Proceedings of CIDOC2012 - Enriching Cultural Heritage, Helsinki, Finland, June 10-14, 2012.
http://network.icom.museum/cidoc/resources/cidoc-digital-library.html

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

Eero Hyvönen[1,3], Thea Lindquist[2], Juha Törnroos[1,3], and Eetu Mäkelä[1]
Aalto University (1), University of Colorado Boulder (2), University of Helsinki (3)
http://www.seco.tkk.fi/

Abstract

Events are an essential component of cultural heritage (CH) Linked Data (LD): they link actors, places, times, objects, and other events into larger narrative structures, providing a rich basis for semantic searching, recommending, analysis, and visualization of CH data. This paper argues that shared vocabularies (gazetteers, ontologies) of events, such as the "Battle of Normandy" or "Crucifixion of Jesus", are necessary to facilitate the aggregation and linking of heterogeneous content from various collections. For example, biographies, histories, photos, and paintings often reference or depict events. A set of general requirements for an event gazetteer is presented, based on the needs of publishing, aggregating, and reusing cultural heritage content as Linked Data. After this, a metadata model addressing the presented requirements for representing historical events is outlined. The model is being applied in a case study aimed at developing an event ontology for World War I (WWI). Our goals from an end-user perspective are twofold: 1) Facilitate event-based cataloging for curators in memory organizations; 2) Utilize semantic event descriptions and narrative event structures in end-user applications for searching and linking documents and other content about WWI, and for structuring and visualizing them.

1 INTRODUCTION

Ontologies (Staab, Studer, 2009) have been used in CH portals as 1) metadata schemas for harmonizing content and 2) as value vocabularies for populating schema property (element) values. For example, Dublin Core[1], the Europeana Data Model (EDM)[2], and CIDOC CRM[3] are metadata schemas, and the Getty value vocabularies[4] TGN, AAT, and ULAN provide values for places, objects, and actors in such schemas. Events have also been proposed as a basis for harmonizing cultural heritage content. The CIDOC CRM standard and LIDO harvesting schema[5], for example, are based on the idea of representing events in terms of related places, times, actors, and objects. There are value vocabularies for gazetteers of the same. However, interestingly, shared-value vocabularies for events are still lacking (even though many thesauri may contain terms for events). This paper 1) argues that specific gazetteers for events with rich semantic structure are needed, 2) proposes a model for such a vocabulary based on

[1] http://dublincore.org/
[2] http://pro.europeana.eu/edm-documentation
[3] http://www.cidoc-crm.org/
[4] http://www.getty.edu/research/tools/vocabularies/
[5] http://network.icom.museum/cidoc/working-groups/data-harvesting-and-interchange/what-is-lido.html

requirements for publishing cultural heritage content as Linked Data, and 3) presents an overview of a case study where the model is applied to represent World War I events.

In the following, requirements for an event gazetteer and then a general metadata model for events are presented and discussed, and the case study described. In conclusion, contributions of the work are summarized, related work discussed, and directions for further research outlined.

2 REQUIREMENTS FOR AN EVENT GAZETTEER OF HISTORY

2.1.1 Requirements for Usage

We argue, that an event gazetteer, i.e., a history ontology, is useful for the following types of use cases:

1. A **metadata schema**. The ontology defines a metadata model that can be used to represent historical events in cultural heritage applications. Various schemas for representing events have been developed before; in this paper a schema for historical events is suggested based on related ideas in other schemas.
2. A domain **ontology / gazetteer**. The individual events, conforming to the metadata schema, can be used as a gazetteer for indexing historical cultural heritage content, such as a photograph taken during a particular battle or a weapon used in that battle. To our knowledge event gazetteers have not been developed in a systematic fashion, thus promoting this use case is a contribution of this paper. Our case study focuses on WWI events to demonstrate how an event-based approach might be employed in a historical context.
3. A **data repository**. The ontology with its content can be used as a data source of its own about history, and be linked with related datasets. Additional data and descriptions about the events can be linked to the event URIs.

