History on the Semantic Web as Linked Data
— An Event Gazetteer and Timeline for World War I

Eero Hyvönen, Thea Lindquist, Juha Törnroos, and Eetu Mäkelä
Semantic Computing Research Group
Aalto University, University of Colorado Boulder, and University of Helsinki
Outline

- Research Goals and Means
- Context: Web of Linked Data
- Focus and Contributions of This Paper
- Requirements for an Event Gazetteer of History
- Metadata Elements for Historical Events
- An Event Gazetteer and Timeline for WWI
- Semantic Portal Application
Research Goals and Means

- **Goals**
  - Enriching other cultural heritage content by linking data
    - E.g. paintings, books, artifacts, news, photographs, …
  - Providing more comprehensive history data online for a variety of audiences
  - Providing finer grained, multi-valued data about history
  - Detecting differences in views/opinions to support historical research
  - Answering history questions, based on Linked Data, that cannot be answered based on separate datasets alone

- **Means**
  - Linked Data paradigm and technology as the basis: focus on events
  - Aggregating local/regional/national perspectives on events into a global view
  - Aggregating "subjective" views/opinions about the events
  - Link history with CH collections, ontologies and linked datasets
Linked Data ≈ Web of Data ≈ Semantic Web

- **Web of Pages** (traditional web)
  - Links connect pages
  - Used by browsers
- **Web of Data** (semantic web)
  - Links connect concepts (e.g. magpie -> bird)
  - Used by semantic web browsers and applications
- WWW contains both webs
  - Web of Pages for **humans**
    » **WWW** World Wide Web
  - Web of Data for **machines**
    » **GGG** Giant Global Graph
Using WWW and GGG

WWW

GGG

(Anja Jentzsch, 2012)
Linked Open Data Story so Far
Starting 2005

(Tim Berners-Lee)
New Datasets Linked with Others

(Tim Berners-Lee)
(Tim Berners-Lee)
Linked Open Data 2009
Linked Open Data Cloud 2010:
http://linkeddata.org
Linked Open Data Cloud 2011: http://linkeddata.org

onki.fi: Finnish municipalities
## LOD Cloud Statistics, Sept 2011

<table>
<thead>
<tr>
<th>Domain</th>
<th>Number of datasets</th>
<th>Triples</th>
<th>%</th>
<th>(Out-)Links</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Media</td>
<td>25</td>
<td>1,841,852,061</td>
<td>5.82 %</td>
<td>50,440,705</td>
<td>10.01 %</td>
</tr>
<tr>
<td>Geographic</td>
<td>31</td>
<td>6,145,532,484</td>
<td>19.43 %</td>
<td>35,812,328</td>
<td>7.11 %</td>
</tr>
<tr>
<td>Government</td>
<td>49</td>
<td>13,315,009,400</td>
<td>42.09 %</td>
<td>19,343,519</td>
<td>3.84 %</td>
</tr>
<tr>
<td>Publications</td>
<td>87</td>
<td>2,950,720,693</td>
<td>9.33 %</td>
<td>139,925,218</td>
<td>27.76 %</td>
</tr>
<tr>
<td>Cross-domain</td>
<td>41</td>
<td>4,184,635,715</td>
<td>13.23 %</td>
<td>63,183,065</td>
<td>12.54 %</td>
</tr>
<tr>
<td>Life sciences</td>
<td>41</td>
<td>3,036,336,004</td>
<td>9.60 %</td>
<td>191,844,090</td>
<td>38.06 %</td>
</tr>
<tr>
<td>User-generated content</td>
<td>20</td>
<td>134,127,413</td>
<td>0.42 %</td>
<td>3,449,143</td>
<td>0.68 %</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>295</strong></td>
<td><strong>31,634,213,770</strong></td>
<td><strong>%</strong></td>
<td><strong>503,998,829</strong></td>
<td><strong>%</strong></td>
</tr>
</tbody>
</table>
Focus of this Paper: Historical Events Fostering Linked Data

- Events are a basis for modern metadata models
  - Generic event-centric schema ontologies
  - CIDOC CRM, LIDO,…
- Events = semantic glue linking cultural heritage (CH) content
  - Persons, organizations
  - Places
  - Times
  - Artifacts
  - Intangible things
  - Other events
- We consider vocabularies of particular historical events
  - “Crucifixion of Jesus”, “Coronation of Napoleon I”, “World War I”
There are controlled vocabularies / gazetteers for:
- Persons, organizations
  » E.g. ULAN, authority files
- Places (modern & historical)
  » E.g. TGN, Geonames, GNS, local databases
- Artifacts
  » E.g. AAT, MASA/MAO,
- ...

