

Panel Proposal: Linked Open Data and the First World War

Rob Warren
Big Data Institute
Dalhousie University
Halifax, NS, Canada
rhwarren@dal.ca

The year 2014 marks the centenary of the First World War. This war is arguably the first “industrial” war and as a consequence, one of the first where extensive documentation is available. This abundance of scanned documents requires GLAM (Galleries, Libraries, Archives and Museums) organisations to make them available in a format that is both searchable and discoverable. Given the breadth and depth of the data available, this is not a trivial problem.

Besides rich commemorative events and lowering the cost of access to primary sources for scholars, this digital availability of large archival collections enables the creation of interlinked data-sets about the people, events and places of the Great War. This centenary also coincides with the availability of Semantic Web technologies and opportune funding that has allowed archival institutions to make their document collections available through the web.

This event is a follow up to an initial “birds of a feather” session and dinner at LODLAM2013 in Montreal, Canada. That event was primarily institutional in nature and we propose that this panel should build on the momentum generated by presenting “in use” cases for different projects and issues in creating Linked Open Data about the Great War.

We believe that this panel has value in both its timeliness as a commemoration event and as a test case of what distributed researchers, archivists and librarians can achieve using Linked Open Data. The area is much hyped but still needs a reference test case that can show the value.

Some of the issues that will be discussed with by the panellist:

1. Support for the recording of differing viewpoints about events by reconciling the underlying elements of agreement and framing an effective debate.
2. How to support the recording of partially unknown, imprecise and uncertain information by providing URI for their existence, reference and discussion.
3. Linked Open Data allows for the separation of terms that represent the *thing* from other properties such as the name, location, shape or even identity. This is an opportunity to localize information for consumption across cultures and languages.

Speakers

1. *Lasting Remembrance: Newfoundland, Labrador and War*, Kathryn Rose (kathrynr@mun.ca) and Dean Seeman (dseeman@mum.ca), Memorial University, St-Johns, Canada.

Kathryn Rose is the History Liaison Librarian at Memorial University of Newfoundland and has been working to create a unified representation of the library’s holding about the Royal Newfoundland Regiment.

2. *‘Small ontologies, loosely joined’: developing linked open data for the First World War*, Mia Ridge (mia.ridge@open.ac.uk), Trinity College Dublin, Dublin, Ireland.

1. PANEL ORGANIZATION

Mia is a research fellow with the CENDARI project at Trinity College Dublin. She is Chair of the Museums Computer Group (MCG) and a member of the Executive Council of the Association for Computers and the Humanities (ACH). She maintains a wiki of collection of museum, gallery, library, archive, archaeology and cultural heritage APIs for machine-readable open cultural data.

3. *Mapping the Western Front: using British and German Coordinates Systems to Cross-Reference Archival Resources*, Rob Warren (rhwarren@dal.ca), Big Data Institute, Halifax, Canada.

Rob Warren is a Post Doctoral Fellow at the Big Data Institute and Adjunct Professor in Mathematics and Statistics at Carleton University. His research interests focus on Linked Open Data, social network analysis, knowledge discovery in databases and ontologies. An engineer by training, he had previously held research posts in industry, government and academia both in North America and in Europe. He is a researcher on the Muninn Project which is one of the two largest Linked Open Data project on the First World War and part of the Linked Data Cloud.

4. *Europeana 1914-1918, User-Generated Content and Linked Open Data*, Jill Cousins (Jill.Cousins@europeana.eu), Europeana Foundation, The Hague, The Netherlands.

Jill Cousins is Executive Director of the Europeana Foundation and Director of The European Library. Prior to this, she worked as a researcher with her own company specialising in providing business information to large corporations. She has been involved in several international publishing industry bodies such as CrossRef and COUNTER.

1 Panel Organization

The panel will consist of a series of talks each 10 minutes in length with a 5 minute question period. A break will take place 45 minutes into the panel for coffee and or necessities as presented in Table 1.

Time	Description
15 minutes	1st presentation and questions
15 minutes	2nd presentation and questions
15 minutes	3rd presentation and questions
10 minutes	Break and informal chat
15 minutes	4th presentation and questions
20 minutes	Group panel and discussion

Table 1: Proposed timetable for panel

The last presentation will be a 20 minutes discussion panel of all of the speakers. To stimulate discussion, each panellist will be asked to respond to a series of prepared questions and/or respond to challenges from the audience.

