai *kapteeni Karhunen*. Henkilönäkymästä ei ole mahdollisuutta siirtyä suoraan toisen henkilön sivulle. Se voitaisiin tuottaa hakemalla henkilöverkosta esimerkiksi yhteisten valokuvien, sotilasarvojen tai joukko-osastojen kautta toisiinsa kytkeytyviä henkilöitä.



Kuva 6.1: Mannerheimin olinpaikkoja valokuvamateriaalin perusteella.

6.3 Kontribuutio

Toimijaontologialuokkien perustana toimii historiallisen tiedon tallentamisen kehitetty CIDOC CRM -standardi. Standardi tarjoaa selkeän kehyksen toimija-tapahtuma-mallille; henkilöille, ryhmille sekä näitä tarkentaville tapahtumille. Toisaalta tapahtumamallia käytettiin vain aihepiiriin keskeisesti kuuluvien asioiden mallintamiseen. CIDOC CRM -standardiin perustuva Semanttinen kansallisbiografia ja toisaalta Open Memory Project ja Menehtyneet-tietokanta esittävät historiaa yksittäisen henkilön kautta. Semanttisen kansallisbiografian ja Menehtyneet-tietokannan kaikkia tietoja ei muunnettu osaksi Sotasammon toimijaontologiaa, koska ne voidaan hakea alkuperäisestä kannasta federoidulla SPARQL-kyselyllä. Erillään pitämistä voidaan myös perustella mahdollisella tarpeella päivittää tietokantoja. Mu-

ninn Military Ontology tarjosi esimerkin sotilastermistön semanttiselle esittämiselle. Tässä työssä laaditulle henkilöiden ja ryhmien ontologialle ei ollut suoraa esikuvaa, vaan se koottiin yhdistämällä periaatteita useammasta eri tarkoitusta palvelleesta ratkaisusta.

Aivan samoin on Sotasampo-projekti saattanut yhteen sotavuosien 1939–1945 aineistoa Suomen osalta. Toisesta maailmansodasta on valtavasti tietoa ympäri maailman; eri maissa, eri kielillä, eri formaateissa ja usein eriävistä näkökulmistakin julkaistuna. Tiedon kerääminen ja yhdenmukaistaminen semanttisesti jäsennellyksi avoimen datan tietojulkaisuksi on valtava työkenttä, mutta vähitellen toteutuessaan siitä voisi syntyä globaali toisen maailmansodan kattava tietokanta. Sotahistorian tapahtumissa ovat yksittäiset henkilöt ja heistä muodostetut ryhmät keskeisellä sijalla. Tässä työssä esitetty toimijaontologiamalli ei ehkä sellaisenaan sovellu kaikkiin käyttötarkoituksiin, mutta sen toivotaan toimivan esimerkkinä ihmisten ja ryhmien toiminnan kuvaamisesta ja inspiroivan tutkijoita kehittämään ratkaisua eteenpäin.

Lähteet

- [1] *Suomen rintamamiehet 1939–1945 -kirjasarja*. Etelä-Suomen kustannus, 1972.
- [2] *Suomi Sodassa - Talvi- ja jatkosodan tärkeät päivät*. Valitut Palat, 2. painos, 1983. ISBN 9519078940.
- [3] Beckett, D., Berners-Lee, T., Prud'hommeaux, E. ja Carothers, G. RDF 1.1 Turtle, Terse RDF Triple Language, W3C Recommendation 25 February 2014. <https://www.w3.org/TR/turtle/> viitattu 29.9.2016.
- [4] Bennett R., Hengel-Dittrich C., O'Neill E.T. ja Tillett B.B. VIAF (Virtual International Authority File): Linking die Deutsche Bibliothek and Library of Congress name authority files. In *World library and information congress: 72nd IFLA general conference and council*, 2006.
- [5] Berners-Lee, T. Design issues: Linked data (2006). <http://www.w3.org/DesignIssues/LinkedData.html>, viitattu 16.06.2016.
- [6] Biografiakeskus, Suomalaisen Kirjallisuuden Seura. Suomen kansallisbiografia -kirjasarja. <http://www.kansallisbiografia.fi/kirjat/?p=2>, viitattu 20.10.2016.
- [7] Brazzo, L. ja Mazzini, S. From the Holocaust Victims Names to the Description of the Persecution of the European Jews in Nazi Years: the Linked Data Approach and a New Domain Ontology. http://www.bygle.net/wp-content/uploads/2015/04/Open-Memory-Project_3-1.pdf, viitattu 29.9.2016.
- [8] Brickley, D. ja Guha, R. RDF Schema 1.1, W3C Recommendation 25 February 2014. <https://www.w3.org/TR/rdf-schema/>, viitattu 25.10.2016.
- [9] Brickley, D. ja Miller, L. FOAF vocabulary specification 0.98. *Namespace document*, 9, 2012. <http://xmlns.com/foaf/spec/>, viitattu 29.9.2016.

