

Master theses

**A LOD-based georeferencing tool for historic
maps**

Alber Sánchez

Director: Dr. Simon Scheider
Betreuer: Dr. Carsten Keßler
Institute for Geoinformatics

Abstract

Historic maps represent a snapshot of a past reality no longer available, making them an interesting subject of research. Usually, maps are treated as regular text documents avoiding their specific properties but librarians acknowledge that including such information would improve search results. For achieving this purpose, they are turning to technology.

This thesis presents a software tool a library could use to georeference historic maps in order to extend traditional document searching capabilities. The inclusion of map metadata such as spatial coverage, publication date and contents open opportunities to extend the historic map's context.

This tool was developed in response to a request from the *Linked Data for eScience Services* project which is part of the *Linked Open Data University of Münster* initiative. Both project and initiative are developing tools and solutions related to the web of data including some specifically tailored for the *Universitäts und Landesbibliothek* of Münster where this new tool will be integrated in the library's workflow.

Contents

1 Introduction

As part of their foundational mission, libraries keep historic documents which play different roles in modern scientific research. Making these resources available to the public while preserving them is a challenge librarians worldwide are tackling by using modern technologies.

Due to its spatial nature, historic maps deserve special treatment as library resources in order to expose their complexities. Accessing historic maps and relate them to others (modern or historic as well) have the potential to improve and speed research. The importance of historic maps could be illustrated by quoting a few examples of famous maps:

- The "*Lienzo de Quauhquechollan*" was made in the 16th century and it is perhaps the oldest map of Guatemala. This map tells a conquest story but instead of presenting the Spanish Empire perspective, it presents the indigenous view point¹.
- The map of Napoleon's march of 1812 by Charles Minard is a valuable tool for understanding the development and consequences of such war campaign²..
- The map of 1854 cholera outbreak in London by John Snow is an example of the use of maps and statistics to describe a phenomenon³.

Why are these historic maps so important? Why are they subject of study even today? This theses can help researchers to answer these questions by allowing librarians to link those maps and their contents to other data sources. This way, these maps could be easily found and better understood helping libraries achieving their objectives.

1.1 Historic maps

In this work we are aware of the many definitions of a map, like "*a particular human way... of looking at the world*", "*a distinct mode of visual representation*", "*a mirror of nature*" or "*a mechanism for defining social relationships* and its many roles (cultural text, promotional device, instrument of sovereignty, authoritarian image) [?].

¹<http://www.lienzo.ufm.edu/>

²http://en.wikipedia.org/wiki/Charles_Joseph_Minard

³http://en.wikipedia.org/wiki/1854_Broad_Street_cholera_outbreak

Until the 90s, the accepted map definition did not include references to the map use. Map functions include [?]: information carrying, explanation of what is in a particular place, enhancing spatial knowledge, spatial knowledge communication, decision support and change of social behaviour due to map use.

However, it is common for maps being used in a different way from what the maker had in mind and it is completely up to the user to use the map in one or other way.

Map content. -Geo -Map stuff Map use. Map location (posters, atlas, book, on Tv, on computer screen, on internet)

But, for the prurpose of this work, the map is a data source which can be linked to other data in order to be discoverable by users in a library context; a map can be referenced to space and time and its content can be linked to other resources in the web. The interpretation and implications of such content is left to map users to discover.

What is the difference among a historic map and a map?

We assume the difference lies on time. Maps age until at some point, they stop being interesting for their precision

When does a map stop being a map and start being a historic map?

When does a map stop being a map and start being a historic map?

We assume in this thesis that historic map mus meet 2 conditions

Historic map uncertantly, a place for fantasy?

Because modern maps are updated constantly, did we stop creating historical maps?

maps as databases are a response to constant change and the need for precision?
A map which never grows old?

other stuff besides maps can be georeferenced as well —————

1.2 LODUM, LIFE & ULB

LODUM⁴ is the implementation by the University of Münster of *open access* and *open data* practices to its non-sensitive data. LODUM initiative is leaded by the *Institute for Geoinformatics*⁵ and the *Muenster Semantic Interoperability Lab*⁶. Together with the *Universitäts und Landesbibliothek*⁷ (ULB) they launched the LIFE project⁸ which aims to improve interdisciplinary collaboration between science and education through spatio-temporal information sharing. This project develops standards-based services and applications to capture, catalogue, manage, search, find, access and link spatio-temporal data being emphatic in linking under the belief that this allows users to make better spatio-temporal-thematic queries.

1.3 Linked Open Data

Linked Data is a set of good practices and principles for publishing structured data in the web and linking it to other data sources [?] [?] and the expected result of it is "a web of things in the world, described by data on the web" [?]. The principles consist on using Uniform Resource Identifiers (URI) for naming things, using Hypertext Transfer Protocol (HTTP) for reaching data, using standards for structuring and querying the data (like Resource Description Framework - RDF⁹ and SPARQL¹⁰) and linking the data to other data [?]. If the data is published in an openly available way then its said it is Linked Open Data (LOD).