2.1.2 Requirements for Linked Data

Our general goal is to publish historical information as Linked (Open) Data (LD) (Heath, Bizer, 2011). Data can be published on the web at increasing levels of openness and linking as characterized by the 5-star system proposed by Tim Berners-Lee[6]:

- ★ 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.

[6] http://www.w3.org/DesignIssues/LinkedData.html

In our case, the goal is to create a 5-star publication of events, which sets various requirements regarding the data model and publication. In our model, the following design choices on levels 1-5 were made, respectively:

★ The W3C semantic web standards based on RDF are used[7], as is customary in the Linked Data community.
★ The Creative Commons Attribution License CC 3.0[8] (or similar related licensing models) is proposed to guarantee maximal usage of the data. This means that it is free for anyone to share, copy, modify, distribute and transmit the work, and to remix the data to adapt the work, including for commercial uses.
★ The W3C and other non-proprietary and open standards are used as required.
★ Dereferencable HTTP URIs are used, according to LD principles[9], so that both RDF and HTML representations of the events can be returned to machine and human users, respectively.
★ The data is linked to related repositories, such as to those in the Linked Open Data cloud (e.g., DBpedia[10]).

2.1.3 Requirements for Interfaces and APIs

The whole dataset is to be made available via the following interfaces:

1. **Human interface.** Human end-user-interface for searching and browsing the data.
2. **Linked Data browsing.** Linked Data browser interface based on URI dereferencing. This interface makes the data set browsable by the various tabulators and LD browsers created for the Web of Data.
3. **SPARQL-endpoint**[11]**.** Published for querying the data in a standard way for, e.g., mash-up applications.
4. **Download**. Possibility of downloading the data as an RDF data dump is provided. This possibility is useful for, e.g., external users willing to analyze, modify, or enrich the dataset, or use it in offline and other applications.

2.1.4 Requirements for Identifiers

A key concept on the Semantic Web is the notion of unique identities called *resources*. Each real-world or imaginary thing, say “World War I” or “unicorn”, as well as things on the web, say the homepage of a person, must have an indentity and identifier before we are able to speak, i.e., represent (meta)data, about it. Identities, i.e., resources, are referenced to by Universal Resource Indentifiers (URI). The idea is that events are given and identified by language-neutral identifiers for machine use, not by their names or titles in a particular natural language that are used for human consumption in applications. Using URIs makes it possible to differentiate entities with similar names and decouple machine semantics from human languages, which enables, e.g., language-independent multilingual applications. For example, consider the WWI event “Battle of Albert” in 1914. When indexing data about it, it is not clear what language should be used. For example, “Bataille d'Albert” in French or “Albertin taistelu”

[7] http://www.w3.org/standards/semanticweb/
[8] http://creativecommons.org/licenses/by/3.0/
[9] http://www4.wiwiss.fu-berlin.de/bizer/pub/LinkedDataTutorial/
[10] http://dbpedia.org/
[11] http://www.w3.org/standards/semanticweb/query

in Finnish could be used in addition to the English title. The solution is to use neutral URIs for the underlying concept, such as:

http://dbpedia.org/resource/Battle_of_Albert_%281914%29[12] (2.1)

This URI is actually used in DBpedia[13] for this particular battle, and although it is based on English, it is language neutral in the sense the URI indeed has different labels "Bataille d'Albert (1914)" in French and "Battle of Albert (1914)" in English attached to it, and separate from the URI. Depending on the HTTP request header, this URI can be resolved into

http://dbpedia.org/page/Battle_of_Albert_%281914%29

in a HTML page (e.g., when the URI is used in an ordinary web browser) or

http://dbpedia.org/data/Battle_of_Albert_%281914%29

to access the RDF representation of the concept.