- [10] Canadiana. Out of the Trenches. <http://www.canadiana.ca/en/pcdhn-lod/>, viitattu 1.11.2016.
- [11] Centro di documentazione ebraica contemporanea. Shoah Vocabulary Specification. <http://dati.cdec.it/lod/shoah/reference-document.html>, viitattu 27.10.2016.
- [12] Cyganiak, R., Wood, D., Lanthaler, M., Klyne, G., Carroll, J. J. ja McBride, B. RDF 1.1 Concepts and Abstract Syntax. W3C Recommendation, 2014. <http://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/>, viitattu 25.10.2016.
- [13] Davis, I. Linked Data for Ian Davis. <http://iandavis.com/id/me.html>, viitattu 30.10.2016.
- [14] Davis, I. ja Galbraith, D. Bio: A vocabulary for biographical information. 2003. <http://vocab.org/bio/>, viitattu 19.5.2016.
- [15] Davis I. ja Vitiello, E. Jr. RELATIONSHIP: A vocabulary for describing relationships between people. *RDF Vocabulary Specification*, 2005. <http://vocab.org/relationship/>, viitattu 29.9.2016.
- [16] DBpedia Association. DBpedia. <http://dbpedia.org/>, viitattu 30.10.2016.
- [17] DCMI Usage Board. DCMI Metadata Terms, 2012. <http://dublincore.org/documents/dcmi-terms/>, viitattu 25.10.2016.
- [18] Doerr, M. The CIDOC conceptual reference module: an ontological approach to semantic interoperability of metadata, 24(3): sivut 75–92. *AI magazine*, 2003.
- [19] Friedrich-Meinecke-Institut (FMI), Center for Digital Systems (CeDiS) ja Bavarian State Library (BSB). 1914–1918 Online, International Encyclopedia of the First World War. <http://www.1914-1918-online.net>, viitattu 1.11.2016.
- [20] Gartner, R. ja Hedges, M. CENDARI: establishing a digital ecosystem for historical research. In *2013 7th IEEE International Conference on Digital Ecosystems and Technologies (DEST)*, sivut 61–65. IEEE, 2013.
- [21] Goldsmith, J. Wikipedia API 1.4.0 for Python, 2016. <https://wikipedia.readthedocs.org/en/latest/>, viitattu 29.9.2016.
- [22] Google. AngularJS. <https://angularjs.org/>, viitattu 14.04.2016.

- [23] Guarino, N., Oberle, D. ja Staab, S. What Is an Ontology? Handbook on Ontologies, International Handbooks on Information Systems, 2. painos, sivut 1–17, 2009.
- [24] Hyvönen, E., Alonen, M., Ikkala, E. ja Mäkelä, E. Life stories as event-based linked data: Case semantic national biography. In *Proceedings of ISWC 2014 Posters & Demonstrations Track*. CEUR Workshop Proceedings, October 2014. <http://ceur-ws.org/Vol-1272/>, viitattu 24.8.2016.
- [25] Hyvönen, E., Heino, E., Leskinen, P., Ikkala, E., Koho, M., Tamper M., Tuominen, J. ja 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)*, sivut 758–773. Springer-Verlag, 2016.
- [26] Hyvönen, E., Tuominen, J., Alonen, M. ja 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*, pages 226–230, 2014.
- [27] Hyvönen, E., Tuominen, J., Mäkelä, E., Heino, E., Leskinen, P., Ikkala, E., Koho, M. ja Tamper, M. Sotasampo, talvi- ja jatkosota semanttisessa webissä. <http://seco.cs.aalto.fi/projects/sotasampo/sotasampo-esittely-2015-11-07.pdf>, viitattu 14.10.2016.
- [28] ICT Policy Support Programme (ICT PSP). Europeana Collections 1914–1918. <http://www.europeana-collections-1914-1918.eu>, viitattu 1.11.2016.
- [29] Ikkala, E. Suomalainen historiallisten paikkojen ja karttojen ontologiapalvelu. Master's thesis, Aalto University, School of Electrical Engineering, Degree Programme of Automation and Systems Technology, August 2016.
- [30] ImageMagick Studio LLC 1999-2016. ImageMagick. <http://www.imagemagick.org/script/index.php>, viitattu 13.10.2016.
- [31] Kansallisarkisto. Talvisodan organisaatiokortisto, digitoitu aineisto. [toimitettu sähköisenä].
- [32] Kansallisarkisto. Senaatin kartastot. http://wiki.narc.fi/portti/index.php/Senaatin_kartastot, viitattu 14.10.2016.

