

GeoTwain: Geospatial analysis and visualization for researchers of transculturality

Matthias Arnold, Konrad Berner, Peter Gietz, Kilian Schultes, Roland Wenzlhuemer

Cluster of Excellence "Asia and Europe", University of Heidelberg

<surname>@asia-europe.uni-heidelberg.de

Abstract

In our research cluster on transculturality many projects deal with georeferenced data. This paper offers an introduction to the new project GeoTwain that works on visualization techniques for such data. Based on an analysis of the global telegraph network as an example of transcultural research using georeferenced data, the paper derives user requirements by combining experiences gained from previous projects and specifies a set of solutions. It is the aim of GeoTwain to provide easy visualization of 4-D-information based on Google Earth and to grasp spatial relationships embedded in historical evidence to analyse, recombine and disaggregate geo-referenced historical data without having to use more specialized and highly complex GIS Tools. Envisioned visualization with GeoTwain allows for fast and efficient assessment of georeferencing's analytical potential in any given case; it also allows the user to carefully weigh further investments in data enrichment in relation to expected findings. Both the development and application of GeoTwain are embedded in the broader research environment infrastructure called Heidelberg Research Architecture (HRA).

1. Introduction

Among historians, Historical GIS is often regarded as a nice-to-have, work intensive method of visualizing interrelations that have already been discovered and proved in a "conventional" way. While this common reservation may be of some relevance in cases where only a limited amount of geo-referenced data has to be assessed, human cognitive capacity is usually not able to grasp spatial relationships and large interactive networks formed by thousands of interrelated datasets. This holds all the more true in the case of transcultural (historical) research that (a) necessitates the integration of both macro and micro levels of research and analysis and (b) emphasizes the dynamic and interactive character of its research objects, all of which are embedded in what Heidelberg University's Cluster of Excellence "Asia and Europe in a Global Context: Shifting Asymmetries in Cultural Flows" [1] conceptualized as (trans)cultural flows.

As part of the research cluster a virtual research environment – the Heidelberg Research Architecture (HRA) – is being developed with the aim to provide researchers with a set of databases and analytical tools that support their work. Within the HRA, GeoTwain is a first attempt at overcoming the problem of visualizing masses of geo-referenced data produced on different levels of transcultural research. In this context it is important to point out that GeoTwain does not at all aim at replacing established GIS software, but rather seeks to provide an intuitive and easy-to-use tool that visualizes data with a minimum of user input on a well-known platform. The project intends to contribute to that "eureka moment when somebody sees data mapped for the first time" [2, p 18]. Several of the application's functions will be encapsulated as Web Services, so that other components of the HRA will be able to make use of them. This approach follows a new paradigm known as Service Oriented Architecture [3], that is also beginning to gain ground within the e-Humanities, e.g. in the TextGrid-Project [4]. Such project architecture allows for the integration of different tools according to different needs.

2. The Scenario: Global telegraph network

One of the case studies tested in the development of GeoTwain is the visualization of the emergence of a global telegraph network in the second half of the nineteenth century. This particular case study has been selected because it fulfils both requirements mentioned in the introduction: (a) it integrates the macro level (global submarine cable network) with the micro level (national and regional domestic telegraph networks) and (b) seeks to visualize the patterns of use of global telegraphy (and therefore the information flows) in the nineteenth century.

2.1 Historical background

A few decades after the invention of electric telegraphy, a global telegraph network started to develop in the middle of the nineteenth century. The speed of communication between places connected through this network increased dramatically and eventually pushed the already ongoing process of

globalization into a new qualitative dimension. The structure of the global telegraph network, however, closely resembled the socioeconomic and geopolitical order of its time. Access to the network was unevenly distributed, from a global, as well as a local perspective. While some regions, such as Western Europe or North America, boasted a well-developed telegraphic infrastructure and privileged information access, other areas such as Asia, Latin America and particularly Africa were often only marginally connected to the global information network. Yet even within the well-connected regions, network access was primarily a privilege of the wealthy and the educated due to the high costs and specialized skills necessary to participate in global telecommunication. Accordingly, both structure and use of the global telegraph network were asymmetrical and enhanced as well as perpetuated existing inequalities.

2.2 Visualization and geo-referencing

Contemporary telegraph maps already depict at least the structural asymmetries and show that some world regions were better served by telegraph cables than others. But a detailed analysis that goes beyond the surface of simply tracing the course of the cables cannot be achieved without geo-referencing the available data on network structure and usage. There are two main reasons for this: contemporary maps (and their modern reproductions) usually focus either on a global perspective showing the routes of the main submarine cables only, or they zoom in on a specific territory, e.g. London or another part of England, and depict the local telegraph structure in great detail. They cannot, however, do both and represent the global telegraph network with its global trunk routes as well as its fine local branches in its entirety (see Fig. 1). And yet such a holistic perspective is necessary in order to properly evaluate the functioning of the network, since for its efficiency it depended on the immediate availability (locally developed branches) and the capacity to quickly convey messages over long distances (the trunk routes).