Using URIs has many obvious benefits for linking data. If different content providers index their data using shared URIs (ontologies), then the distributed data can be linked together automatically in order to enrich it (Hyvönen et al., 2009). For example, if the URI (2.1) is used as a "keyword" when indexing documents and objects data about the battle in museums or libraries, then these collection items can be linked together, to Wikipedia articles in different languages about the Battle of Albert, and to other resources related to the URI via, e.g., the Linked Open Data Cloud. The global domain name infrastructure of the WWW guarantees that only the owner of the domain (here dbpedia.org) can introduce synonymous URIs, which prevents semantic confusion on a global scale. Finally, the URIs not only provide unique identifiers but also a mechanism for finding information about the identity through HTTP. For example, by typing the URI in the browser, a DBpedia page is opened with labels and content about the battle from Wikipedias in different languages, subject terms, etc. By using the Linked Data referencing mechanism, a URI can unfold into an RDF description for machine usage. It should be noted, however, that in some URI schemes, such as URN (Uniform Resource Name)[14], the identifier may not be used as an address to retrieve data related to the resource without an additional service, e.g., a URN resolver. Such services virtually duplicate the domain name infrastructure of the web that is already operational and free to use. As a result, URNs and similar URI schemes are not recommended or used by the Linked Data community. [15]

One challenge in using URIs is that there are often several URIs for a concept already in use. Using a single global ontology for URIs would be an optimal solution, but the reality is that different repositories and communities will continue using different identifiers for the same things. For example, there may be data about WWI or the Battle of Albert in Freebase[16] or in the Imperial War Museum databases, national land surveying organizations have their own

[12] Here "%281914%29" is the URL encoded representation of "(1914)".
[13] http://dbpedia.org/, the RDF version of Wikipedia content is used as the central hub in the Linked Data cloud.
[14] http://tools.ietf.org/html/rfc2141
[15] See http://www.w3.org/2001/tag/doc/URNsAndRegistries-50.html for a detailed discussion on why HTTP URIs are preferred over schemes such as URN.
[16] http://www.freebase.com/

established identifiers for places, and so on. The LD solution to address this issue is to create ontology alignments (mappings) between repositories defining what URIs refer to the same concept in different RDF stores (typically owl:sameAs mappings) or overlap in meaning.

Another challenge is that there are different ways of constructing URIs, so how does one determine what naming policy to use? There are no clear rules, but a fundamental principle for minting URIs is that "cool URIs do not change"[17]. Therefore, it is a good policy to use URIs that are neutral in terms of meaning and language. This can be a problem, e.g., in (2.1), since the URI is based on English (why not e.g. in French since the battle took place in France?), or if it is later discovered that the Battle of Albert actually took place in a year other than 1914. Any later change in the URI is a problem for systems already using the old URI. On the other hand, mnemonic names with a meaning make URIs easier to use from a human perspective.

2.1.5 Requirements for Core Metadata

We next outline requirements for a minimal metadata schema for a gazetteer of historical events, based on the requirement that metadata is used on the Semantic Web by both humans and machines.

From the human perspective, a minimal metadata schema should include the means for understanding and identifying precisely what the events in the gazetteer actually are, and for disambiguating between similar looking events. The name of an event alone does not necessarily disambiguate events, in the same way as people (e.g., different John Smiths) and places (e.g., Paris in France vs. Paris in Texas) cannot be identified by name alone. Therefore, names such as the "Battle of Albert" cannot in general be used for term labels in a thesaurus or gazetteer. For example, even in the context of WWI there are actually three distinct "Battles of Albert":

1. Battle of Albert (1914) (25–29 September 1914) - encounter battle during the Race to the Sea
2. Battle of Albert (1916) (1–13 July 1916) - opening phase of the Battle of the Somme (1916)
3. Battle of Albert (1918) (21–23 August 1918) - opening phase of the Second Battle of the Somme 1918

Furthermore, there probably have been conflicts in the Albert referenced here at other times and in other places named Albert.

Events are, from an ontological perspective, spatiotemporal phenomena that take place in time. This suggests that the following minimal metadata should be associated with each event:

1) Short **name** (in appropriate languages) for humans to identify and refer to the event.
2) **Time** of the event. Since the time of an event is often imprecise or our knowledge about it is incomplete, a flexible and rich time ontology is needed.
3) **Place** of the event. This disambiguates, e.g., phasesor parallel occurrences of an event taking place simultaneously at different locations.