Controlled vocabularies for events are missing!
- We propose semantic web "event gazetteers"
Contributions of this Paper

- Analyze requirements for an event gazetteer
  - From Linked Open Data (LOD) perspective
- A metadata model for representing events
- Application case study: World War I
  - Gazetteer of WWI events
  - Semantic portal for WWI
Requirements for an Event Gazetteer of History
Requirements for Usage

- Use as a gazetteer
  - Shared reference IDs glue data together
- Use as a data repository
  - Event timeline with descriptions
- Use metadata schema for representing events
Requirements for Linked Data

- Tim Berners-Lee’s 5 star criteria for Linked Data
  - Data is available as structured data, e.g., as an Excel sheet instead of as an image scan of a table, so that it can be reused.
  - Data is available on the Web (in whatever format) under an open license.
  - Non-proprietary open formats are used, e.g., CSV (Comma Separated Values) format instead of Excel’s own proprietary format.
  - HTTP URIs are used to identify things, so that people can point to the data and serve RDF from it.
  - Data is linked internally and externally to other data to provide context.
● Earning 5 stars in our case:
  ★ The structured semantic data based on RDF, SKOS, OWL is used
  ★ The Creative Commons Attribution License (or similar) is used
  ★ The W3C and other non-proprietary and open standards are used
  ★ Dereferencable HTTP URIs are used, according to LD principles
    ★ RDF for machines
    ★ HTML for humans
  ★ The data is linked to related repositories
    ★ E.g. DBpedia and Geonames in Linked Open Data cloud
    ★ Collections of WWI materials
Requirements for Interfaces and APIs

- **Human interface.** Human end-user-interface
  - For search engines and browsers
- **Linked Data browsing.** Linked Data browser interface based on URI dereferencing
  - LD browsers created for the Web of Data.
- **SPARQL endpoint.** For querying the data in a standard way for, e.g., mash-up applications
- **Download.** Downloading the data as an RDF data dump
Requirements for Identifiers

- Use HTTP Universal Resource Identifiers (URI)
  - Unique **identifiers** for anything (inside or outside WWW)
    » Based on Domain Name System of WWW
      ● E.g. http://www.cidoc-crm.org/
  - Like web **addresses**
    » URI identifier also tells where the data is
      ● Can be used e.g. in a browser!
    » Not all URI schemas have this property
      ● E.g. URN identifiers
      ● Not recommended for Linked Data!
- **Persistence** - cool URIs do not change
Example: Challenges in Identifying Events

- How to represent the "Battle of Albert" in WWI
- What language to use?
  - Bataille d'Albert? (fr)
  - Albertin taistelu? (fi)
- Which battle?
  - Battle of Albert (1914) (25–29 September 1914)
    » encounter battle during the Race to the Sea
    » opening phase of the Battle of the Somme, 1916
  - Battle of Albert (1918) (21–23 August 1918)
    » opening phase of the Second Battle of the Somme, 1918
- Is "Battle of Albert" a Web page, a concept and/or identifier?
Linked Data Solution in DBpedia: Dereferencing "Battle of Albert, 1914"

- Web of Data (GGG) for machines:
  - URI identifier
    » http://dbpedia.org/resource/Battle_of_Albert_%281914%29
  - Web page URL
    » http://dbpedia.org/page/Battle_of_Albert_%281914%29
  - (Meta)data on the web
    » http://dbpedia.org/data/Battle_of_Albert_%281914%29

- Web of Pages (WWW) for humans:
  - The actual Wikipedia information page
La Battaglia di Albert iniziò il 25 settembre 1914 come parte della Corso al mare durante la prima guerra mondiale. È immediatamente precedente alla prima battaglia della Marna ed alla prima battaglia dell’Aliseo. La 10 Armata francese iniziò ad ammassarsi ad Amiens a partire dalla metà di settembre, ed il 25 iniziò a spingersi verso est.

http://dbpedia.org/page/Battle_of_Albert_%281914%29
The Battle of Albert began on September 25, 1914 as part of the Race to the Sea during World War I. It directly followed the First Battle of the Marne and the First Battle of the Aisne as progress toward advancing the trench lines to the sea continued. The French Tenth Army began to assemble at Amiens from mid-September and on September 25 began to push eastwards.