LOD inherits Web capabilities by using already existing Web standards from webpages in data. The use of URIs and HTTP avoid data silos because they enable navigation and data discovery while the use of RDF allows the description of both data and schema (using RDF vocabularies¹¹) enabling the fusing of different data sources and expressive query capabilities. Besides, anyone can publish data and because of its interlinking, LOD is a global data graph where data is self describing since vocabularies used are RDF as well and in case of new terms, their meaning can be found at runtime just by following the term's URI [?].

What is this for historic maps?

⁴Linked Open Data University of Münster <http://lodum.de/>.

⁵<http://ifgi.uni-muenster.de/>

⁶<http://musil.uni-muenster.de/>

⁷<http://www.ulb.uni-muenster.de/>

⁸Linked Data for eScience Services <http://lodum.de/life/>

⁹<http://www.w3.org/RDF/>

¹⁰<http://www.w3.org/TR/rdf-sparql-query/>

¹¹RDF Schema is used to describe RDF vocabularies <http://www.w3.org/TR/rdf-schema/>

——AQUI VOY——

-To link stuff, stuff must exist in first place -Stuff matching -In space -In time -
In other dimensions (difficult) -Historic map as a facebook (context, understanding,
images as symbols of events not instants, pe napoleon map) -The map as a place
where space, time and stuff can meet for understanding.

SPAR ONTOLOGIES

1.4 Research question

At the same time, from Geoinformatics viewpoint, publishing historic maps raises questions like:

1. How to publish historic maps as Linked Open Data?
2. What metadata should be used for tagging historic maps?
3. What are the specific challenges when georeferencing historic maps (i.e distortions, accuracy, spatial reference systems)?
4. How can historic map content be exploited and for what purpose?

This Master thesis faces the first three question and it explores number four. For achieving this purpose, it will use Linked Open Data (LOD) as means of publishing spatial information and linking additional resources in the context of the semantic web.

-quote book -quote simon, kouppinnen, kessler -trends -Cools stuff -Bad stuff -
SPARQL? xml encoding?

————— 2 ways to deal with spatial relations explicit - LOD, the point is
inside - SPARQL query - algorithms, the relation is realized on demand - GeoSPARQL
the same with time stSPARQL

-example vocabularies

THE APPLICATION -It's a proof of concept -Features -Failures -Future work/perspectives

1.5 Existing tools

<http://dme.arcs.ac.at/annotation/> Allows comments on user-drawn features and selecting tagging. It also works for video

— <https://github.com/dbpedia-spotlight/dbpedia-spotlight/wiki> <https://github.com/dbpedia-spotlight/dbpedia-spotlight/wiki>
<http://dme.arcs.ac.at/annotation/linked-data.html> <http://dbpedia-spotlight.github.io/demo/>
<https://github.com/annotorious/annotorious/wiki/Getting-Started> <http://annotorious.github.io/annotorious/wiki/Getting-Started.html> <https://github.com/annotorious/annotorious/wiki/Getting-Started:-Standalone-Version> <http://okfnlabs.org/annotator/> <http://annotateit.org/>

— Named-entity recognition Name resolution — AUGmenting Europeana Content with Linked Open Data Resources (article) and Europeana Connect media annotation prototype (website) Open layers -Georeferencing (it doesn't seem to do re-sampling) -KML overlay -Annotation

<http://dme.arcs.ac.at/annotation/> <http://dme.arcs.ac.at/annotation/econnect-annotation-showcase.html>

It uses named entity recognition and LOD Index What about the PROJ? Annotates the image or the map? Annotates video CLodtags They're developing a module for open layers but it's not finished yet

2 The georeferencer application

3 Closing

conclusion, future work, lesson learned Spatio temporal sparql would make things easier. Follow the same patten as GeoSPARQL, include a vocabulary and operations. The operations could be Allen's algebra. Check the article and OGC - W3C for proposals.

— IT NEEDS WORK

Librarians are be the source of user scenarios which will be compiled into software as a LOD-enhanced georeferencing tool for historic maps. Regarding question number two, this thesis will find alternatives for exposing metadata standards as LOD. For doing so, existing LOD-Vocabularies will be explored and new ones will be proposed when necessary. To answer question three, a web application will be prototyped in order to georeference published while matching user stories. This application will allow LOD-tagging of spatial properties of historic maps, according to results of the first two questions. Additional data (i.e control points, accuracy measurements and the features of interest) will be added to a LOD-triple store enriching the semantic web with data about the map. Finally, the answer to question four will demonstrate if LOD is the right choice. Maps convey a large amount of information open to different interpretations depending on the observer's study field and background. It is expected using LOD allows smooth integration and interoperability of historic maps in scientific research.

Ich versichere hiermit, dass ich meine Diplomhausarbeit „*A LOD-based georeferencing tool for historic maps*“ selbstständig und ohne fremde Hilfe angefertigt habe, und dass ich alle von anderen Autoren wörtlich übernommenen Stellen wie auch die sich an die Gedankengänge anderer Autoren eng anlehrenden Ausführungen meiner Arbeit besonders gekennzeichnet und die Quellen zitiert habe.

Münster, den (Abgabedatum)

Alber Sánchez