[17] http://www.w3.org/TR/cooluris/

A principle of thesauri construction (Aitchison et al., 2000) is to make the terms of a thesaurus in one language unique in order to avoid problems of synonymy. In the case of an event gazetteer, such terms can be constructed in a natural way by using the pattern:

label (*time*, *place*)

In Wikipedia, only time is used as a qualifier in brackets, e.g., "Battle of Albert (1914)". However, the form "Battle of Albert (1914, France)" would probably be more useful in a more general context since places are important in characterizing events for humans, and it is possible that two distinct events with the same time designation could take place at the same time in different locations. For example, "World War I (1914-1918, France)" is different from "World War I (1914-1918, Great Britain)".

Unique full names for events with the pattern *label* (*time*, *place*) can be created, if not provided manually, by machines based on the short event name, time, and place. The full name, however, should be used by applications and ontology services as the preferred label for human consumption.

The values for time should be instances of time, i.e. RDF resources with their own structure, rather than literal date-expressions. In this way, it is possible to represent more complex expressions of time, such as intervals with uncertainty. For example, CIDOC CRM time span instances (class E52 Time-Span) can have several properties, such as a beginning, an ending, and shortest and longest duration. In the same way, the place(s) where an event occurs is expressed in terms of places with internal structure, including, e.g., labels in different languages, topological relationships, etc. In this way events can be linked with other content by shared or overlapping instances of time and place.

3 METADATA ELEMENTS FOR HISTORICAL EVENTS

This section presents a metadata model for representing historical events in an event gazetteer. The elements in the suggested model are presented in table 3.1. We decided to base our model on SKOS[18], the W3C standard for simple vocabularies. It is assumed that entries in the gazetteer conform to this standard including its integrity conditions. There are obvious reasons for grounding the vocabulary to SKOS: 1) end users are familiar with it and can more easily use standards, 2) tools such as editors, data validators, data stores, and publishing platforms can be shared and used based on standards, 3) vocabularies can be aligned and merged more easily if they are based on shared standards, and 4) standard models are usually well-crafted and already accepted, so the decision to use them in legacy systems can be made more easily. SKOS has been used especially in representing simple hierarchical thesauri. In this case narrower/broader hierarchies can be used for representing partonomies of events, i.e., decomposition of events into their sub-events. We hypothesize that the model can be applied to event gazetteers as well with appropriate extensions in RDF and OWL.

The event metadata in table 3.1 can be divided into 1) core metadata, 2) subject descriptions, 3) descriptions of the inter-event narrative structure, and 4) administrative metadata

[18] http://www.w3.org/2004/02/skos/

(documentation properties). For each element the label is presented (in English), and its meaning is briefly described, identifier URI defined (with a name space; the default is our own namespace not yet minted), cardinality constraint given (e.g., 1..n means that the element has at least one value), element value range is constrained, and actual values are characterized. Namespace skos refers to the SKOS standard, geo to http://www.w3.org/2003/01/geo/, and dc to Dublin Core.

Metadata type	**Label**	**Meaning**	**Identifier**	**Cardinality**	**Range**	**Value**
Core elements						
Name	name	Short event name	:name	0..1	Literal	string@languge
	full name	Full event name	skos:prefLabel	1	Literal	string@languge
Description	description	Description of the event	dc:description	0..1	Literal	string@languge
Time	time	Time span of the event	:time	1	Time instance	Time URI
Place	place	Place where the event took place	:place	1	Place instance	Place URI
	point		geo:point	0..1	Point instance	Point URI
	path		geo:line	0..1	Line instance	Line URI
	area		geo:polygon	0..1	Polygon instance	Polygon URI
Subject						
Subject matter	related actor	Actor involved in the in the event	:subjectActor	0..n	Actor instance	Actor URI
	related time	Other time related to the event	:subjectTime	0..n	Time instance	Time URI
	related place	Other place related to the event	:subjectPlace	0..n	Place instance	Place URI
	related event	Other event related to the event	:subjectEvent	0..n	Event instance	Event URI
	related topic	Topic concept related to the event	:subjectConcept	0..n	Concept	Topic URI
	related object	Individual object related to the event	:subjectObject	0..n	Object instance	Object URI
classification	event type		:eventType	0..n	Concept	Classification URI
Narrative						
Event hierarchy	is contained in	Larger event	skos:broader	0..n	Event instance	Event URI
	contains	Narrower event	skos:narrower	0..n	Event instance	Event URI
Event succession	next event	Next event	:next	0..n	Event instance	Event URI
	previous event	Previous event	:previous	0..n	Event instance	Event URI
Causal structure	cause	Cause of the event	:cause	0..n	Event instance	Event URI
	effect	Effect of the event	:effect	0..n	Event instance	Event URI
Administrative						
	Documentation properties	SKOS Documentation properties	skos:*xxx*	0..n	Literal	string@language