Conventional maps/visualizations fall short of such requirements. Only with the help of a powerful geo-referencing tool can we visualize both the myriad local branch lines, telegraph stations and the flows to, from and between them, together with the international routes, their varying capacities and the telegraphic traffic flowing through them. Connecting all these stations and lines allows us to conduct a thorough (Social) Network Analysis, the result of which can be fed back into the database. Only by linking such valuable metadata with the geo-references structural data we will eventually be able to gain new insights

into both the structure and the usage of the global telegraph network in the nineteenth century.

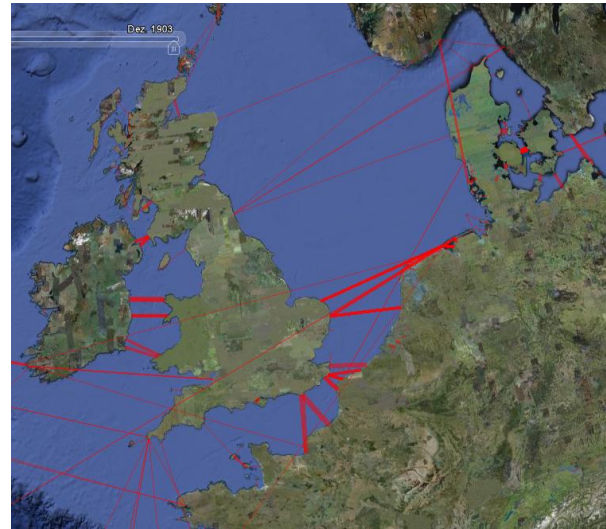


Figure 1: Submarine telegraph cable network connecting Great Britain with the world, 1903

This approach was put to the test and yielded very encouraging results in a recently published article by one of the research team members, Roland Wenzlhuemer [5]. Due to the overwhelming wealth of data and the necessity to manually identify and enter every place and every line into the database, the article could only offer results for a limited case study: London’s telegraphic development in 1868. However, there is a wealth of data available for this specific period of observation (in great detail at least for Great Britain, India and other parts of the British Empire) that would allow for a global analysis. Expanding the scope in such a way is only possible to achieve with the help of geo-referencing tools and databases that support the researchers in identifying historical place names and connecting them to existing metadata. The wealth of data along with the lack of any definite spelling or naming standards is already a problem when analysing British data only; it becomes even more difficult when investigating regions such as India.

3. User requirements

Based on the problems encountered in our research on the global telegraph network and several other transcultural research projects located at Heidelberg University – such as “The Real Fiction of Unreal Equality: Networking the International System”, “Trading Diasporas – Trading Secrets: Bans, Customs and Smuggling in the Hanseatic and Venetian Trade Systems (1291-1450)”, or the itineraries of the “Kurfürsten” (the electors) – we deduced several general user requirements.

3.1 Automatically find a place and display it on Google Earth

Getting the geographic coordinates for a location requires normalisation of its place name. For example, it is necessary to find standard modern equivalents for ancient locations and their spelling variants. Also, when working on the micro level of villages and hamlets, there are often no consistent spelling standards. The biggest challenge lies in place names that are ascribed to more than one location. In such cases, the tool needs to visualize the respective probabilities based on additional information, which is manually entered by the user and/or automatically calculated like "all places are in Great Britain". Alternatively, the data is offered in chronological order (e.g. an itinerary).

3.2 Display dynamics of change by incorporating Google Earth's Timeline-Feature

Leaving aside all political questions about 2D-projections and scale [6], GeoTwain needs to visualize shifting spatial patterns, dynamics of change and interactions between places, such as the information flows between stations in the global telegraph network, on the 3D-globe of Google Earth including its Timeline-Feature in order to function in a transnational research context. We differentiate between three main functions: (1) the visualization of findings; (2) quality control: in addition to the plain visualization of findings, GeoTwain can help to locate errors in the data. Wrong timestamps and/or places can be recognized and corrected much easier than in tables or spreadsheets; (3) new theories: the mapped display of geo-referenced data should facilitate the recognition of spatial patterns embedded in the source material, which can often not be found with data aggregated in a table or a database only, especially if there are thousands of datasets and, therefore, places to be considered. "The ability of GIS to integrate, analyze, and visually represent spatially referenced information is inspiring historians to combine sources in new ways, to make the geographical context an explicit part of their analysis, to reexamine familiar evidence, and to challenge long-standing historical interpretations." [7]