Lasting Remembrance: Newfoundland, Labrador and War

Kathryn Rose and Dean Seeman
Memorial University
St. John's, Newfoundland and Labrador, Canada
kathrynr,dseeman@mun.ca
<http://www.library.mun.ca/nlwar.html>

Abstract

Memorial University of Newfoundland project “Lasting Remembrance: Newfoundland, Labrador and War,” integrates historical GIS data with archival resources, digital images, digitized newspapers and historical information about the Newfoundland Regiment during the First World War. The project will assist scholars, students and community members in their discovery and understanding of primary and secondary sources related to Newfoundland and Labrador’s First World War experience. It will provide scholars with unprecedented access to unique NL resources, and will equip faculty with a valuable teaching tool.

This project is one of the first major digital humanities projects that the library has undertaken. Four years ago, the library started to prepare information and investigate challenges inherent in making our data speak to other data by making it available in linked data format. The project has already taken significant historical and archival research and converted it to structured data as the initial basis of the resource.

There are three components to the portal. One component contains a map of Newfoundland and Labrador communities, as they existed during the Great War, which directs users to digitized newspapers and biographies of servicemen from those communities. Memorial University Libraries has a very strong collection of provincial daily newspapers, and has spent considerable energy and resources focusing on the digitization of issues published during the war years.

The second component will commemorate the individual soldiers who joined the Royal Newfoundland Regiment. The database will eventually contain records for all servicemen, including members of the Regiment, the Navy, and women who served overseas as nurses and members of the Voluntary Aid Detachment. Entries include references to their biographies, service, battles fought, decorations, and the cemetery where the soldier was buried.

The third component is a timeline function that includes a textual timeline and an interactive, map-based, visual timeline. Using ESRI’s ArcGIS technology, users can follow the Royal Newfoundland Regiment through the war as they move between theatres, front, and trenches. Events and areas noted on the map will offers users the opportunity to draw on descriptive essays written by faculty, related media, archival sources, and library holdings at Memorial. We hope to be able to include a crowd sourcing component, so that end-users can enrich the content of the portal.

The platform allows users to browse events and resources on a global scale, enriching this exploration with historical-spatial context not often present in lectures or text books. Students will be able to watch the progress of the war in a time-lapsed global view of battles and shifting alliances, see where Newfoundlanders and Labradorians traveled during periods of war, and how these events impacted communities throughout the province.

This presentation will further consider how this information being collected, curated, and stored will connect with other information on the Web. The presentation will contextualize this process

in terms of progression towards linked data production and participation. Additional issues will be addressed, including the conversion of our data to URIs, sources to use for existing URIs, and whether to create our own URIs for entities not represented in these sources. The benefits of linked data in terms of this project will be discussed: is the ideal of our data interoperating with other data in a linked data context worth the effort and infrastructure, or is a localized research tool sufficient?

‘Small Ontologies, Loosely Joined’: Developing Linked Open Data for the First World War

Mia Ridge
Trinity College Dublin
Dublin, Ireland
mia.ridge@open.ac.uk

Abstract

The centenary of the outbreak of the First World War has seen a vast number of related historical records digitised and published online and an increase in the number of people interested in researching the lives of ancestors who were affected by the Great War. While the availability of data varies hugely between combatant countries, many military records are limited to the date, place and battalion or regiment in which a particular soldier enlisted. The task of finding and interpreting World War One records can be daunting for the novice researcher, who must find a way to learn about military hierarchies, unit movements and engagements before they can begin to understand the experience of a particular soldier in the war.

Linked open data (interrelated data-sets published on the open web[6]) could underpin the development of tools to support historians. This paper reports on a short-term project (September - December 2014) undertaken during a visiting Research Fellowship under the First World War strand of CENDARI project, which aims to provide historians with tools for contextualising and sharing their research[1][2]. The Fellowship project aimed to create tools to provide more context for a given battalion or regiment name by providing information on related higher military units (brigades, divisions, armies and theatre of war) and the related places and activities in which they were engaged.