Table 3.1 A schema for event gazetteers of history

This list of elements and the model is not complete for describing semantics of events, but rather presents a set of key elements for gazetteer use that can be extended by additional descriptions in different applications. For example, events can be described not only by a text description but also by using pointers to sources such as Wikipedia pages, images, videos, etc.

3.1.1 Core elements

The core elements in Table 3.1 describe the essential properties for event identification and description, i.e., the name(s) of the event for human consumption, time when the event occurred, and place where the event took place. Place references are taken from a place ontology, and representation of geodata in terms of points, paths, and polygons (namespace "geo") is based on the GeoRDF system[19]. For representing time instances, e.g., the Time Ontology[20] or the CIDOC CRM time span system can be used.

3.1.2 Subject metadata

Subject metadata describes the event in terms of references to ontological structures (instead of using literal keywords). The meaning of the elements is similar to those used in the LIDO XML schema for a single "subject set" descriptions. When using Linked Data, it is advisable to take value references from shared domain ontologies, because shared resources link related content objects automatically with each other in the underlying semantic network. Major categories of subject annotations are:

- Actors related to the event, such as Napoleon I.
- Time related to the event, other than the time at which it actually took place. For example, a treaty made at a time may come in effect only at a later time.
- Related place, other than the place of the event. For example, a battle in colonial Africa may be related to European countries.
- Related event that is not related to the event being described by the narrative and causal metadata (discussed below). For example, the idea of the computer was invented independently by different persons (in slightly different ways) at different but related occasions.
- Topic described in terms of general keyword concepts, such as "painting", "jacket" or "declaration".
- Other individual things related to the event, such as a particular artifact, say the painting "Mona Lisa" in the Louvre or the jacket "Nelson's Trafalgar Jacket" in the National Maritime Museum), or a document (say "Declaration of Independence"). References to these things can be drawn upon from an ontology that contains instance concepts, such as DBpedia.
- Classification. Events can be classified along different dimensions. For example, the event of publishing Picasso's painting "Guernica" in Paris in 1937 can be related to both art history and political history.

Subject descriptions could in principle be represented by a single property, say "subject" of Dublin Core. However, we decided to separate different aspects of subject description in order to encourage content creators to consider explicitly the subject matter from different ontological viewpoints. Furthermore, URIs for different subject matters are drawn from ontologies of a very different kind, and it is technically simpler to provide the cataloging services needed for gazetteer construction, such as autocompletion and concept fetching, for each description field separately for each ontology.

3.1.3 Narrative metadata

Events are related to other events through narrative relations. Important general relations already in use in various event metadata models include subevent relationships decomposing events into hierarchical partonomies in time. Events also have succession relationships. For example, a war typically has successive phases that follow each other, such as declaring war,

[19] http://www.w3.org/wiki/GeoRDF
[20] http://www.w3.org/TR/owl-time/

fighting battles, and making peace, with each phase breaking down into subevents. For example, during WWI the "Race to the Sea" breaks down into a series of battles, including the Battle of Albert (1914).

