

Geocoding, Publishing, and Using Historical Places and Old Maps in Linked Data Applications

Esko Ikkala¹, Eero Hyvönen^{1,2}, and Jouni Tuominen^{1,2}

¹ Semantic Computing Research Group (SeCo), Aalto University, Finland

² HELDIG – Helsinki Centre for Digital Humanities, University of Helsinki, Finland

<http://seco.cs.aalto.fi/projects/histoplaces/en/>
firstname.lastname@aalto.fi

Abstract. This paper presents a Linked Open Data brokering service prototype Hipla.fi for using and maintaining historical place gazetteers and maps based on distributed SPARQL endpoints. The service introduces several novelties: First, the service facilitates collaborative maintenance of geo-ontologies and maps in real time as a side effect of annotating contents in legacy cataloging systems. The idea is to support a collaborative ecosystem of curators that creates and maintains data about historical places and maps in a sustainable way. Second, in order to foster understanding of historical places, the places can be provided on both modern and historical maps, and with additional contextual Linked Data attached. Third, since data about historical places is typically maintained by different authorities and in different countries, the service can be used and extended in a federated fashion, by including new distributed SPARQL endpoints (or other web services with a suitable API) into the system.

Keywords: historical place, old map, linked data, crowdsourcing, geocoding

1 Relating Historical Information to Geographic Locations

Historical documents and content include references to historical places that provide an essential context for the data. However, historical places cannot necessarily be found on modern maps and gazetteers, but only on old maps from a matching time period. Dealing with historical geographical places and gazetteers³ [9] adds a temporal dimension and the notion of change to Geographic Information Systems (GIS). Many, if not most, historical places, such as Carthago or Czechoslovakia, do not exist anymore on modern maps or have at least changed substantially over the time.

Linked Data publishing principles [3] and geospatial place ontologies [1] are becoming popular in georeferencing [5], i.e., in relating information to geographic locations in information sciences. Ontologies define classes and individuals for representing geographic regions, their properties, and mutual topological and other relationships. Interoperability of dataset contents in terms of geographical places can be fostered by

³ A gazetteer is a geographical dictionary or directory used in conjunction with a map or an atlas.

sharing place resource URIs in different applications, preferably already when cataloging and annotating data.

To facilitate geographic information retrieval, data analysis, and visualization of historical data, old placenames on old maps need to be geocoded. This paper presents a solution to this with a prototype implementation supporting crowdsourced placename geocoding as Linked Data. A public service⁴ was established, integrated with Map Warper⁵, an open source map georectifying tool developed at the Public Library of New York. New place instances can be compared with existing ones in the underlying Linked Data repository (ontology) to foster reuse and in order to prevent creation of multiple instances of the same place. Metadata about the maps is stored in a Linked Data repository in similar way to places, which facilitates using maps in applications via a SPARQL endpoint.

As a pilot use case, we show how the Hipla.fi data service has been applied in creating a semantic portal for Second World War Data [8] dealing with places in pre-war and contemporary Finland.

2 Prototype Implementation: Hipla.fi

In this section we show how the Hipla.fi service is used in practice. Fig. 1 depicts the user interface, providing the end user with the following functionalities:

Searching places For finding, disambiguating, and examining historical places, there is an autocomplete search input field (a). By using the checkboxes above (b) the user can select which datasets (e.g., TGN, Suggested New Places) are included in the search results. The results are grouped based on their dataset, and they can be examined as follows:

1. Hovering the cursor over the search results shows where the places are, the corresponding marker bounces on the map.
2. Clicking a search result label or the corresponding map marker opens the info window of the place, showing its context (c).
3. Clicking the menu button on a result row (a) shows the place data in a Linked Data browser for investigating the data in detail.

Multiple dataset browsing If the user does not know the name of the place, but she has some idea where the place is located, she can pan and zoom the map view to the area. After this it's possible to use "View all places on current map view" button next to (b) on the left. This way places from different datasets connected to Hipla.fi are rendered on the map, and the user can check if the place exists already in some of the datasets. Places from different datasets are dataset-wise color-coded, which makes it possible to compare places in different gazetteers.