3.3 Representation of Possible Pasts – visualizing (un-)certainty

All historical research needs to reflect on the universal warning that "all visualizations of information are abstractions, which provide useful approximations of the real world." [8, p. 2024]. Depending on the source material, there are different challenges to be met in order to make the degree of

probability of any visualization explicit and avoid the danger of representing a higher precision than is justified by the historical source. The latter can often be seen in geo-visualizations [9, p. 85-88 and 1, p. 19]. GeoTwain allows the user to assess the visualized data, to leave room of alternative interpretations and to discuss other *Possible Pasts*.

(1) Accuracy and precision of geospatial information: Often the location is not a clearly defined place, but e.g. an approximation or a vaguely defined area ("near Heidelberg"). In addition, the source itself, e.g. a forged medieval document, can lack credibility and thus add uncertainty. In the case of the global telegraph network, for example, some uncertainty stems from the fact that the data has been evaluated by hundreds of different station masters – all with different standards and attitude toward the task.

(2) Accuracy and precision of various temporal ranges, the "temporal footprint": The boundaries of the linear span of time within which a certain feature exists are always fuzzy. The accuracy of dates found in historical sources, e.g. in the *Regesta Imperii* giving information about when exactly Frederick II stayed in a particular place, vary from an exact day (16 March 1217) to the notion of a year (1217) without any further specification – to say nothing of information such as "first half of the 13th century" or "in the Middle Ages".

4. Solutions

The GeoTwain project and some related activities in the core development of HRA, offer and develop a number of solutions for the above mentioned user requirements, either as web applications or as Web Services that can be combined in workflows.



Figure 2: Travels of Frederick II showing data from *Regesta Imperii* in Google Timeline

A comparatively small research project about the travels of Frederick II (1194-1250) served as the first case study to test the functionality of Google Earth's timeline feature. Two thousand place names associated with Frederick's various whereabouts were taken from

the *Regesta Imperii* database [10] and manually given coordinates. These data were converted to a KML [11] file. The result was a timeline that allows the user to follow Frederick's travels and get direct links to the entries of the *Regesta Imperii* database (see Fig. 2). With the help of this visualization it becomes clear that several contemporary entries provided false information, because too much distance was covered in too short a time.

Some fields of research within the Cluster focus on medieval European places. The related archival material is primarily written in Latin. This issue was solved by an online service based on data from *Orbis Latinus* [12] to locate Latin place names. The "geoparsing" [9, ch. 5] application converts these Latin names into their modern equivalents, looks up their geo-coordinates and displays them using Google Earth.

A more complex service can digest an entire text as input and will then produce a list of all locations featured in the text. The application can then use the visualization engine to display a Google Earth map of all featured places.

Another tool that will be tested in the telegraph scenario will check and compare location names with controlled vocabularies or authority files such as the Getty Thesaurus of Geographic Names (TGN). The TGN is especially interesting as it does not only serve as a reference for names offering spelling variants, but also positions the location within a geo-hierarchy and provides its coordinates.

A further aim of the GeoTwain project is to develop a web application that converts user-uploaded tables into the Google Earth native KML format. The data including geo-referenced place marks and timestamps can then be visualized in Google Earth making full use of the timeline functionality.

Problems of indistinctive locations can be automatically addressed by accessing information stored in the afore-mentioned databases, such as historical names tagged with chronological information on their usage and validity. Another approach to resolve such ambiguities is to pinpoint the search on specific geographical or historical regions. These spatial focuses can be activated interactively: the user chooses among numerous pre-defined regions and hierarchies defined in external thesauri or ontologies. If the tool is not able to correctly identify a location or place, all available alternatives are shown as a list or in an interactive map to let the user manually pick the correct place.

Additionally, a dynamically produced timeline can help to disambiguate locations, e.g. if one of the possible telegraph stations in the expanding network is

geographically and chronologically located between two identified stations. In the case of chronologically linear data, probabilities can sometimes even be automatically inferred by calculating distances.

The application is also able to visualize the fuzziness of converted data by applying different colours and shapes to place marks. The visualization can finally be enhanced by adding adjusted historical maps ("rubbersheeting") that users can choose as an appropriate background layer.

So far a proprietary XML format is being used within these applications, but other possible standards, especially GML [13] and ADL gazetteer protocol [14 and 9] are currently being evaluated. Since GML proves to be very flexible, but also quite complex, ADL as a more lightweight approach will be easier to implement. As ADL has been implemented in GML, it will be possible to move to a bigger subset of the latter. Both can also interoperate with KLM.