However, the project quickly uncovered the lack of structured, openly available data-sets online that would support such a tool. There are several reasons for this. The development of machine-readable data-sets (through application programming interfaces, or APIs) is often promised as part of museum, library and archive projects published WWI records, but sometimes fails to appear. The amateur historians and special interest groups that have created much pre-existing material on WWI do not have the skills or access to infrastructure that would allow them to link their data to formal linked data vocabularies. Those few projects that have published their ontologies (models of concepts or relationships) for WWI are often highly context-specific, developed following close reading of historical sources and designed to meet the needs of particular projects[3] [4]. Once relevant ontologies have been found, it is often difficult to obtain the contextual information that would enable another researcher to assess an ontology against their own requirements. Is this an inherent property of linked open data platforms, or can it be ameliorated with more thoughtful documentation? In 2009, David Weinburger wrote, ‘a system that is composed of lots of small ontologies loosely joined and multiple ontologies covering the same fields in different ways will capture more knowledge and be more robust than single ontologies that cover huge fields’[5]. To what extent does the state of linked open data about World War One confirm or problematise this statement?

This paper reports on the success (or otherwise) of an attempt to develop a linked open data-set of information about Allied battalions in World War One through a combination of collaboratively

developing an ontology for modelling military units, the close reading of a range of historical sources, and crowdsourcing the task of populating the resulting ontology. It will discuss the tools currently available for designing and populating ontologies and consider the impact of the chosen platform on the project and collaborators. Is it possible to capture the deep knowledge of citizen historians about often highly specialised aspects of military history and World War One with technologies currently available? In many ways, the creation of ontologies to describe WWI is a process of turning closely read little data into big data. If each of these specialised ontologies focus on a particular aspect of WWI history, can they be linked to each other in ways which respect the historical contingency and messiness of their original sources and research questions? And why have so few official WWI projects in museums, libraries and archives published the authority files or simple lists of names that underlie their projects as structured data for re-use by other projects, when it could be a relatively simple task? What tools are required to support the publication of the specialised ontologies that would help link individual WWI projects into a wider cloud of linked open data about the First World War? Finally, what tensions and expectations between the neat world of computer science and ‘big data’, and the untidy, small, localised, specialised practices of history are revealed or concealed when turning small pieces into big data?

References

- [1] Cendari. Undated. ‘First World War’. Cendari: Collaborative European Digital Archive Infrastructure. <http://www.cendari.eu/research/first-world-war-studies/>.
- [2] CENDARI WP6 team. 2014. ‘Guidelines for Ontology Building’. Cendari: Collaborative European Digital Archive Infrastructure.
- [3] Gruber, Thomas R. 1995. ‘Toward Principles for the Design of Ontologies Used for Knowledge Sharing?’. *International Journal of Human-Computer Studies* 43 (5-6): 90728. doi:10.1006/ijhc.1995.1081.
- [4] Törnroos, Juha, Eetu Mäkelä, Thea Lindquist, and Eero Hyvönen. 2013. ‘World War 1 as Linked Open Data’. Submitted for review at *Semantic Web Interoperability, Usability, Applicability*. <http://www.semantic-web-journal.net/content/world-war-1-linked-open-data>.
- [5] Weinberger, David. 2009. ‘The Dream of the Semantic Web - KMWorld Magazine’. *KM World*, March. <http://www.kmworld.com/Articles/News/News-Analysis/The-dream-of-the-Semantic-Web-52764.aspx>.
- [6] World Wide Web Consortium (W3C). Undated. ‘Linked Data’. World Wide Web Consortium (W3C). <http://www.w3.org/standards/semanticweb/data.html>.

Mapping the Western Front: Using British and German Coordinates Systems to Cross-Reference Archival Resources

Rob Warren
Big Data Institute
Dalhousie University
Halifax, NS, Canada
rhwarren@dal.ca
<http://rdf.muninn-project.org/>

Abstract

The Western Front is one of the well known locations of the Great War where several armies engaged in Trench Warfare over the course of several years.

Mapping is still at its infancy in this age, with cavalymen still carrying a sketchboard onto which they were supposed to record enemy dispositions as they skirted and flanked enemy lines. Maps, if available at all commercially, would be at a very large scale, recording only major features or towns. With the fighting now occurring within entrenched position, both British and German high commands needed small scale maps with which to coordinate operations.