View on historical maps The "Maps" (b) tab provides a list of old maps that intersect the current map view. The map images are fetched from Hipla.fi's Map Warper

⁴ <http://hipla.fi>

⁵ <https://github.com/timwaters/mapwarper>

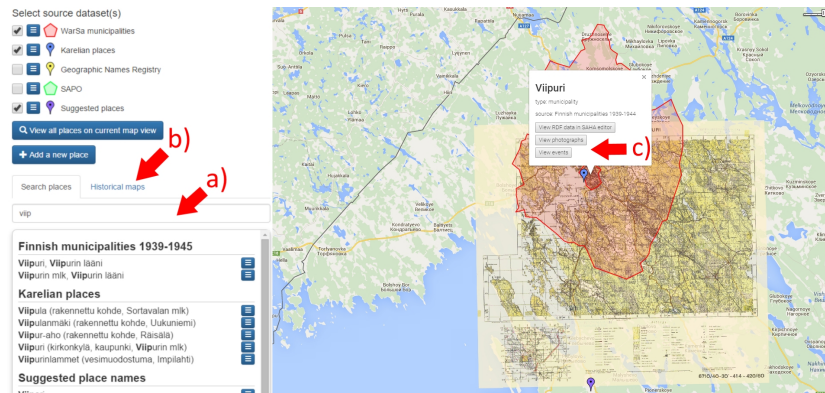


Fig. 1. Hipla.fi user interface.

georectifying service⁶ and their metadata is queried with SPARQL from the map RDF graph of the Hipla.fi service. Each map has a checkbox for rendering the map on the main map view, a thumbnail image, information about map series, scale and type, and a link to view the map in Map Warper. All map series are visible by default, but with the map series button it is possible to filter maps series-wise. Once one or more historical maps have been selected with the checkboxes for viewing, the opacity of the historical maps can be adjusted with the slider that is located on the top right corner of the map. If the user pans or zooms the main map view, clicking on the "Refresh map list" button updates the map list.

View contextual data When the user selects a place, the resource can be browsed using the Linked Data browser SAHA⁷ to see its detailed structure. Furthermore, contextual data (c) is provided connecting the place to other relevant data sources using an infobox.

Suggesting new placenames If the place at hand does not exist in any of the datasets connected to HIPLA, the user can submit a place suggestion by clicking the "Add a new place" button and filling the place details form. Coordinates for the new place suggestion can be selected from the Google map view, and it is possible to use historical map sheets for setting the coordinates. Finally the user must select the target dataset for the place suggestion. After the "Save changes" button is clicked, the new place suggestion is available for all the users of the service. This mechanism prevents the creation of duplicate place suggestions entries.

New datasets can be added to the Hipla.fi service by providing their configuration to the system. The needed information include 1) the SPARQL endpoint URL, 2) a SPARQL query for the autocomplete search, and 3) a HTML template for rendering a SPARQL result in the autocompleted result list. In addition, another SPARQL query and a HTML template can be supplied for providing contextual data for the user when a place is selected.

⁶ <http://mapwarper.onki.fi>

⁷ <http://seco.cs.aalto.fi/services/saha/>

The system was implemented using the Linked Data Finland platform⁸ [7], based on Fuseki⁹ with a Varnish Cache¹⁰ front end for serving the Linked Data. The end-user interface of Hipla.fi is a lightweight HTML5 single page map application, which provides access to multiple data sources with SPARQL queries and autocomplete search functionality using typeahead.js¹¹. Embedded Google Maps view is used to visualize historical places.

3 Application Case: An Ontology of World War II Places

This section presents an application of the Hipla.fi prototype in the WarSampo Portal¹², a system for publishing collections of heterogeneous, distributed data about the Second World War on the Semantic Web. The WarSampo Portal allows both historians and laymen to study war history and destinies of their family members in war from different interlinked perspectives.

The war zone between Finland and the Soviet Union during the WW2 was annexed to the Soviet Union after the war, and modern maps have only Soviet or Russian names, making it impossible to use modern gazetteers to describe primary source data of the war, such as photographs, articles, war diaries, etc., in which original Finnish placenames are used. To provide the missing target ontology for named entity linking of WW2 related materials, a historical geo-ontology of placenames and maps covering the war years 1939–1945 was created.

The ontology was built by combining and populating the Hipla.fi service with six data sources: 1) National Archives of Finland's map application data of 612 wartime municipalities, 2) the Finnish Spatio-Temporal Ontology describing the regions of the Finnish municipalities in different times¹³, 3) a dataset of geocoded Karelian map names (34,000 map names with coordinates and place types), 4) the current Finnish Geographic Names Registry (800,000 places), 5) Historical Senate atlas (ca. 1900), and 6) Karelian maps (1928–1951).

Named entity linking of placenames was used to automatically link [4] 160 000 photo captions, over 1000 principal event descriptions, 95 000 death records, 4500 war prisoner records, and 3400 magazine articles to geographic locations. The resulting data is available as 5-star Linked Open Data at the Linked Data Finland service¹⁴, with content negotiation, a SPARQL endpoint, and additional services for reusing the data.

Using the automatically generated links it was possible to build the WarSampo Places Perspective¹⁵ for viewing WarSampo contents on both modern and historical maps. The Places Perspective was implemented by re-using Hipla.fi user interface components.

