Language, Cultural Influences and Intelligence in Historical Gazetteers of the Great War

Institute for Big Data Analytics

Robert Warren
rhwarren@dal.ca
Big Data Institute
Dalhousie University

Bo Liu
boliu@dal.ca
Big Data Institute
Dalhousie University

DALHOUSIE UNIVERSITY
FACULTY OF COMPUTER SCIENCE

Abstract: Historical gazetteers trace locations that have been long forgotten while allowing for the cross-referencing of locations across different documents. In this work, we present the problem of managing a gazetteer of geometries, features and names during the Great War on the Western Front. The careful tracking of provenance information and the novel use of existing semantic web standards allows for the discovery of both the quality of the cartographic work done by both sides and the cultural influences between belligerents.

Introduction

The availability of plentiful and electronically accessible data sources is driving a renewed interest in historical gazetteers as a means of locating places that have been forgotten or changed over time. Besides providing spatial contexts to toponymy, gazetteers in the Linked Open Data (LOD) contexts can be used for information discovery and integrating different data sets.

In this poster we review some issues in handling complex historical gazetteer databases that come about when binarizing (transforming scanned images of maps into digital features) maps dating from the Great War. Complexity and size are not new issues in gazetteers but linking them through Semantic Web technologies with machine generated data is creating new opportunities in integrating multiple data sources.

A gazetteer is defined as a "geographical dictionary" in the Webster dictionary. Originally it was a simple alphabetical index that would name a location and possibly provide a locator coordinate on a separate map. In a linked digital setting, the gazetteer is expected to handle multiple maps, cultures, historical eras and possibly reach across different projects and geo-spatial databases, which have different names for the same features.

The resolution of these bodies is important is that it opens access to a number of historical documents that were not intended to reference the gazetteers such as personal letters and diaries. These documents are rich in local knowledge that is not available from official sources that focus on large scale events.

Label as a Class

An example of differing nomenclature is an German- held trench that was part of the Hindenburgh line during the Battle of the Somme in the Great War. It was named Regina Trench by the Canadian units that were attacking it while called Staufen Riegel by the defending German units.

In this case, the data obtained from both Canadian and German documents references the same feature but with different names. At the cost of additional complexity we can trace whether the name of the location implies either the German or Canadian experience of the event. Furthermore while both views concern the same military trench (feature) within time and space, the provenance of the different nomenclature is recorded as being from two different surveying sources.

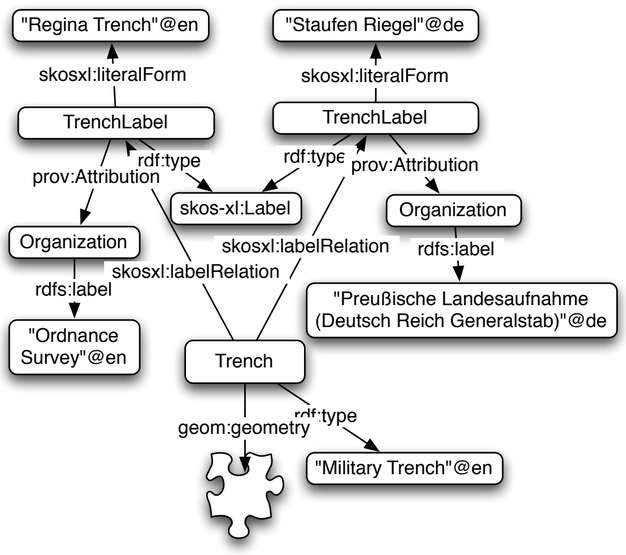

Figure 1 - The underlying linked open data structure that support multi-cultural labelling as well as provenance determination.