In addition to hierarchical and linear relationships, events may also have causal relationships that explain why events follow each other. Representing such relationships is needed, because successive or hierarchical events are not necessarily causally related.

The model presented does allow representation of multiple narratives that share same events, without mixing the story lines. In the same way, different opinions about causality of events cannot be represented, even if history is in many ways controversial, and different authors may have different opinions about it. We assume that applications where such representations are needed would create independent narrative event graphs of their own, using references to the events of the shared gazetteer. In the gazetteer, only non-controversial generally agreed facts should be represented, e.g., that Battle of Albert is a sub-event of WWI.

3.1.4 Administrative metadata

Administrative metadata records the process and authorship of developing the ontology. Such provenance information is important for, e.g., maintaining the quality of the gazetteer.

4 AN EVENT GAZETTEER AND TIMELINE FOR WORLD WAR I

In order to test and evaluate in practice the model of event gazetteers presented, a case study project was started in the autumn 2011 aimed at developing and utilizing an event gazetteer for WWI. This project involves computer scientists from Aalto University and a library subject specialist/domain expert from the University of Colorado Boulder (CU). The primary dataset used is CU's WWI Collection Online, which comprises around 1,200 titles (over 55,000 pages) published from 1914 to 1920, originating from a variety of geopolitical regions and covering a wide range of topics. Since the impact of "total war" on civilian populations is a current area of focus for scholarly research and the topic was well-represented in the collection, we are concentrating on the civilian experience of war and occupation in Belgium in WWI.

We are not only are planning to publish the collection metadata as Linked Data but also creating a WWI event gazetteer to facilitate the annotation of and deep linking of concepts among various WWI collections and related datasets. The strategy developed is intended to be adaptable for use with other historical domains and datasets, such as those related to conflicts like the U.S. Civil War or World War II. In particular, we are also working on an ontology of Finnish history. An early version was presented in (Hyvönen et al., 2007) and is already in use in the CultureSampo[21] portal (Hyvönen et al., 2009; Mäkelä et al., 2012).

Our objective with the WWI gazetteer work is to enhance access to topics, people, places, and timeframes in the collection and create context for the documents by establishing links between data points in the collection, datasets incorporated into the project, and external data sources like DBpedia and Freebase. Among the datasets converted to RDF are the collection metadata, standard event vocabularies from the Imperial War Museum, information on German atrocities

[21] http://www.kulttuurisampo.fi/

in Belgium[22], the German army hierarchy, and Belgian geographical units. Deep linking in this subset, combined with an intelligent user interface, will allow us to demonstrate the types of complex questions that can be answered to meet user needs in a specialized domain, such as: Is the scale of the atrocity incidents involving German troops in Belgium accurately reflected in the atrocity literature? What divisions of the German army were involved in the most violent incidents? What was the geographic distribution of deportations from the Belgian provinces? We hope this functionality will lead to a richer understanding of the many forces shaping the WWI period.

Our work utilizes the existing FinnONTO ontology and ontology service infrastructure (Hyvönen, 2010), and extends it with a specialized vocabulary on WWI Belgium as well as a general framework for representing WWI events. These vocabularies are meant to be shared by other projects, providing a kind of "semantic glue" to bind separate datasets together and allowing searching/browsing between them. The semantic SAHA annotation tool combined with the HAKO search engine generator (Kurki, Hyvönen, 2010) is used for both content creation and publishing. We are also using the ARPA tool (Sinkkilä et al., 2011) to automate part of the annotation process by extracting named entities from the documents. One of our ultimate goals is to contribute history data to the Linked Open Data cloud that could serve as a hub for other WWI collections and foster interoperability among them. Development of a customized WWI web portal based on the faceted HAKO portal engine is underway to facilitate searching and browsing collection data by topics, people, places and time periods, and to represent the collection metadata in visual and interactive ways.

The resulting gazetteer will be published as an event catalog for human use; and as web service APIs and a SPARQL endpoint for machine use. The SAHA editor integrated with the publishing front-end HAKO will facilitate immediate publication of the gazetteer as an end-user application, where WWI content can be searched by using a faceted search engine, browsed using a semantic recommendation system, and visualized, e.g., via timelines and maps.

