WarSampo
Finnish WW II on the Semantic Web

UiT Digital Humanities Conference, Nov 30, 2017

Prof. Eero Hyvönen
Aalto University and University of Helsinki
Semantic Computing Research Group (SeCo)
Helsinki Centre for Digital Humanities (HELDIG)

http://seco.cs.aalto.fi/
http://heldig.fi
WarSampo team and collaborators

- Erkki Heino
- Petri Leskinen
- Esko Ikkala
- Mikko Koho
- Minna Tamper
- Jouni Tuominen
- Eetu Mäkelä

- National Archives, Defense Forces, War Museum, Association for Military History in Finland, Bonniers, National Land Survey, Kaatuneiden Muistosäätiö, Association for Cherishing the Memory of the Dead of the War, Finnish Cultural Foundation, Ministry of Education and Culture, Suomi Finland 100 Programme
Outline

• Vision
  – WW2 on the Semantic Web

• Realization
  – Infrastructure: WarSampo LOD data service
  – Applications: WarSampo semantic portal
What is the Most Devastating Catastrophe in Human History?
Description: World War II map of Europe providing details about the most devastating war in the world.
How Much Data is Available about WW2 as Linked Open Data?
~ 0
"We learn from history, that we learn nothing from history"

Georg Wilhelm Friedrich Hegel
WarSampo tries to prove that Hegel was wrong!

The more we know about the price of the war, the less there will be wars.
Goals of WarSampo

• Create a shared **LOD Infrastructure of WW2**
  – For historians to use in research
  – For the public to understand and know Finnish history
  – To promote peace and Finnish identity as a nation

• Create **Applications of the LOD Service**
  – To support Digital Humanities research
  – Cultural Heritage publication for citizens
  – To support collaborative Citizen Science

• **Develop Semantic Web Technology**
  – Data Production and Publication
    • Annotation, enriching, linking, validation, LD publishing
  – Data analysis, visualization, and knowledge discovery
Why infrastructure?

"Intellectuals solve problems - geniuses prevent them"  
Albert Einstein
Realization
Challenges:
Content Complexity & Production
Problem 1: Cultural Content Complexity
- Heterogenous and Interlinked

Artifacts
Maps
Encyclopedia
Videos
Narratives
Literature
Buildings
Music
Cultural sites
Biographies

Gallen-Kallela, Akseli (1865 - 1931)

Fine arts
Problem 2: Cultural Content Production System
- Distributed and Independent

Museums

Land survey

Archieves

Web 2.0 sites

Linked Data

Media

Citizens

Libraries
Solution: Linked (Open) Data
"Sampo" Model for Semantic CH Portals

Ontology Infrastructure

Semantic Metadata

Content Providers

Land survey

Museums

Web 2.0 sites

Archives

Linked Data

Media

Libraries

Citizens
In Principle a Piece of Cake but ...
Local Content Creation

RDF content creation process for a local data source.
Global Content Aggregation

Global data aggregation of local RDF data sources.
WarSampo: two components

Data Service for Linked Open Data
http://ldf.fi

Applications based on the service
http://sotasampo.fi
LDF.fi: 7-star Linked Data Service Model
http://www.ldf.fi/dataset/warsa

Linked Data Finland
Living Laboratory Data Service for the Semantic Web

This site is the Living Laboratory of the Linked Data Finland research initiative, conducted by the Semantic Computing Research Group at Aalto University in collaboration with University of Helsinki and a large consortium of Finnish public organizations and companies.

Our goal is to make life easier for both publishers as well as consumers of structured data on the Web. We base our work on the Linked Data paradigm and stack of standards, which combines an expressive, semantic data model (RDF) with standardized access mechanisms (SPARQL and live HTTP URIs).

5-star Linked Data

The baseline of our work is the 5-star Linked Data model, proposed originally by Tim Berners-Lee.

* Make data available on the Web in whatever format.
** Make data available as structured data (e.g., Excel instead of an image scan of a table).
*** Use non-proprietary formats (e.g., CSV instead of Excel format).
**** Use URIs to denote things, so that people can point at your data.
****** Link your data to other data to provide context.

7-star Linked Data Service

However, in our opinion, providing 5-star Linked Data is just the beginning. To actually make use of the datasets, consumers need more support in getting to know and access them, as well as a better grasp of their quality and provenance. To this end, we extend the model with two additional stars:

****** Provide your data with a schema and documentation so that people can understand and re-use your data easily.
******* Validate your data and denote its provenance so that people can trust the quality of your data.