The Battle of Albert began on September 25, 1914 as part of the Race to the Sea during World War I. It directly followed the First Battle of the Marne and the First Battle of the Aisne as progress toward advancing the trench lines to the sea continued. The French Tenth Army began to assemble at Amiens from mid-September and on September 25 began to push eastwards. De Castelnau, under the command of Joseph Joffre, launched a frontal attack on the German lines near Albert after attempts to stretch the line northward failed. De Castelnau's immediate resistance and counterattack as the German Sixth Army had reached Bapaume on September 26 and advanced to Thiepval on the 27th, in the midst of what was to become the Somme battlefield of 1916. The German aim was to drive westwards to the English Channel, seizing the industrial and agricultural regions of Northern France, cutting off the supply route of the British Expeditionary Force (BEF) and isolating Belgium. Neither side was able to make any decisive ground and the battle around Albert ended around September 29 as the fighting moved northwards towards Arras and Lille and into West Flanders. This confrontation and those to follow were deemed as the fighting settled into prolonged trench warfare.

The Battle of Albert began on September 25, 1914 as part of the Race to the Sea during World War I. It directly followed the First Battle of the Marne and the First Battle of the Aisne as progress toward advancing the trench lines to the sea continued. The French Tenth Army began to assemble at Amiens from mid-September and on September 25 began to push eastwards. De Castelnau, under the command of Joseph Joffre, launched a frontal attack on the German lines near Albert after attempts to stretch the line northward failed. De Castelnau's immediate resistance and counterattack as the German Sixth Army had reached Bapaume on September 26 and advanced to Thiepval on the 27th, in the midst of what was to become the Somme battlefield of 1916. The German aim was to drive westwards to the English Channel, seizing the industrial and agricultural regions of Northern France, cutting off the supply route of the British Expeditionary Force (BEF) and isolating Belgium. Neither side was able to make any decisive ground and the battle around Albert ended around September 29 as the fighting moved northwards towards Arras and Lille and into West Flanders. This confrontation and those to follow were deemed as the fighting settled into prolonged trench warfare.
Battle of Albert (1914)

This article does not cite any references or sources. Please help improve this article by adding citations to reliable sources. Unsourced material may be challenged and removed. (April 2008)

The Battle of Albert began on September 25, 1914 as part of the Race to the Sea during World War I. It directly followed the First Battle of the Marne and the First Battle of the Aisne as progress toward advancing the trench lines toward the sea continued.

The French Tenth Army began to assemble at Amiens from mid-September and on September 26 to push eastwards. De Castelnau, under the command of Joffre, launched a frontal attack on the German lines near Albert after attempts to stretch the line northward failed. De Castelnau was met with immediate resistance and counterattack as the German Sixth Army had reached Ripon on September 26 and advanced to Thiepval on the 27th. In the midst of what was to become the Somme battlefield of 1916. The German aim was to drive westward to the English Channel, seizing the industrial and agricultural regions of Northern France, cutting off the supply route of the British Expeditionary Force (BEF) and isolating Belgium.

Neither side was able to make any decisive headway and the battle around Albert ended on September 29 as the fighting moved northwards toward Arras and Lille and into West Flanders. This confrontation and those to follow were deemed draws as the fighting settled into prolonged trench warfare.

This World War I article is a stub. You can help Wikipedia by expanding it.

This article about a battle in French history is a stub. You can help Wikipedia by expanding it.

This article about a battle in German history is a stub. You can help Wikipedia by expanding it.