⁸ <http://ldf.fi>

⁹ http://jena.apache.org/documentation/serving_data/

¹⁰ <https://www.varnish-cache.org>

¹¹ <http://twitter.github.io/typeahead.js/>

¹² <https://www.sotasampo.fi/en>

¹³ <http://seco.cs.aalto.fi/ontologies/sapo/>

¹⁴ <http://www.ldf.fi/dataset/warsa>

¹⁵ <https://www.sotasampo.fi/en/places/>

4 Related Work and Discussion

This paper presented Hipla.fi, a service for brokering historical places from distributed Linked Data gazetteers on historical and contemporary maps. There are several gazetteers of historical places on the web, such as The Historical Gazetteer of England's Place-names¹⁶, Gazetteer for Scotland¹⁷, the Danish service DigDag¹⁸ for finding historical administrative areas with polygons on maps, the Dutch services Gemeentegeschiedenis.nl¹⁹ and Histopo.nl²⁰, and the Alexandria Digital Library Gazetteer [6].

Thesauri of historical places, published as Linked Data, include the Getty TGN of some 1.5 million records and Pleiades²¹ [2] for ancient places. Pelagios projects²² develop APIs and GUIs for multiple historical gazetteers, such as Pleiades. DBpedia²³ contains masses of Linked Data of historical and contemporary places while GeoNames focuses on modern places. VIAF²⁴ brokers mutually aligned authority files, including historical placenames, from various national libraries around the world in Linked Data form, and from some additional open data sources, such as DBpedia and Wikidata.

The big challenge when working with placenames is that they are highly ambiguous (polysemy). There can be dozens or even hundreds of places around Finland with the same name, which presents a serious challenge for, e.g., automatic linking of events to places based on the description texts of events. Utilizing place type information is one partial solution to this problem. For example when linking the placename references in WarSampo datasets to resources in the place ontology the following order of priority was used: 1) municipality, 2) town, 3) village, 4) body of water. House names were most ambiguous, and they were not used in automatic linking. They would however be useful, if the linking is made by manually.

Another major difficulty has been that different geographic data sources, such as maps used as the basis for geocoding, are overlapping, producing multiple instances of same places. A partial solution to this issue was to remove duplicate placenames in advance, when two places shared a name, were close to each other, and had the same place type. However, there remain cases where it is not possible to differentiate between multiple placenames without manual work.

These challenges indicate that it is important to support both manual and automatic geocoding. The Hipla.fi service combines different geographic data sources into a unified view, which enables efficient search and comparison of possibly overlapping data sources.

¹⁶ <http://www.placenames.org.uk>

¹⁷ <http://www.scottish-places.info>

¹⁸ <http://www.digdag.dk>

¹⁹ <http://www.gemeentegeschiedenis.nl>

²⁰ <http://histopo.nl>

²¹ <http://pleiades.stoa.org>

²² <http://commons.pelagios.org>

²³ <http://www.dbpedia.org>

²⁴ <http://viaf.org>

Acknowledgements Hanna Hyvönen rectified Hipla.fi maps and Eetu Mäkelä contributed in creating gazetteers. Our work was supported by the Finnish Cultural Foundation and the Wikidata Project of Wikimedia Finland.

References

1. Ashish, N., Sheth, A. (eds.): Geospatial Semantics and Semantic Web: Foundations, Algorithms, and Applications. Springer-Verlag (2011)
2. Elliott, T., Gillies, S.: Digital geography and classics. *Digital Humanities Quarterly* 3(1) (2009)
3. 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*, Morgan & Claypool (2011)
4. Heino, E., Tamper, M., Mäkelä, E., Leskinen, P., Ikkala, E., Tuominen, J., Koho, M., Hyvönen, E.: Named entity linking in a complex domain: Case second world war history. In: *Proceedings, Language, Technology and Knowledge (LDK 2017)*. pp. 120–133. Springer-Verlag (2017)
5. Hill, L.: *Georeferencing: The geographic associations of information*. MIT Press (2009)
6. Hill, L., Frew, J., Zheng, Q.: Geographic names: The implementation of a gazetteer in a georeferenced digital library. *D-Lib* 5(1) (1999)
7. Hyvönen, E., Tuominen, J., Alonen, M., Mäkelä, E.: Linked Data Finland: A 7-star model and platform for publishing and re-using linked datasets. In: *The Semantic Web: ESWC 2014 Satellite Events, Revised Selected Papers*. pp. 226–230. Springer-Verlag (2014)
8. Hyvönen, E., Heino, E., Leskinen, P., Ikkala, E., Koho, M., Tamper, M., Tuominen, J., Mäkelä, E.: WarSampo data service and semantic portal for publishing linked open data about the second world war history. In: *Proc. of ESWC 2016*. Springer-Verlag (2016)
9. Southall, H., Mostern, R., Berman, M.L.: On historical gazetteers. *International Journal of Humanities and Arts Computing* 5(2), 127–145 (2011)