This added support should come with as little extra work as possible to the data publisher. Our hypothesis is that a lot of this can be done automatically, basing on the Linked Data core. A data publisher needs only to provide their data in the RDF format, and the LDF.fi portal will do the rest automatically. See the overview paper (in ESWC 2014 Proceedings, Springer-Verlag) for some more details about the underlying ideas.

More info: [Hyvönen et al., ESWC P&D, 2014]
## WarSampo Datasets

<table>
<thead>
<tr>
<th>#</th>
<th>Name</th>
<th>Providing organization</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Casualties of WW2</td>
<td>National Archives</td>
<td>94,700 death records</td>
</tr>
<tr>
<td>2</td>
<td>War diaries</td>
<td>National Archives</td>
<td>13,000 war diaries of troops</td>
</tr>
<tr>
<td>3</td>
<td>Photos &amp; films</td>
<td>Defence Forces</td>
<td>160,000 photos &amp; films</td>
</tr>
<tr>
<td>4</td>
<td>Kansa Taisteli magazine articles</td>
<td>The Assoc. for Military History in Finland &amp; Bonnier</td>
<td>3,357 articles of veteran soldiers</td>
</tr>
<tr>
<td>5</td>
<td>Karelian places</td>
<td>Jyrki Tiittanen / National Land Survey</td>
<td>32,400 places of the annexed Karelia</td>
</tr>
<tr>
<td>6</td>
<td>Karelian maps</td>
<td>National Land Survey</td>
<td>47 wartime maps of Karelia</td>
</tr>
<tr>
<td>7</td>
<td>Senate atlas</td>
<td>National Archives</td>
<td>404 historical maps of Finland</td>
</tr>
<tr>
<td>8</td>
<td>Municipalities</td>
<td>National Archives</td>
<td>625 wartime municipalities</td>
</tr>
<tr>
<td>9</td>
<td>Organization cards</td>
<td>National Archives</td>
<td>ca 500 army units &amp; ca 300 persons &amp; 642 battles</td>
</tr>
<tr>
<td>10</td>
<td>National Biography</td>
<td>Finnish Literature Society</td>
<td>ca 500 biographies of wartime persons</td>
</tr>
<tr>
<td>11</td>
<td>Wartime events</td>
<td>War history books</td>
<td>1,000 events</td>
</tr>
<tr>
<td>12</td>
<td>Persons</td>
<td>War history books, Wikipedia</td>
<td>2,600 persons</td>
</tr>
<tr>
<td>13</td>
<td>Army units</td>
<td>War history books</td>
<td>3,200 army units</td>
</tr>
</tbody>
</table>
Conceptual Data Model Core: Extending CIDOC CRM
WarSampo Linked Open Data Cloud

12M triples knowledge graph
Applications
Developers View to Linked Data: Rich Internet Applications (RIA)

Application 1

Application 2

Application N

SPARQL End Point

Linked Data Service
WWW Standard Model

Client Side (Browser)

Server Side
Semantic Portal http://sotasampo.fi
8 Perspectives to War

In-use since Nov 2015
10 000 users in 3 days

Select a perspective to search and browse the WarSampo data

1. Events
   - Events of the Finnish War
   - Search for and browse events with related linked data

2. Persons
   - Persons and individuals involved in the Finnish War
   - Search for and browse individuals with related linked data

3. Army Units
   - Armies and units involved in the Finnish War
   - Search for and browse army units with related linked data

4. Places
   - Locations and places during the Finnish War
   - Search for and browse places with related linked data

5. Deaths
   - Deaths and fatalities during the Finnish War
   - Search for and browse deaths and fatalities with related linked data

6. Memoirs
   - Memoirs and personal accounts of the Finnish War
   - Search for and browse personal accounts with related linked data

7. Photos
   - Photographs depicting the Finnish War
   - Search for and browse photographs with related linked data

8. Cemeteries
   - Cemeteries and burial sites associated with the Finnish War
   - Search for and browse cemeteries with related linked data

More info: [Hyvönen et al., ESWC 2016; Koho et al., WHiSe 2016]
Perspective 1: Events on Maps and Timeline
Perspective 2: Persons

Reassembling the war history on individual soldiers!
Perspective 3: Army Units
Perspective 4: Historical Places