- [33] Kansallisarkisto. Sotapäiväkirjat, 2016. <http://wiki.narc.fi/portti/index.php/Sotap%C3%A4iv%C3%A4kirjat>, viitattu 24.8.2016.
- [34] Kansallisarkisto. Sotapäiväkirjat-kokoelma, 2016. <http://digi.narc.fi/digi/dosearch.ka?atun=65.SARK>, viitattu 24.8.2016.
- [35] Kansallisarkisto. Sotilaslyhenteet, 2016. <http://www.arkisto.fi/fi/aineistot/arkistolaitoksen-aineistot/aineistot-ja-kokoelmat/puolustushallinnon-aineistot-5/sotilaslyhenteet>, viitattu 24.8.2016.
- [36] Koho, M., Hyvönen, E., Heino, E., Tuominen, J., Leskinen, P. ja Mäkelä, E. Linked death - representing, publishing, and using second world war death records as linked open data. In Harald Sack, Giuseppe Rizzo, Nadine Steinmetz, Dunja Mladeníć, Sören Auer, and Christoph Lange, editors, *The Semantic Web: ESWC 2016 Satellite Events, sivut 3–14*. Springer-Verlag, 2016.
- [37] Kopsa, P., Arkistolaitos. Puolustusvoimain joukot 1941-1945 peitelukui-
na, 2016. <http://www.arkisto.fi/uploads/Aineistot/kopsa%5B1%5D.pdf>, viitattu 7.10.2016.
- [38] Kurki, J. Toimijaontologiat ja niiden käyttö semanttisessa webissä. Pro gradu, Helsingin yliopisto, 2011.
- [39] Sotatieteen laitos. *Talvisodan historia 1*. WSOY, Suomi, 2. painos, 1978.
- [40] Leskinen, J. ja Juutilainen, A. *Jatkosodan pikkujättiläinen*. WSOY, Suomi, 2005.
- [41] Maanmittauslaitos. Karjalan kartat. <http://www.maanmittauslaitos.fi/aineistot-ja-palvelut/palvelut/karjalan-kartat>, viitattu 14.10.2016.
- [42] Mannerheim-ristin ritarien säätiö. Mannerheim-ristin ritarit, 2016. <http://mannerheim-ristinritarit.fi/ritarit>, viitattu 24.8.2016.
- [43] Manninen, Ohto. Neuvostojoukot Suomen rintamalla 1939—1940. <http://www.helsinki.fi/aluejakulttuurintutkimus/tutkimus/II.5.OhtoManninen.Joukkojenvahvuudet.pdf>, viitattu 14.04.2016.
- [44] Martini R. G., Araújo C., Librelotto G.R. ja Henriques P.R. *A Reduced CRM-Compatible Form Ontology for the Virtual Emigration Museum*. Springer International Publishing, Cham, 2016. ISBN 978-3-319-31232-3.

- [45] McGuinness, D. L. ja van Harmelen, F. OWL Web Ontology Language Overview. W3C Recommendation, 2004. <http://www.w3.org/TR/2004/REC-owl-features-20040210/>, viitattu 30.10.2016.
- [46] Miles, A. ja Bechhofer, S. SKOS Simple Knowledge Organization System Reference. <https://www.w3.org/TR/2009/REC-skos-reference-20090818/>, viitattu 25.10.2016.
- [47] Mäkelä, E. ja Hyvönen, E. SPARQL SAHA, a Configurable Linked Data Editor and Browser as a Service. In *Proceedings of the ESWC 2014 demonstration track*, Springer-Verlag, May 2014.
- [48] Mäkelä, E., Törnroos, J., Lindquist, T. ja Hyvönen, E. WW1LOD: an application of CIDOC-CRM to World War 1 linked data, International Journal on Digital Libraries, sivut 1–11, 2016. ISSN 1432-1300. <http://dx.doi.org/10.1007/s00799-016-0186-2>.
- [49] Puolustusvoimien Kuvakeskus. Anygraaf Oy. SA-kuva-arkisto, etulinjasta kotirintamalle 1939–1945, 2016. <http://sa-kuva.fi>, viitattu 24.8.2016.
- [50] Savolainen E., Finn Lectura. Suomen kielen äänne-, muoto- ja lauseoppia: 3.7.3 Appositio- eli substantiiviattribuutti. <https://fl.finnlectura.fi/verkkosuomi/Syntaksi/sivu373.htm>, viitattu 30.10.2016.
- [51] Smith, R. An overview of the Tesseract OCR engine. 2007. <https://research.google.com/pubs/pub33418.html>, viitattu 13.10.2016.
- [52] Stevenson, A., Baylis, L. ja Palmer, J. WW1 Discovery. <http://ww1.discovery.ac.uk>, viitattu 1.11.2016.
- [53] Suomen Sotahistoriallinen Seura ry, Hallitus. Kansa Taisteli -lehdet 1957—1986, 2014. <http://www.sshs.fi/siteneews/view/-/nid/92/ngid/1>, viitattu 24.8.2016.
- [54] Svensson, P. The Landscape of Digital Humanities. Digital Humanities Quarterly, 2010. <http://digitalhumanities.org/dhq/vol1/4/1/000080/000080.html>, viitattu 29.9.2016.
- [55] Bootstrap Core Team:. Bootstrap. <http://getbootstrap.com/>, viitattu 14.11.2016.
- [56] The Getty Foundation. Getty Vocabulary Program ontology. <http://vocab.getty.edu/ontology>, viitattu 14.04.2016.