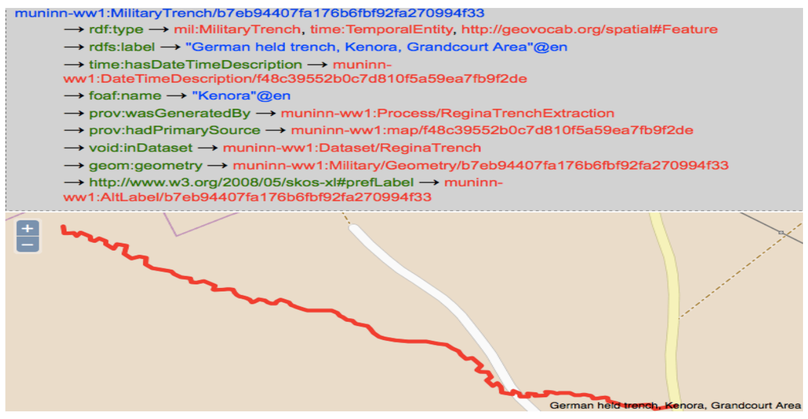

Figure 2 - The trench geometry as extracted from a British Trench Map with both German and Canadian labelling.

Sharing Geometries

Big Data, or more accurately the Big Data of online archives, means that these coping mechanisms are no longer possible. The data binarized from thousands of maps can re-position a feature over a dozen times simply because of changing survey techniques, even before change and movement is taken in account. This creates the possibility of comparing the quality of different belligerent's battlefield surveys through a query since the original geometry information is preserved.

Figure 3 - Extract from Map 36C (M-001570, IWM)

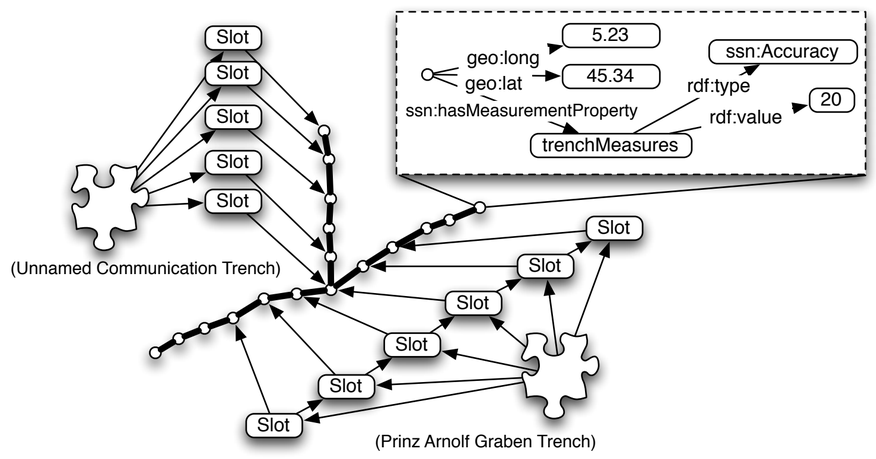

Figure 4 is a simple representation of how the individual points of the geometry polygons can be shared across different features. The reverse is also possible with different geometries being recorded for the same features along with their provenance information.

Tracking Provenance

Few cultures are insular and it is common for one organization to borrow artifacts from another. In the case of maps of the Great War it was common for armies to label their maps with the feature names of their adversaries to facilitate the orientation of their troops. The tracking of cross-cultural references within war maps gives us an insight into the intelligence apparatus of each Army.

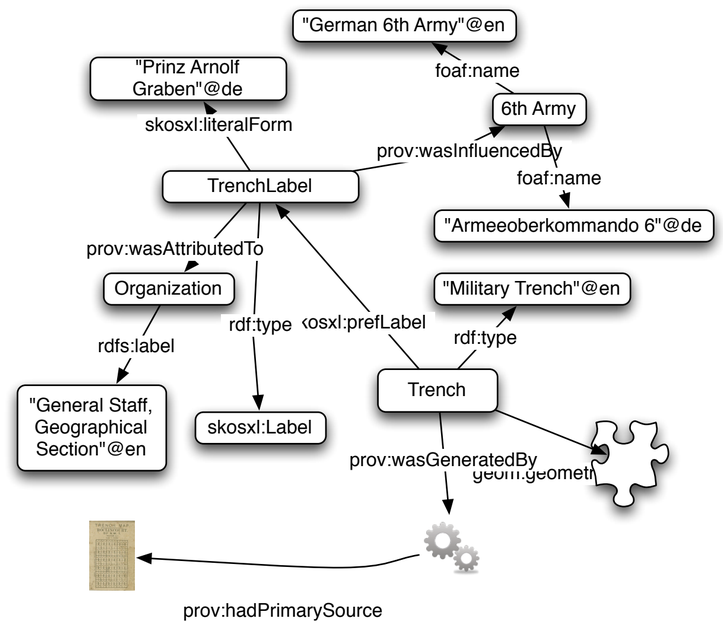

Figure 5 - A representation of the structure used to track not only the name assigned to the feature but what influenced the Geographic Section's choice of labels, which is in this case the German 6th Army's nomenclature.