The British enjoyed the advantages of defending a friendly country and obtained a set of official plates from the Belgian government in exile. From this, they extended the projection into France on an as-needed basis. The British General Staff Geographic Section grew organically from elements of the British Ordnance Survey, which meant that a group of specialists was available to produce maps standardized across all British and Dominion units. While some baffling errors remain unexplained, an imperial grid was used over a metric map, the trench map proved to be useful tool to the British and Dominion forces.

The Germans were handicapped by having to create a mapping system from scratch using whatever cartographic material was captured or confiscated. Further, each German Army had to fend for itself in terms of mapping material for its area of operation which resulted in three different mapping systems being in use, including one coordinate system that was occasionally used backwards by a neighbouring army. Eventual support through the general staff does not seem to have helped the situation, nor were the initial map sets extended past immediate tactical necessities, which may be indicative of the mindset of the General Staff.

This paper explores the notion of place, feature and geometry in the context of the Great War using Linked Open Data. In previous works [1, 2, 3], the translation of obsolete military coordinates through API's (Application Program Interface) was previously covered. We review here their use as an efficient and effective means of indexing archival documents about the war. Most war diaries, operations orders and dispatches in British and Dominions records refer to locations using both named features and coordinates. This permits the geo-referencing of each statements within a document to find the current location in question while segmenting the document according to different spatial component. This allows for distant reading of documents by the location of interest instead of as a linear document.

This geo-referencing also allows for the translation from one coordinate system to another, linking the archival sources of different organisations together through features referenced instead of topic or

keyword. Currently most of the data being referenced in this way deals with British and Canadian units owing to the limited amount of German Army documentation available digitally.

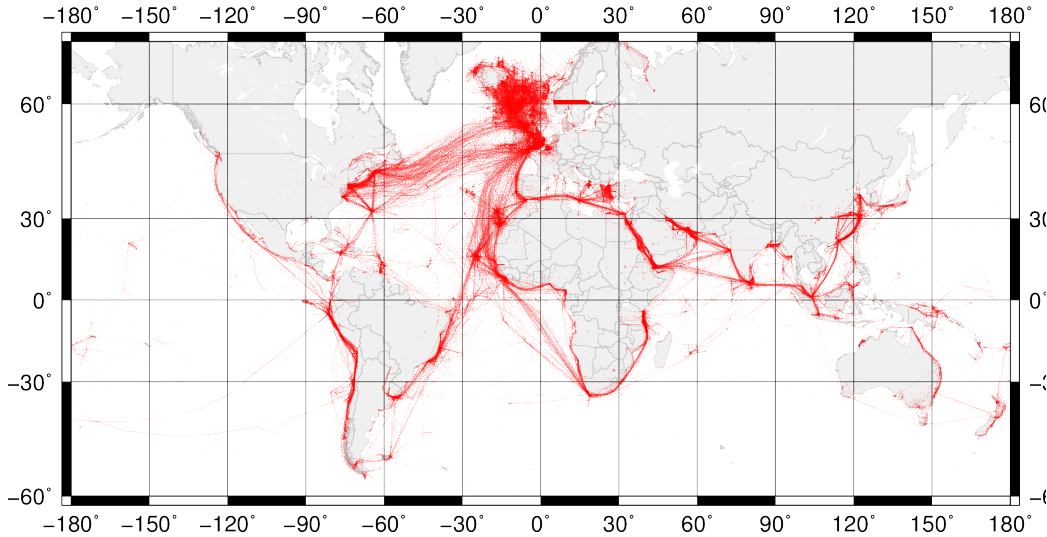


Figure 1: All known geo-referenced points within the Muninn SPARQL endpoint.

To visually explain the scale of the problem, Figure 1 plots all of the locations within the Muninn database. As the database deals with multiple units, countries and theatres of war, not all of this information is of interest to scholars and the absolute amount of data that is not relevant to the scholar's needs increases. Thus, the ability to index information geographically enables us to reduce the cost of this search for relevancy.

Unlike modern longitude and latitude references the coordinate system of the British Army was hierarchical and the length of the coordinate determined the precision of the location being given. This feature of the coordinate system in use further provides a means of searching for a particular location within documents of the Great War since through a Linked Open Data interface the specification of a large area coordinate, such as 36C.S, would contain coordinates 36C.S.22 and 36C.S.24 and retrieve documents referencing both locations.