- [57] W3C SPARQL Working Group. SPARQL 1.1 Overview, W3C Recommendation 21 March 2013. <http://www.w3.org/TR/sparql11-overview/>, viitattu 30.10.2016.
- [58] Warren, R. Creating specialized ontologies using Wikipedia: The Muninn experience. *Berlin: Proceedings of Wikipedia Academy: Research and Free Knowledge (WPAC2012)*, 2012. https://wikipedia-academy.de/w/images/wikipedia-academy-2012/0f/21_Paper_Robert_Warren.pdf, viitattu 28.11.2016.
- [59] Warren, R. Military ontology specification - 1.1, 2015. <http://rdf.muninn-project.org/ontologies/military.html>, viitattu 29.9.2016.
- [60] Wikimedia Foundation. Wikidata. <https://www.wikidata.org/>, viitattu 30.10.2016.
- [61] Wikipedia. Sotilasarvot — Wikipedia, 2016. https://fi.wikipedia.org/w/index.php?title=Sotilasarvot_Suomen_puolustusvoimissa&oldid=16027553, viitattu 24.8.2016.
- [62] Wikipedia. Jääkäriliike — Wikipedia, 2016. <https://fi.wikipedia.org/w/index.php?title=J%C3%A4k%C3%A4k%C3%A4riliike&oldid=15907120>, viitattu 24.8.2016.

Liitteet

Liite 1: Toimijaontologian skeema (RDF)

Toimijaontologiaan sisältyy seuraavat ominaisuudet ja luokat. Sanasto on esitetty Turtle-muodossa [3].

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix xml: <http://www.w3.org/XML/1998/namespace> .
@prefix xmls: <http://www.w3.org/2001/XMLSchema#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix sch: <http://schema.org/> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix crm: <http://www.cidoc-crm.org/cidoc-crm/> .
@prefix dct: <http://purl.org/dc/terms/> .
@prefix dctype: <http://purl.org/dc/dcmitype/> .
@prefix mil: <http://rdf.muninn-project.org/ontologies/
  organization#> .

@prefix : <http://ldf.fi/schema/warsa/> .
@prefix wso: <http://ldf.fi/warsa/sources/> .
@prefix wev: <http://ldf.fi/warsa/events/> .
@prefix wra: <http://ldf.fi/warsa/actors/ranks/> .
@prefix wac: <http://ldf.fi/warsa/actors/> .
@prefix wti: <http://ldf.fi/warsa/events/times/> .
@prefix wdi: <http://ldf.fi/warsa/diaries/> .
@prefix wme: <http://ldf.fi/warsa/medals/> .

## CIDOC CRM

crm:E5_Event skos:prefLabel "Tapahtuma"@fi .

crm:E7_Activity skos:prefLabel "Toiminta"@fi .
```

```
crm:E9_Move skos:prefLabel "Liike"@fi .
crm:E21_Person skos:prefLabel "Henkilö"@fi .
crm:E24_Physical_Man-Made_Thing skos:prefLabel "Esine"@fi .
crm:E31_Document skos:prefLabel "Dokumentti"@fi .
crm:E39_Actor skos:prefLabel "Toimija"@fi .
crm:E55_Type skos:prefLabel "Tyyppi"@fi .
crm:E66_Formation skos:prefLabel "Joukko-osaston
    muodostaminen"@fi .
crm:E67_Birth skos:prefLabel "Syntymä"@fi .
crm:E68_Dissolution skos:prefLabel "Joukko-osaston
    lopettaminen"@fi .
crm:E69_Death skos:prefLabel "Kuolema"@fi .
crm:E74_Group skos:prefLabel "Ryhmä"@fi .
```

Properties

```
crm:P100_was_death_of a rdf:Property ;
    rdfs:subPropertyOf crm:P11_had_participant ;
    rdfs:domain crm:E69_Death ;
    skos:prefLabel "kuollut"@fi ;
    rdfs:range crm:E21_Person .

crm:P98_brought_into_life a rdf:Property ;
    rdfs:subPropertyOf crm:P11_had_participant ;
    rdfs:domain crm:E67_Birth ;
    skos:prefLabel "syntynyt"@fi ;
    rdfs:range crm:E21_Person .

crm:P143_joined a rdf:Property ;
    rdfs:range crm:E39_Actor ;
    skos:prefLabel "liittyjä"@fi ;
    rdfs:subPropertyOf crm:P11_had_participant .

crm:P144_joined_with a rdf:Property ;
    skos:prefLabel "liittynyt osastoon"@fi ;
    rdfs:range crm:E74_Group .

crm:P95_has_formed a rdf:Property ;
```

```
skos:prefLabel "muodostettu"@fi ;
rdfs:subPropertyOf crm:P11_had_participant ;
rdfs:range crm:E74_Group .

crm:P3_has_note skos:prefLabel "huomautus"@fi .

crm:P4_has_time-span skos:prefLabel "tapahtuma-aika"@fi .

crm:P7_took_place_at skos:prefLabel "tapahtumapaikka"@fi .

crm:P70_documents skos:prefLabel "dokumentoi"@fi .

crm:P70i_is_documented_in skos:prefLabel "dokumentoitu"@fi ;
owl:inverseOf crm:P70_documents .

crm:P96_by_mother skos:prefLabel "äiti"@fi .

crm:P97_from_father skos:prefLabel "isä"@fi .

crm:P107_1_kind_of_member skos:prefLabel "henkilön rooli"@fi
.

crm:P141_assigned skos:prefLabel "myönnettiin"@fi .

dct:description skos:prefLabel "kuvaus"@fi .

dct:source skos:prefLabel "lähde"@fi .

## ACTORS

: MilitaryPerson a rdfs:Class ;
skos:prefLabel "Military Person"@en, "Sotilashenkilö"@fi
;
rdfs:subClassOf crm:E21_Person .