5 DISCUSSION

Several semantic models have been proposed for representing events and the relationships between them, such as EVENT (Raimond, Abdallah, 2007), LODE ontology (Shaw, 2010), and SEM (van Hage et al., 2011). The CIDOC CRM ISO vocabulary standard is based on events, associating at its top level physical things, conceptual objects, actors, places, and time spans. Event-Model-F is a foundational OWL ontology for modeling events based on DOLCE (Scherp et al., 2009). An ontology for narrative event structures was presented in (Junnila et al., 2007).

Research on developing history ontologies and markup languages, and visualizing events on timelines also exists. A model of how history can be represented in an ontology, with map visualizations, is presented in (Nagupal et al., 2005). An earlier version of our own history ontology model was presented in (Hyvönen et al., 2007) and is currently in use in the

[22] Thanks are due to John Horne and Alan Kramer of Trinity College Dublin, who gathered and analyzed the atrocity data and granted permission to include it in the project (J. Horne and A. Kramer, German Atrocities, 1914: A History of Denial [New Haven: Yale University Press, 2001], Appendix 1, 435-439)

CultureSampo portal[23]. The Papyrus project developed ontologies for bridging the gap between CH collections and their historical attributes in news archives, based on CIDOC CRM (Robertson, 2009). Visualization using historical timelines is discussed, e.g., in (Jensen, 2003; Stab et al., 2010), and knowledge representation of narratives in (Tuffield, 2006; Zarri, 2003).

Elements used in our simple event and narrative model can be found in these previous models. This paper contributes to the existing literature by first arguing that event gazetteers and domain ontologies for history are needed for semantic annotations in the same way as keyword thesauri (Aitchison et al., 2000), geographical historical and contemporary gazetteers (Southall et al., 2911), and authority files for persons and organizations (Taylor, 2006; Kurki, Hyvönen, 2009). We then developed a set of requirements for such ontologies from a Linked Data perspective, and proposed a simple SKOS-based metadata model for an event ontology for history. Applying and testing the model in the WWI case study, including content creation and application design, is still underway, and success in practice cannot be evaluated. However, the infrastructure and tools are ready to use and the initial model already contains well over 300 events. Among the datasets converted to RDF are the collection metadata, standard event vocabularies from the Imperial War Museum, information on German atrocities in Belgium[1], the German army hierarchy, and Belgian geographical units. An initial application interface exists, too. Thus far results have been encouraging.

References

Aitchison J., Gilchrist A., Bawden D.: Thesaurus Construction and Use: a Practical Manual. Europa Publications, London, 2000.

Heath T., Bizer C.: Linked Data: Evolving the Web into a Global Data Space (1st edition). Synthesis Lectures on the Semantic Web: Theory and Technology, 1:1, 1-136. Morgan & Claypool, 2011.

van Hage, W.R., Malaise, V., Segers, R., Hollink, L., Schreiber, G.: Design and Use of the Simple Event Model (SEM). Web Semantics: Science, Services and Agents on the World Wide Web 9(2), 2011.

Hyvönen, E., Alm, O., Kuittinen, H.: Using an Ontology of Historical Events in Semantic Portals for Cultural Heritage. Proceedings of the Cultural Heritage on the Semantic Web Workshop at the 6th International Semantic Web Conference (ISWC 2007), Busan, Korea, 2007. http://www.seco.tkk.fi/publications/2007/hyvonen-et-al-history-2007.pdf

Hyvönen, E., Mäkelä, E., Kauppinen, T., Alm, O., Kurki, J., Ruotsalo, T. Seppälä, K., Takala, J., Puputti, K., Kuittinen, H., Viljanen, K., Tuominen, J., Palonen, T., Frosterus, M., Sinkkilä, R., Paakkarinen, P., Laitio, J., Nyberg, K.: CultureSampo - Finnish Culture on the Semantic Web 2.0. Thematic Perspectives for the End-user. Proceedings, Museums and the Web 2009, Indianapolis, USA, April 15-18, 2009.