Help improve this article

Did you find what you were looking for? Yes No
Requirements for Core Metadata

- Minimal metadata to identify an event unambiguously
  - Name
  - Time
  - Place
  - Description
- Visualized by pattern name(time, place)
- Examples:
  - Battle of Albert (1914, France)
  - World War I (1914-1918, Belgium)
  - World War I (1914-1918, Germany)
Metadata Elements for Historical Events
<table>
<thead>
<tr>
<th>Metadata type</th>
<th>Label</th>
<th>Meaning</th>
<th>Identifier</th>
<th>Cardinality</th>
<th>Range</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Core elements</td>
<td>Name</td>
<td>Short event name</td>
<td>:name</td>
<td>0..1</td>
<td>Literal</td>
<td>string@language</td>
</tr>
<tr>
<td></td>
<td>full name</td>
<td>Full event name</td>
<td>skos:prefLabel</td>
<td>1</td>
<td>Literal</td>
<td>string@language</td>
</tr>
<tr>
<td></td>
<td>Description</td>
<td>Description of the event</td>
<td>dc:description</td>
<td>0..1</td>
<td>Literal</td>
<td>string@language</td>
</tr>
<tr>
<td></td>
<td>Time</td>
<td>Time span of the event</td>
<td>:time</td>
<td>1</td>
<td>Time instance</td>
<td>Time URI</td>
</tr>
<tr>
<td></td>
<td>Place</td>
<td>Place where the event took place</td>
<td>:place</td>
<td>1</td>
<td>Place instance</td>
<td>Place URI</td>
</tr>
<tr>
<td></td>
<td>point</td>
<td>Place where the event took place</td>
<td>geo:point</td>
<td>0..1</td>
<td>Point instance</td>
<td>Point URI</td>
</tr>
<tr>
<td></td>
<td>path</td>
<td>Place where the event took place</td>
<td>geo:line</td>
<td>0..1</td>
<td>Line instance</td>
<td>Line URI</td>
</tr>
<tr>
<td></td>
<td>area</td>
<td>Place where the event took place</td>
<td>geo:Polygon</td>
<td>0..1</td>
<td>Polygon instance</td>
<td>Polygon URI</td>
</tr>
<tr>
<td>Subject</td>
<td>related actor</td>
<td>Actor involved in the event</td>
<td>:subjectActor</td>
<td>0..n</td>
<td>Actor instance</td>
<td>Actor URI</td>
</tr>
<tr>
<td></td>
<td>related time</td>
<td>Other time related to the event</td>
<td>:subjectTime</td>
<td>0..n</td>
<td>Time instance</td>
<td>Time URI</td>
</tr>
<tr>
<td></td>
<td>related place</td>
<td>Other place related to the event</td>
<td>:subjectPlace</td>
<td>0..n</td>
<td>Place instance</td>
<td>Place URI</td>
</tr>
<tr>
<td></td>
<td>related event</td>
<td>Other event related to the event</td>
<td>:subjectEvent</td>
<td>0..n</td>
<td>Event instance</td>
<td>Event URI</td>
</tr>
<tr>
<td></td>
<td>related topic</td>
<td>Topic concept related to the event</td>
<td>:subjectConcept</td>
<td>0..n</td>
<td>Concept</td>
<td>Topic URI</td>
</tr>
<tr>
<td></td>
<td>related object</td>
<td>Individual object related to the event</td>
<td>:subjectObject</td>
<td>0..n</td>
<td>Object instance</td>
<td>Object URI</td>
</tr>
<tr>
<td></td>
<td>classification</td>
<td>event type</td>
<td>:eventType</td>
<td>0..n</td>
<td>Concept</td>
<td>Classification URI</td>
</tr>
<tr>
<td>Narrative</td>
<td>Event hierarchy</td>
<td>Larger event</td>
<td>skos:broader</td>
<td>0..n</td>
<td>Event instance</td>
<td>Event URI</td>
</tr>
<tr>
<td></td>
<td>contains</td>
<td>Narrower event</td>
<td>skos:narrower</td>
<td>0..n</td>
<td>Event instance</td>
<td>Event URI</td>
</tr>
<tr>
<td></td>
<td>next event</td>
<td>Next event</td>
<td>:next</td>
<td>0..n</td>
<td>Event instance</td>
<td>Event URI</td>
</tr>
<tr>
<td></td>
<td>previous event</td>
<td>Previous event</td>
<td>:previous</td>
<td>0..n</td>
<td>Event instance</td>
<td>Event URI</td>
</tr>
<tr>
<td></td>
<td>cause</td>
<td>Cause of the event</td>
<td>:cause</td>
<td>0..n</td>
<td>Event instance</td>
<td>Event URI</td>
</tr>
<tr>
<td></td>
<td>effect</td>
<td>Effect of the event</td>
<td>:effect</td>
<td>0..n</td>
<td>Event instance</td>
<td>Event URI</td>
</tr>
<tr>
<td>Administrative</td>
<td>Documentation properties</td>
<td>SKOS Documentation properties</td>
<td>skos:xxx</td>
<td>0..n</td>
<td>Literal</td>
<td>string@language</td>
</tr>
</tbody>
</table>
An Event Gazetteer and Timeline for World War I

- Represent the events and story of WWI as LOD
- Publish WWI events as a 5 star RDF LOD repository & service
- Link it with relevant ontologies and datasets
- Joint work of Aalto University, University of Colorado Boulder, and University of Helsinki
  - With an international collaboration network
Datasets