: PoliticalPerson a rdfs:Class ;
skos:prefLabel "Political Person"@en, "Poliittinen henkil
ö"@fi ;
rdfs:subClassOf crm:E21_Person .

: MilitaryUnit a rdfs:Class ;
skos:prefLabel "Sotilasyksikkö"@fi, "Military Unit"@en ;
rdfs:subClassOf crm:E74_Group .

## EVENTS

: MilitaryActivity a rdfs:Class ;
```

```
    skos:prefLabel "Military Activity"@en ;
    skos:prefLabel "Sotatoimi"@fi ;
    rdfs:subClassOf :Activity.

:Battle a rdfs:Class ;
    skos:prefLabel "Battle"@en ;
    skos:prefLabel "Taistelu"@fi ;
    rdfs:subClassOf :MilitaryActivity .

:Bombardment a rdfs:Class ;
    skos:prefLabel "Bombardment"@en ;
    skos:prefLabel "Pommitus"@fi ;
    rdfs:subClassOf :MilitaryActivity .

:Promotion a rdfs:Class ;
    skos:prefLabel "Promotion"@en ;
    skos:prefLabel "Ylennys"@fi ;
    rdfs:subClassOf :MilitaryActivity .

:Wounding a rdfs:Class ;
    rdfs:subClassOf :Event ;
    skos:prefLabel "Wounding"@en ;
    skos:prefLabel "Haavoittuminen"@fi .

:Disappearing a rdfs:Class ;
    rdfs:subClassOf crm:E5_Event ;
    skos:prefLabel "Disappearing"@en ;
    skos:prefLabel "Katoaminen"@fi .

:UnitJoining a rdfs:Class ;
    skos:prefLabel "Unit Joining"@en ;
    skos:prefLabel "Joukko-osaston liittyminen suurempaan
    joukko-osastoon"@fi ;
    rdfs:subClassOf :MilitaryActivity .

:MedalAwarding a rdfs:Class ;
    rdfs:subClassOf crm:E7_Activity ;
    skos:prefLabel "Kunniamitalin myöntäminen"@fi, "Medal
    Awarding"@en .

:PersonJoining a rdfs:Class ;
    skos:prefLabel "Person joining a Military Unit"@en ;
    skos:prefLabel "Henkilön liittyminen joukko-osastoon"@fi
    ;
    rdfs:subClassOf :MilitaryActivity .

:UnitNaming a rdfs:Class ;
    skos:prefLabel "Unit Renaming"@en ;
    skos:prefLabel "Joukko-osaston nimeäminen"@fi ;
```



```

    rdfs:subClassOf crm:E66_Formation, :MilitaryActivity .

:TroopMovement a rdfs:Class ;
    skos:prefLabel "Troop Movement"@en ;
    skos:prefLabel "Joukkojen liikutus"@fi ;
    rdfs:subClassOf :MilitaryActivity,
        crm:E9_Move .

# Properties used:
# (any) crm:P4_has_time-span -> (crm:E52_Time-Span)
# (any) crm:P7_took_place_at -> (:Place / PNR place) place
# (any) crm:P14_carried_out_by -> (Actor) active actor: in
    photography the photographer
# (any) crm:P11_had_participant -> (Actor) any actor involved
# (any) dct:source -> (:Source) source
# (any) skos:prefLabel -> (string) label (same as description
    )
# (any) dct:description -> (string) description
#
# :Death crm:P100_was_death_of -> (Person)
#
# :Birth crm:P98_brought_into_life -> (Person)
#
# :MedalAwarding crm:P141_assigned -> :Medal
# :MedalAwarding crm:P11_had_participant -> (Person) the
    person getting the medal
#
# crm:E66_Formation P95_has_formed -> (:MilitaryUnit) formed
    unit
#
# :UnitNaming P95_has_formed -> (:MilitaryUnit) (re)named
    unit
#
# :TroopMovement P95_has_formed -> (:MilitaryUnit)
    concentrated unit
#
# :UnitJoining P143_joined -> (:MilitaryUnit) the unit
    joining a larger unit
# :UnitJoining P144_joined_with -> (:MilitaryUnit) the larger
    unit with which the
#
    smaller
    unit is joining

## RANKS

:Rank a rdfs:Class ;
    rdfs:subClassOf crm:E55_Type ;
    skos:prefLabel "Military Rank"@en,

```

```
"Sotilasarvo"@fi .

:hasRank a rdf:Property ;
  Rdfs:domain etypes:Promotion ;
  rdfs:range ranks:Rank ;
  skos:prefLabel "Military Rank"@en,
    "Sotilasarvo"@fi .

mil:rankSeniorTo skos:prefLabel "ylempi kuin"@fi .

mil:equalTo skos:prefLabel "vastaa"@fi .

## WAR DIARIES

:WarDiary a rdfs:Class ;
  rdfs:subClassOf crm:E31_Document ;
  skos:prefLabel "War Diary"@en,
    "Sotapäiväkirja"@fi .

## MEDALS

:Medal a rdfs:Class ;
  rdfs:subClassOf crm:E24_Physical_Man-Made_Thing ;
  skos:prefLabel "Kunniamitali"@fi .