It is interesting to note that this effectively clusters the information contained within documents by providing a relative importance to different locations within the battlefield according to each document being inspected. The talk will feature examples based on the battles of Regina trench and Vimy Ridge.

About

Rob Warren is a Post Doctoral Fellow at the Big Data Institute and lead researcher on the Muninn Project one of the two largest Linked Open Data project on the First World War and part of the Linked Data Cloud. He has published a number of vocabularies and ontologies, including the Military and Naval ontologies.

References

- [1] Robert Warren and David Evans. Translating maps and coordinates from the great war. In *Proceedings of the Terra Cognita Workshop at ISWC 2014*, Riva Del Garda, IT, October 2014.
- [2] Robert H. Warren and David Evans. From the trenches - API issues in linked geo data. In *Linking Geospatial Data Workshop*, London, UK, March 2014. World Wide Web Consortium (W3C).
- [3] Robert H. Warren and Bo Liu. Language, cultural influences and intelligence in historical gazetteers of the great war. In *Proceedings of the IEEE International Conference on Big Data 2014 (IEEE BigData 2014)*, Washington, DC, USA, October 2014. Extended Abstract and Poster.

Europeana 1914-1918, User-Generated Content and Linked Open Data

Jill Cousins

Europeana Foundation

The Hague, The Netherlands

Jill.Cousins@europeana.eu

<http://www.europeana1914-1918.eu/>

Abstract

Europeana 1914-1918¹ is an ongoing project that gathers and digitizes European resources on the First World War. Our cultural heritage material comes partly from national collections (the Europeana Collections 1914-1918 project²) and institutional film archives (the European Film Gateway 1914-1918 project³). It is also being contributed by citizens from all over Europe, who contribute their own stories and objects during history road shows organized by the Europeana Foundation, the University of Oxford and local partners such as libraries and museums. To date, the Europeana 1914-1918 portal gathers over 350,000 items, and almost 12,000 user-contributed stories. The portal also aims to provide an integrated access to resources coming from New Zealand, Australia, Canada and America, through the APIs of DigitalNZ, Trove, Canadiana and DPLA, respectively.

This presentation will explain how we have started to tackle the non-trivial problems of providing such integrated access on top of highly heterogeneous resources. Europeana 1914-1918 items are of a very different nature (films, pictures, postcards, shells, bibles...). They also come from different countries, which imply that the definitions (meta-data) are in many European languages. To provide a streamlined search and browsing experience, the partners of the Europeana 1914-1918 project have created a small controlled, multilingual vocabulary that gathers the main topics of interest they could identify, like ‘Western front’, ‘propaganda’, ‘trench life’. The creation of the vocabulary has been facilitated by the availability of the Library of Congress headings as linked data⁴. The partners can start gathering concepts - and English label data by re-using what LoC made available, and completing it by adding translations in the main languages used by the projects. These concepts are then used for browsing through the collections, while the semantic and multilingual data are used in the search engine for completion of user queries, especially their translations in other languages. Such semantic and multilingual auto-completion allows users to find in a trustful way objects, which they would have missed otherwise.

The exploitation of the vocabulary largely relies its availability through a semantic repository service, which Europeana has deployed using the OpenSKOS concept server⁵. Europeana 1914-1918 has served as a prototype for introducing in the more general Europeana portal some enhancements, that are inspired from leading semantic web-based digital heritage projects. On the other hand, as other Europeana objects, the items of the Europeana 1914-1918 benefit from the open access settings

¹<http://www.europeana1914-1918.eu/en>

²<http://pro.europeana.eu/web/europeana-collections-1914-1918>

³<http://www.europeanfilmgateway.eu/1914>

⁴<http://id.loc.gov>

⁵<http://openskos.org/>

via the main Europeana API⁶. They also undergo a semantic enrichment process that is applied to all meta-data aggregated in Europeana. For example, items are connected to related places with the corresponding co-ordinates by means of enrichment with the data made available in the GeoNames⁷ linked open data service.

We will continue similar efforts, connecting our data to other semantic resources made available in the Linked Open Data cloud, notably connecting to data-sets that can make the Europeana 1914-1918 collections serve better the needs of researchers.

⁶<http://pro.europeana.eu/api>

⁷<http://geonames.org/>