- Wikipedia/DBpedia WWI timeline and events
- *Principal Events, 1914-1918* (London, 1922)
  - British government’s view of the war proceedings
  - Appendix provides German, Belgian, French and American views
- Imperial War Museum (London) data
  - Geographical terms relating to the Western Front
  - Event terms relating to WWI
- Horne & Kramer, *German Atrocities, 1914* (New Haven, 2001)
  - Atrocity data
- Tessin, *Deutsche Verbände und Truppen* (Osnabrück, 1974)
  - German army hierarchy
- *Annuaire Statistique de la Belgique* (Brussels, 1922)
  - Government data about Belgian population, economy, etc. for war years
- Polygons of Belgian provinces during WWI
  - HISSTAT project (Universities of Ghent, Brussels, Louvain-la-Neuve and State Archives of Belgium)
● Collections
  – University of Colorado Boulder, WWI Collection Online
  – Over 1,100 titles (55,000 pages)
  – Published 1829-1922, vast majority 1914-1918

● Articles
  – DBpedia
  – Potential collaboration with online encyclopedia project

● Ontologies
  – Annotated event gazetteer of WWI
  – Ontologies of agents, regions, and times
  – Places mapped on Geonames

● Specialized vocabulary relating to WWI Belgium
Annotation Process

- Collection MARCXML -> RDF
  - Transformation by machine
- Manual deep linking based on content
  - Strings in text -> URI references
- Manual creation of additional ontologies and metadata
- Automatic annotation of texts
  - Work in progress
World War I Collection

The University of Colorado at Boulder Libraries' digital World War I collection offers a rich array of primary-source material relating to the War selected from the World War I "Pamphlets" Collection housed in the CU Archives. The collection consists of over 1,200 individual titles representing a staggering range of genres, authors, geopolitical units, and subject matter. A selection of core works from the rich body of materials in the collection is currently available online. This resource offers full-text searching capability along with visualization tools that facilitate alternate avenues of exploration in the collection.
30th ... Fort Thiaumont (Verdun) retaken by French forces (see 23rd, and July 1st).

**Battle of the Strypa ends** (see 11th).
British Government conclude further agreement with the Netherlands Overseas Trust for rationing of Holland (see November 23rd, 1915).

**JULY 1916.**

1st ... **Battles of the Somme, 1916, begin** with (1st/13th) (see November 18th).
Contact patrol, or liaison with infantry, first instituted in the Royal Flying Corps.
Kirmanshah (Persia) reoccupied by Turkish forces (see February 26th, 1916, and March 11th, 1917).

2nd ... **Battle of Baranovichi begins** (see 9th).

3rd ... Russian and Japanese Governments conclude treaty with regard to future policy in the Far East.

7th ... **Italian counter-offensive in the Trentino ends** (see June 16th).
Tanga (German East Africa) occupied by British forces (see November 5th, 1914).
Mr. Lloyd George succeeds Lord Kitchener as Secretary of State for War, Great Britain (see June 5th and December 11th).
**British Government issue Order in Council rescinding Declaration of London of 1909.** French Government issue similar order (see October 29th, 1914, and November 6th, 1914).

* Bulk of position was captured on May 29th, and this date marks the limit of the German advance on Verdun.

[8369]

**Battle of Albert, 1916**
Creating Metadata about WWI Events: Classes and Instances in SAHA Editor

Collaborative annotation project (Aalto & CU Boulder)
The Semantic Portal: Faceted Search for Atrocities

List Map

[remove] no
[remove] yes

- Atrocities in Baranzy (atrocity incident)
- Atrocities in Dinant (atrocity incident)
- Atrocities in Esen (atrocity incident)
- Atrocities in Flénu (atrocity incident)
- Atrocities in Korbeek-Lo (atrocity incident)
- Atrocities in Rotselaar (atrocity incident)
- Atrocities in Somme-Leuze (atrocity incident)
- Atrocities in Zempst (atrocity incident)
Visualizing Search Results on a Map & Timeline
End-user Interface: Events on the Map and Timeline: Battle of Albert 1914

agent
France, Germany

dateModified
2012-06-08T04:42:29.824Z

description
The Battle of Albert began on September 25, 1914 as part of the Race to the Sea. Subsequent action to the Battle of the Marne, 1914 and the Battle of the Aisne, 1914, as progress toward advancing the trench lines to the sea continued.

name
Battle of Albert, 1914

place of action
Albert

time of action
9/25/1914

type of instance
historical event
Conclusions

- History is a promising application area for Linked Data
- Linked Data is based on shared concepts (resources)
- Event gazetteers are missing
- We presented a model for this based on LD requirements
- Evaluation by application is underway
  - Populating a gazetteer/timeline for WWI
  - Creating a semantic portal for WWI
- More info in the paper online:

Acknowledgements

- Michael Ortiz, Martha Hanna
- Fulbright Foundation, Tekes, Finnish Cultural Foundation