## SOURCES

:Source a rdfs:Class ;
  skos:prefLabel "Lähde"@fi, "Source"@en ;
  rdfs:subClassOf crm:E31_Document .
```

Liite 2: Henkilötoimija (RDF)

RDF-esimerkki henkilödatasta.

```

@prefix : <http://ldf.fi/warsa/actors/> .
@prefix atypes: <http://ldf.fi/warsa/actors/actor_types/> .
@prefix crm: <http://www.cidoc-crm.org/cidoc-crm/> .
@prefix etypes: <http://ldf.fi/warsa/events/event_types/> .
@prefix events: <http://ldf.fi/warsa/events/> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix medals: <http://ldf.fi/warsa/medals/> .
@prefix ranks: <http://ldf.fi/warsa/actors/ranks/> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix sources: <http://ldf.fi/warsa/sources/> .
@prefix times: <http://ldf.fi/warsa/events/times/> .

# Perustiedot
:person_294 a atypes:MilitaryPerson ;
  dc:source sources:source5 ;
  owl:sameAs natbib:p92,
    <http://dbpedia.org/resource/Jorma_Karhunen>,
    <http://wikidata.org/entity/Q5482501> ;
  foaf:page <http://fi.wikipedia.org/wiki/Jorma_Karhunen>,
    <http://mannerheim-ristinritarit.fi/ritarit?xm_ksort=
      s&xmid=38> ;
  skos:prefLabel "Jorma (Joppe) Karhunen"@fi ;
  foaf:familyName "Karhunen"^^xsd:string ;
  foaf:firstName "Jorma (Joppe)"^^xsd:string ;
  dc:description ""Jorma "Joppe" Karhunen (17. maaliskuuta
    1913 - 18. tammikuuta 2002 Tampere) oli suomalainen
    upseeri,...""@fi .

# Syntymä
events:birth_294 a crm:E67_Birth ;
  dc:source sources:source10 ;
  crm:P7_took_place_at <http://ldf.fi/pnr/P_10060945> ;
  crm:P4_has_time-span times:time_1913-03-17-1913-03-17 ;
  crm:P98_brought_into_life :person_294 ;
  skos:prefLabel "Jorma (Joppe) Karhunen syntyi"@fi .

# Kuolema
events:death_294 a crm:E69_Death ;
  dc:source sources:source10 ;
  crm:P100_was_death_of :person_294 ;
  crm:P4_has_time-span times:time_2002-01-18-2002-01-18 ;
  crm:P7_took_place_at <http://ldf.fi/warsa/places/
    municipalities/m_place_420> ;
  skos:prefLabel "Jorma (Joppe) Karhunen kuoli"@fi .

```

```
# Ylennys kapteeniksi
events:kapteeni_294 a etypes:Promotion ;
dc:source sources:source10 ;
:hasRank ranks:Kapteeni ;
crm:P11_had_participant :person_294 ;
crm:P4_has_time-span times:time_1941-08-04-1941-08-04 ;
skos:prefLabel "Kapteeni - Jorma (Joppe) Karhunen"@fi .

# Ylennys majuriksi
events:majuri_294 a etypes:Promotion ;
dc:source sources:source10 ;
:hasRank ranks:Majuri ;
crm:P11_had_participant :person_294 ;
crm:P4_has_time-span times:time_1943-08-31-1943-08-31
;
skos:prefLabel "Majuri - Jorma (Joppe) Karhunen"@fi .

# Kuuluminen joukko-osastoon
events:joining_294_459 a etypes:PersonJoining ;
:hasUnit "LLv 24" ;
dc:source sources:source5 ;
crm:P107_1_kind_of_member "Lentueenpäällikkö" ;
crm:P143_joined :person_294 ;
crm:P144_joined_with :actor_459 ;
crm:P4_has_time-span times:time_1939-11-30-1940-03-13 ;
skos:prefLabel "Lentueenpäällikkö, Lentolaivue 24,
1939-40"@fi .

# Kunniamitali
events:medal_83_294 a etypes:MedalAwarding ;
crm:P11_had_participant :person_294 ;
crm:P141_assigned medals:medal_83 ;
crm:P4_has_time-span times:time_1942
-09-08-1942-09-08> ;
skos:prefLabel "Mannerheim-ristin ritari no 92, 2. luokka
"@fi .
```

Liite 3: Joukko-osasto (RDF)

RDF-esimerkki joukko-osaston määrittelystä.

```

@prefix : <http://ldf.fi/warsa/actors/> .
@prefix atypes: <http://ldf.fi/warsa/actors/actor_types/> .
@prefix crm: <http://www.cidoc-crm.org/cidoc-crm/> .
@prefix dc: <http://purl.org/dc/elements/1.1/> .
@prefix dcterms: <http://purl.org/dc/terms/> .
@prefix etypes: <http://ldf.fi/warsa/events/event_types/> .
@prefix events: <http://ldf.fi/warsa/events/> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix natbib: <http://ldf.fi/history/kb/> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix ranks: <http://ldf.fi/warsa/actors/ranks/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix sources: <http://ldf.fi/warsa/sources/> .
@prefix times: <http://ldf.fi/warsa/events/times/> .
@prefix xml: <http://www.w3.org/XML/1998/namespace> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