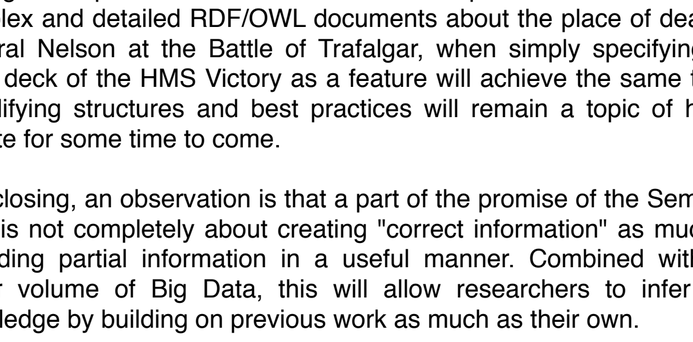

Figure 6 - The extracted trench geometry overlaid over a modern OpenStreetMap background.

Conclusion

In this paper we presented some of our current approaches to recording complex geographical changes to event data from the Great War in a manner that is not lossy.

One of the challenges of both Big Data and the Semantic Web is recording detailed information without creating a new information management problem: it is difficult to convince practitioners to write very complex and detailed RDF/OWL documents about the place of death of Admiral Nelson at the Battle of Trafalgar, when simply specifying the orlop deck of the HMS Victory as a feature will achieve the same thing. Simplifying structures and best practices will remain a topic of heavy debate for some time to come.

In closing, an observation is that a part of the promise of the Semantic Web is not completely about creating "correct information" as much as recording partial information in a useful manner. Combined with the sheer volume of Big Data, this will allow researchers to infer new knowledge by building on previous work as much as their own.

References

[1] T. Kauppinen et al., "Creating and using geospatial ontology time series in a semantic cultural heritage portal," in *ESWC*, ser. Lecture Notes in Computer Science, S. Bechhofer, M. Hauswirth, J. Hoffmann, and M. Koubarakis, Eds., vol. 5021. Springer, 2008, pp. 110– 123.

[2] OGC, "OGC GeoSPARQL - a geographic query language for rdf data," Open Geospatial Consortium, Tech. Rep. OGC 11- 052r4 OGC 11-052r4, September 2012.

[3] C. Stadler, J. Lehmann, K. Ho¨ffner, and S. Auer, "Linkedgeo- data: A core for a web of spatial open data," *Semantic Web Journal*, vol. 3, no. 4, pp. 333–354, 2012.

[4] R. Battle and D. Kolas, "Enabling the geospatial semantic web with parliament and geosparql," *Semantic Web Journal*, 2011.

[5] M. Doerr and G. Hiebel, "Crmgeo: Linking the cidoc crm to geosparql through a spatiotemporal refinement," Institute of Computer Science, FORTH, N. Plastira 1100, Vassilika Vouton, GR70013, Tech. Rep. GR70013, April 2013.

[6] S. Brown and J. Simpson, "The curious identity of michael field and its implications for humanities research with the semantic web," in *IEEE Big Humanities Data*, 2013, pp. 77– 85.

[7] D. A. Randell, Z. Cui, and A. G. Cohn, "A spatial logic based on regions and connection," in *Proceedings 3rd International Conference on Knowledge Representation and Reasoning*, 1992.

[8] L. Isaksen, et al. "Pelagios and the emerging graph of ancient world data," in *Proceedings of the 2014 ACM Conference on Web Science*, ser. WebSci '14. New York, NY, USA: ACM, 2014, pp. 197–201.

Acknowledgements

This paper follows up on gazetteer and geometry problems identified during stimulating conversations with Bettina Waldvogel and Thomas Scharrenbach at the Swiss Federal Institute for Forest, Snow and Landscape Research.