Hyvönen, E.: Developing and Using a National Cross-domain Semantic Web Infrastructure. Semantic Computing (Phillip Sheu, Heather Yu, C. V. Ramamoorthy, Arvind K. Joshi, Lotfi A. Zadeh (ed.)), IEEE Wiley - IEEE Press, May, 2010.

Jensen, M.: Visualizing Complex Semantic Timelines. NewsBlip Research Papers, Report NBTR2003-001, 2003.

[23] http://www.kulttuurisampo.fi/

Junnila, M., Hyvönen, E., Salminen, M.: Describing and Linking Cultural Semantic Content by Using Situations and Actions. In Robering, K., ed.: Information Technology for the Virtual Museum, LIT Verlag, Berlin, 2007.

Kurki, J., Hyvönen, E.: Authority Control of People and Organizations on the Semantic Web. Proceedings of the International Conferences on Digital Libraries and the Semantic Web 2009 (ICSD2009), September, 2009. Trento, Italy.
http://www.seco.tkk.fi/publications/2009/kurki-hyvonen-onki-people-2009.pdf

Kurki, J., Hyvönen, E.: Collaborative Metadata Editor Integrated with Ontology Services and Faceted Portals. Workshop on Ontology Repositories and Editors for the Semantic Web (ORES 2010), the Extended Semantic Web Conference ESWC 2010, Heraklion, Greece, CEUR Workshop Proceedings, http://ceur-ws.org/, June, 2010.

Mäkelä, E., Hyvönen, E., Ruotsalo, T. How to Deal with Massively Heterogeneous Cultural Heritage Data – Lessons Learned in CultureSampo. Semantic Web – Interoperability, Usability, Applicability, Vol. 3, No. 1.

Nagypal, G., Deswarte, R., Oosthoek, J.: Applying the Semantic Web: The VICODI Experience in Creating Visual Contextualization for History. Lit Linguist Computing 20(3), 2005.

Robertson, B. G.: Fawcett: A Toolkit to Begin an Historical Semantic Web. Digital Studies / Le Champ Numerique 1(2), 2009.

Raimond, Y., Abdallah, S.: The Event Ontology, 2007. http://motools.sourceforge.net/event/event.html

Shaw, R.: LODE: An Ontology for Linking Open Descriptions of Events, 2010.
http://linkedevents.org/ontology/

Scherp, A., Franz, T., Saatho, C., Staab, S.: F- a Model of Events Based on the Foundational Ontology DOLCE+DnS Ultralight. In: International Conference on Knowledge Capturing (K-CAP), Redondo Beach, CA, USA., 2009.

Staab S., Studer R. (eds.), Handbook of Ontologies (2nd Edition). Springer-Verlag, 2009.

Stab, C., Nazemi, K., Fellner, D. W.: Sematime - Timeline Visualization of Time Dependent Relations and Semantics. In: Proceedings of the 6th International Conference on Advances in Visual Computing - Volume Part III (ISVC'10), Springer-Verlag, 2010.

Southall H., Mostern R., Berman M. L., On Historical Gazetteers. International Journal of Humanities and Arts Computing, Vol. 5 (2011), pages 127-145.

Taylor, A. Introduction to Cataloging and Classification. Library and Information Science Text Series, Libraries Unlimited, 2006.

Tuffield, M., Millard, D., Shadbolt, N.: Ontological Approaches to Modelling Narrative. In: 2nd AKT DTA Symposium, 2006.

Sinkkilä R., Suominen O., Hyvönen E.: Automatic Semantic Subject Indexing of Web Documents in Highly Inflected Languages. Proceedings of the 8th Extended Semantic Web Conference (ESWC 2011). Springer-Verlag, 2011.

Zarri, G.P.: Semantic Annotations and Semantic Web Using NKRL (Narrative Knowledge Representation Language). In: Proc. of the 5th International Conference on Enterprise Information Systems, Angers, France (ICEIS 2003), 2003. pp. 387-394.