# Osasto:
:actor_2477 a atypes:MilitaryUnit ;
  dc:description "Jalkaväkirykmentin 17 perusti Satakunta-Hämeen sotilaslääni 14. lokakuuta 1939 julistetun YH:n nimellä toteutetun Suomen puolustusvoimain tosiasiallisen liikekannallepanon yhteydessä. Yksikön perustaneeseen sotilaslääniin kuului Tampereen ja Porin sotilaspiirit."@fi ;
  dc:source sources:source1, sources:source3, sources:source7, sources:source10 ;
  owl:sameAs <http://www.wikidata.org/entity/Q18693478> ;
  foaf:page <https://fi.wikipedia.org/wiki/Jalkav%C3%A4kirykmentti%2017%20%28talvisota%29> ;
  skos:altLabel "JR 7"^^xsd:string ;
  skos:prefLabel "Jalkaväkirykmentti 7"@fi .

# Osaston perustaminen:
events:formation_2477 a crm:E66_Formation ;
  :placed_at "Tampere"^^xsd:string ;
  dc:source sources:source1, sources:source3 ;
  crm:P4_has_time-span times:time_1939-10-14-1939-10-14 ;
  crm:P7_took_place_at <http://ldf.fi/warsa/places/municipalities/m_place_420> ;
  crm:P95_has_formed :actor_2477 ;
  skos:altLabel "JR 17"^^xsd:string ;
  skos:prefLabel "Jalkaväkirykmentti 17"@fi .

```

```
# JR 7 osaksi 3. Divisioonaa:
events:joining_2477_2468 a etypes:UnitJoining ;
  dc:source sources:source1 ;
  crm:P143_joined :actor_2477 ;
  crm:P144_joined_with :actor_2468 .

# I/JR 7 osaksi JR 7:a:
events:joining_2478_2477 a etypes:UnitJoining ;
  dc:source sources:source1 ;
  crm:P4_has_time-span times:time_1939-10-14-1939-10-14 ;
  crm:P143_joined :actor_2478 ;
  crm:P144_joined_with :actor_2477 .

# Komentaja K.A. Heiskanen
events:joining_72_2477 a etypes:PersonJoining ;
  dc:source sources:source1, sources:source5, sources:
    source10 ;
  crm:P107_1_kind_of_member "Komentaja" ;
  crm:P143_joined :person_72 ;
  crm:P144_joined_with :actor_2477 ;
  crm:P4_has_time-span times:time_1939-10-14-1940-03-13 ;
  skos:prefLabel "komentaja, JR 17" .

# Osaston keskittäminen:
events:concentration_2477 a etypes:TroopMovement ;
  :placed_at "Kaipiainen"^^xsd:string ;
  dc:source sources:source3 ;
  crm:P4_has_time-span times:time_1939-10-21-1939-10-22 ;
  crm:P7_took_place_at <http://ldf.fi/pnr/P_10381450> ;
  crm:P11_had_participant :actor_2477 ;
  skos:prefLabel "JR 17:n keskittäminen"@fi .

# Nimenvaihdos JR 17 - JR 7:
events:naming_2354 a etypes:UnitNaming ;
  dc:source sources:source1, sources:source3 ;
  crm:P4_has_time-span times:time_1939-12-31-1939-12-31 ;
  crm:P95_has_formed :actor_2477 ;
  skos:altLabel "JR 7"^^xsd:string ;
  skos:prefLabel "Jalkaväkirykmentti 7"@fi .

# Osaston lopettaminen:
events:dissolution_2477 a crm:E68_Dissolution ;
  crm:P4_has_time-span times:time_1940-04-23-1941-06-24 ;
  crm:P99_dissolved :actor_2477 ;
  skos:prefLabel "JR 17:n lakkauttaminen"@fi .

# Osaston sotapäiväkirja:
diaries:diary_2198 a diaries:WarDiary ;
```

```
dcterms:hasFormat <http://digi.narc.fi/digi/slistaus.ka?
  ay=37103> ;
crm:P4_has_time-span times:time_1940-01-01-1940-04-23 ;
crm:P70_documents :actor_2477 ;
skos:prefLabel "Jalkaväkirykmentti 7. 1.1.1940-23.4.1940"
  ^^xsd:string .
```

Liite 4: Sotilasarvo (RDF)

RDF-esimerkki sotilasarvon määrittelystä.

```

@prefix : <http://ldf.fi/warsa/actors/ranks/> .
@prefix dc: <http://purl.org/dc/elements/1.1/> .
@prefix dcterms: <http://purl.org/dc/terms/> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix mil: <http://rdf.muninn-project.org/ontologies/
  organization#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .

:Majuri a :Rank ;
  dc:description "Major is a military rank of commissioned
    officer status, with corresponding ranks existing in
    many military forces ..."@en,
    "Majuri (lyhenne maj.) on useissa eri asevoimissa ..."
    @fi ;
  :level 13 ;
  dcterms:isPartOf :Esiupseeri ;
  mil:rankSeniorTo :Kapteeni ;
  owl:sameAs <http://dbpedia.org/page/Majuri> ;
  skos:altLabel "maj" ;
  skos:prefLabel "Majuri"@fi, "Major"@en ;
  foaf:page <https://fi.wikipedia.org/wiki/Majuri> .