5. Conclusion

GeoTwain provides researchers in transcultural studies with an easy way to visualize 4-D-information in a Google Earth environment and allows the user to grasp spatial relationships embedded in historical evidence without having to explore the technological depths of ESRI's Arc GIS or other GIS Tools. GeoTwain does not at all aim at replacing other GIS software by trying to provide a platform for viewing a "geographically-integrated, connected world history, which can be queried at any scale from the global to the individual to improve understanding." [8, p. 2016] Rather, it seeks to open up a quick and efficient way to analyze, recombine and disaggregate geo-referenced historical data. Based on the preliminary visualizations done by GeoTwain the user can then evaluate whether the prospective results justify further efforts in GIS data processing.

We believe that the aggregation of different geo-referenced sources paves new ways of research, teaching and e-learning and hope that GeoTwain may be a modest contribution to bridge the still existing gap between historians and geographers [15, 16]: "[Historical GIS] compels writers to think graphically and forces spatial thinkers to come to grips with the subtlety of historical texts." [2, p. 3]

References

- [1] Asia and Europe in a Global Context: Shifting Asymmetries in Cultural Flows, University of Heidelberg, <<http://www.asia-europe.uni-heidelberg.de>>
- [2] KNOWLES, ANNE KELLY/HILLIER, AMY (ed.): Placing History. How Maps, Spatial Data, and GIS are Changing Historical Scholarship, Redlands, Calif. 2008.

- [3] ARSANJAN, ALI, Service-oriented modeling and architecture - How to identify, specify, and realize services for your SOA, IBM developerWorks, Nov 2004, <<http://www.ibm.com/developerworks/library/ws-soa-design1/>>
- [4] GIETZ, PETER, et.al. TextGrid and eHumanities, Second IEEE International Conference on e-Science and Grid Computing (e-Science'06), p. 133-141.
- [5] WENZLHUEMER, ROLAND: Metropolitan Telecommunication: Uneven Telegraphic Connectivity in 19th-Century London, in: Social Science Computer Review 2009, vol. 27, no. 3, p. 437-451, <<http://ssc.sagepub.com/cgi/content/abstract/27/3/437>>
- [6] MONMONIER, MARK S.: Rhumb Lines and Map Wars. A Social History of the Mercator Projection, Chicago, Ill. 2004.
- [7] KNOWLES, ANNE KELLY: Introducing Historical GIS, in: Past Time, Past Place. GIS for History, ed. by ANNE KELLY KNOWLES, Redlands, CA 2002, p. XI-XX.
- [8] OWENS, J. B.: Toward a Geographically-Integrated, Connected World History: Employing Geographic Information Systems (GIS), in: History Compass 5/6 (2007), p. 2014-2040, online <<http://dx.doi.org/10.1111/j.1478-0542.2007.00476.x>>.
- [9] HILL, LINDA L.: Georeferencing. The Geographic Associations of Information (Digital Libraries and Electronic Publishing), 2nd edition, Cambridge, Mass. 2009.
- [10] Bayerische Staatsbibliothek: Regesta Imperii online, V,1,1 Friedrich II. <http://mdz.bib-bvb.de/regesta-imperii/kapitel/ri05_fic1881_kap10>
- [11] WILSON, Tim [ed.]: OGC® KML, version 2.2.0, 2008-04-14, <<http://www.opengeospatial.org/standards/kml>>
- [12] Orbis Latinus. Lexikon lateinischer geographischer Namen des Mittelalters und der Neuzeit, Großausgabe, bearb. und hrsg. von Helmut Plechl unter Mitarbeit von Sophie-Charlotte Plechl, Braunschweig 1972.
- [13] PORTELE, CLEMENS [ed.]: OpenGIS Geography Markup Language (GML) Encoding Standard, version 3.2.1, OGC 07-036, 2007-08-27, <<http://www.opengeospatial.org/standards/gml>>
- [14] Alexandria Digital Library Project: Guide to the ADL Gazetteer Content Standard, version 3.2, February 26, 2004, <<http://www.alexandria.ucsb.edu/gazetteer/ContentStandard/version3.2/GCS3.2-guide.htm>>
- [15] BAKER, ALAN: Geography and History, Cambridge, United Kingdom 2003.
- [16] BODENHAMMER, DAVID J.: History and GIS: Implications for the Discipline, in: Placing History. How Maps, Spatial Data, and GIS are Changing Historical Scholarship ed. by ANNE KELLY KNOWLES, Redlands, Calif. 2008.