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 bipomary, 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.

Label as a Class

An example of differing nomenclature is an German-held trench that was part of the Hindenburg 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 reference the same feature but with different names. At the core of this 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.

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 conditions. The sheer volume of Big Data, this will allow researchers to infer new knowledge by building on previous work as much as their own.

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


Acknowledgements

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