:Komentajakapteeni a :Rank ;
  dc:description ""Komentajakapteeni (lyhenne kom.kapt.)
    on Suomen merivoimien sotilasarvo. Maavoimissa vastaa
    majurin sotilasarvo ..."@fi ;
  :level 13 ;
  mil:equalTo :Majuri ;
  mil:rankSeniorTo :Kapteeniluutnantti ;
  skos:altLabel "komkapt" ;
  skos:prefLabel "Komentajakapteeni"@fi, "Lieutenant
    Commander"@en ;
  foaf:page <https://fi.wikipedia.org/wiki/
    Komentajakapteeni> .

```


Liite 5: Sotapäiväkirja (RDF)

RDF-esimerkki sotapäiväkirjojen määrittelystä.

```
@prefix : <http://ldf.fi/warsa/actors/> .
@prefix atypes: <http://ldf.fi/warsa/actors/actor_types/> .
@prefix crm: <http://www.cidoc-crm.org/cidoc-crm/> .
@prefix dc: <http://purl.org/dc/elements/1.1/> .
@prefix dcterms: <http://purl.org/dc/terms/> .
@prefix etypes: <http://ldf.fi/warsa/events/event_types/> .
@prefix events: <http://ldf.fi/warsa/events/> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix natbib: <http://ldf.fi/history/kb/> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix ranks: <http://ldf.fi/warsa/actors/ranks/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix sources: <http://ldf.fi/warsa/sources/> .
@prefix times: <http://ldf.fi/warsa/events/times/> .
@prefix xml: <http://www.w3.org/XML/1998/namespace> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

# Eversti Simeliuksen päiväkirja:
diaries:diary_13045 a :WarDiary ;
  dcterms:hasFormat <http://digi.narc.fi/digi/dosearch.ka?
    ay=19437.SARK> ;
  crm:P4_has_time-span times:time_1944-09-20-1944-10-05 ;
  crm:P70_documents :person_110 ;
  skos:prefLabel "Eversti Simeliuksen valtuuskunta.
    20.9.1944 - 5.10.1944"@fi .

# Jalkaväkirykmentti 11:n päiväkirja:
diaries:diary_1349 a :WarDiary ;
  dcterms:hasFormat <http://digi.narc.fi/digi/slistaus.ka?
    ay=36294> ;
  crm:P4_has_time-span times:time_1940-01-15-1940-03-28 ;
  crm:P70_documents :actor_1440 ;
  skos:prefLabel "Jalkaväkirykmentti 11.
    15.1.1940-28.3.1940"@fi .
```

Liite 6: Kunniamerkki (RDF)

RDF-esimerkki kunniamerkkityypin määrittelystä.

```
@prefix : <http://ldf.fi/schema/warsa/> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix medals: <http://ldf.fi/warsa/medals/> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .

medals:medal_130 a :Medal ;
    skos:altLabel "MR 1" ;
    skos:prefLabel "Mannerheim-ristin ritari 1. lk."@fi ;
    foaf:page <https://fi.wikipedia.org/wiki/Mannerheim-risti
    > .

medals:medal_83 a :Medal ;
    skos:altLabel "MR 2" ;
    skos:prefLabel "Mannerheim-ristin ritari 2. lk."@fi ;
    foaf:page <https://fi.wikipedia.org/wiki/Mannerheim-risti
    > .
```

Liite 7: SPARQL-esimerkkihaku

Kysely hakee tietyistä osastosta ja kaikista sen alaosastoista määriteltynä ajanjaksona kuolleet, haavoittuneet tai kadonneet.

```

PREFIX events: <http://ldf.fi/warsa/events/>
PREFIX crm: <http://www.cidoc-crm.org/cidoc-crm/>
PREFIX geo: <http://www.w3.org/2003/01/geo/wgs84_pos#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

# Henkilö, koordinaatit ja tapahtumapäivä:
SELECT ?person ?lat ?lon ?date WHERE {
  # Etsitään osastoon liittyvät henkilöt:
  { SELECT DISTINCT ?person WHERE {
    # Osasto (JR 37):
    VALUES ?unit { <http://ldf.fi/warsa/actors/actor_940> }
    # Osastoon tai sen alaosastoon kuuluneet henkilöt:
    ?unit (^crm:P144_joined_with/crm:P143_joined)+ ?person.
    ?person rdf:type/rdfs:subClassOf* crm:E21_Person .
  }
}

# Tapahtumaluokat ja ominaisuudet
VALUES (?eventclass ?rel) {
  (crm:E69_Death crm:P100_was_death_of)
  (events:Wounding crm:P11_had_participant)
  (events:Disappearing crm:P11_had_participant)
}

# Päivämäärät ja paikkakoordinaatit:
?evt a ?eventclass ;
  ?rel ?person ;
  crm:P4_has_time-span/crm:P82a_begin_of_the_begin
    ?date ;
  crm:P7_took_place_at
    [ geo:lat ?lat ; geo:long ?lon ].

# Suodatetaan aikaväli:
FILTER("1939-12-26"^^xsd:date <= ?date
  && ?date <= "1940-01-07"^^xsd:date)

} ORDER BY ?death_date

```