All places with linked data

Search places  Search maps
Zooming into Historical Maps and Events
After clicking on place "Ristimäki" related events (with links) can be found
### Perspective 5: Linked Death

Death records of 95,000 soldiers

More info: [Koho et al., ESWC WhiSe 2016]
Using the Data in Digital Humanities

Analysing military unit data:
- Faceted search for filtering data
- Data analysis & visualizations

Casualties of the 33rd infantry regiment
Four ontologies for linking named entities

More info: [Mäkelä et al., DH 2016]
Perspective 7:
160 000 authentic photographs
New Application Nov 22, 2017:
"Home pages" for 673 war cemeteries & 95 000 casualties

Akka, Kylmäkosken sankarihautausmaa
Perustamisvuosi: 1940-01-10
Hautojen lukumäärä: 115
Nykyinen kunta: Akka
Entinen kunta: Kylmäkoski
Katuosoite: Lehttie 3, 37910 Kylmäkoski
URI: http://ldf.fi/maara/places/cemeteries/0218_1

50 000 visitors in a week
Context Needed for Disambiguation

• Naive linking not feasible
  – E.g. photo captions -> persons
    • Naive string match 1 932 036 links to 53 848 resources
    • Context data used 27 420 links to 16 033 resources

• Contextual data used for disambiguating people
  – Name
  – Ranks and promotion dates
  – Time of death
  – Match length
    • E.g. “eversti P. Talvela” vs. “P. Talvela”
  – Military decorations (mainly the Mannerheim Cross)
  – Military unit
Disambiguating Person Names

1. “Kersantti Leskinen” yields 3 possible matches
2. None of them dead before the date of the photograph
3. Reino Leskinen, however, served in “Sissipataljoona 2”.
4. Seems likely that Reino Leskinen is the person in the photo
5. He happened to die on the very same day
   1. Although the date of the photograph might be incorrect

Data is not complete: casualties and well-known military persons are known but ordinary soldier who survived are not (at least for now)
• A test with links from 50 photos

• Places
  – Precision: 76%
  – Recall: 85%

• Persons
  – Precision: 76%
  – Recall: 76%
Challenge 2: Dynamic Army Unit Ontologies

• Synonymy and ambiguity of unit names in different wars
  – Winter War 1939-1940
  – Continuation War 1941-1944
  – Lapland War 1944-1945

• Units were created, merged, and changed during each war

• Unit names were changed in order to confuse the enemy!
Conclusions: Strategic Steps Taken

1. Establish collaboration network

2. Project and Funding

3. Prototype

4. Sustainable ecosystem
Conclusions: Linked Data Makes a Difference

- **End-user’s perspective**
  - Global view to heterogeneous, distributed contents
  - Automatic content aggregation
  - Semantic search
  - Semantic browsing and recommendations
  - Other intelligent services (knowledge discovery, personalization, visualization, …)

- **Content publisher’s perspective**
  - Distributed content creation
  - Enriching each other’s contents semiantically
  - Automated link maintenance
  - Shared content publication channel
  - Reusing aggregated content in other applications
• War history is a very promising domain for Linked Data

• WarSampo aims at creating **LOD infra for WW2**
  – For creating applications
  – For Digital Humanities research
But the Lunch is not Free

- More collaboration is need -> complicates work
- Integration of semantic portals with legacy systems
- Manual annotations are costly and may not scale up
- Automatic annotation lowers data quality
Thanks!

More info and publications on the web:

Semantic Computing Research Group (SeCo)
Making computers and the web more intelligent and interoperable

WarSampo:
Finnish World War II on the Semantic Web

Goal: Understanding History and Promoting Peace

According to Georg Wilhelm Friedrich Hegel we learn from history that we learn nothing from history. Hopefully this is not the case for the Second World War (WW2), now that fighting has started again even within the borders of Europe in Ukraine. One way to promote peace is to make reliable data about the war openly available for everybody to learn.

WarSampo is the next step in our series of "Sampo" portals based on Linked Data, including CulturoSampo, BookSampo, and TravelSampo, and continues our earlier works on modeling the First World War as Linked Data. The project started in autumn 2014 and is finished in 2017, by the Centennial of Finland's Independence, December 6th.

Figure: Mäkituoto artillery fires at the Battle of Hanko in 1942. Finnish Wartime Photograph Archive, Defence Forces.