

“Visualizing Uncertainty: How to Use the Fuzzy Data of 550 Medieval Texts?”

Stefan Jänicke, Institut für Informatik, Universität Leipzig

David Joseph Wrisley, Department of English, American University of Beirut

Our research project *Visualizing Medieval Places* has brought together a computer scientist and a literary historian. We use the web-based tool *GeoTemCo* (Jänicke, 2012) to visualize thousands of place names against a focusable timeline. The resultant geospatial-temporal visualization is a way for the researcher to analyze space and time in a historical corpus of literature. The ideal user interface will allow manipulation of the visualization by (1) dynamically changing the thresholds for both visualizing and suppressing given types of uncertainty in the geospatial and temporal dimensions and (2) adding or removing facets (e.g., particular genres, time ranges) to broaden or constrain the data amount to be displayed. These means of user interaction will hopefully allow for controlled visualization of literary data, and will facilitate the formation of nuanced, supportable hypotheses about time and space in literature. Figure 1 illustrates the current user interface.

The data set is being built using a corpus of nearly 550 medieval French texts. Unlike English or Classics, scholars of medieval French have few electronic texts at their disposal. Furthermore, spelling variance of toponyms in medieval vernacular texts poses a significant challenge to any semi-automatic extraction. We are considering combining our geospatial data of French places with Latin place names to build a bilingual gazetteer for use by the digital medievalist community in the future. For now, the toponyms (and their variants) are being manually harvested from a canonical reference work, the *Table des noms propres* (Flutre, 1962); they are subsequently disambiguated and geocoded. Unlike narratologically-inspired digital literary geographies such as *The Literary Atlas of Europe* (Hurni, Piatti et al., 2012) constructed on close readings of fictional and vernacular spaces, the data model for *Visualizing Medieval Places* includes only real geographical place names. It shares more with the GIS-based analysis of unstructured texts found in the Lancaster project entitled *Spatial Humanities* (Gregory et al., 2012). Since Flutre’s work does not fully represent the variety of textual communities and genres of medieval French, we are also extracting place names from name indices in selected critical editions. The first subset of data points from Franco-Italian literature is virtually complete. The project uses the crowd-sourced *Archives de littérature du moyen âge* (Brun, 2012) to enrich the metadata about the texts.

Using the data has proved problematic since so many aspects of it are uncertain. Situating the composition of medieval texts in a specific time and place can be, at best, speculative. Date formats of traditional scholarship have been represented in idiosyncratic ways (e.g. between 1095-1291, first half of the 14th century, before 1453). Likewise, the toponyms found in these works are difficult for various reasons: they are unmappable, they can refer to multiple places, or they designate ancient Greco-Latin or medieval geographical zones no longer found on the contemporary geo-map.

The visualization of uncertainty is a hot topic in the visualization community. Despite a broad set of applications in this field, there are still no straightforward solutions for displaying multiple, overlapping kinds of uncertainty within one set of visual interfaces. Drawing upon a long list of uncertainty types (Griethe et al. 2006), a data item within our project might be said to embody two basic kinds of uncertainty. The first uncertainty is one of “lineage,” by which we mean the reliability of the text source. Certainty values for lineage can simultaneously affect the representation of data items in both dimensions, the geospatial and temporal. The second uncertainty is one of “accuracy” referring to the granularity of place or time, that is, to the distinct-sized intervals in which a value can lie. Again, granularity impacts both dimensions, the geospatial (with units such as landmarks, localities, regions, countries, continents) and the temporal (years, eras, as well as upper- or lower-bounded time declarations). Unlike Rees (Rees, 2012) who uses primarily transparency to depict uncertain information, we need to investigate multiple visual metaphors that represent several dimensions of uncertainty in a clear manner.

Visualizing distinct, overlapping geographical entities with different certainty values represents a major challenge for the project. Inspired by MacEachren’s overview of existing methods for the geospatial (MacEachren et al., 2005), we suggest testing pairwise mixtures of color hue, texture, saturation, and transparency as well as other features such as pop-up text, backgrounds, overlapping/non-overlapping shapes to encode lineage and accuracy uncertainties. In Figure 2, we use distinct shapes to encode objects with distinct geospatial accuracies. Although some work addresses the problem of temporal uncertainty for small datasets (Zuk et al., 2005), insufficient research on large-scale temporal uncertainty is available. Expecting thousands of overlapping temporal values of different granularity on the timeline, we need to create novel visualization approaches. Figure 3 illustrates one solution for dealing with aggregated temporal uncertainty where increased saturation designates a higher degree of certainty.

In our short presentation, we will demonstrate the tool and some strategies for simultaneously visualizing various aspects of the data.

The project hopes to bring attention back to hundreds of unread works of the period (Moretti, 2005)—perhaps even spawning new close readings of them based on their “interspatiality”—the referentiality of places they share with other texts—but also to encourage students and scholars to experiment with visualizing spatial clusters and patterns.

Keywords: Geospatial Visualization, Temporal Visualization, Uncertainty, Medieval French Literature, Adapting Digital Tools, Spatiality, Fuzzy Data, Lineage, Accuracy.

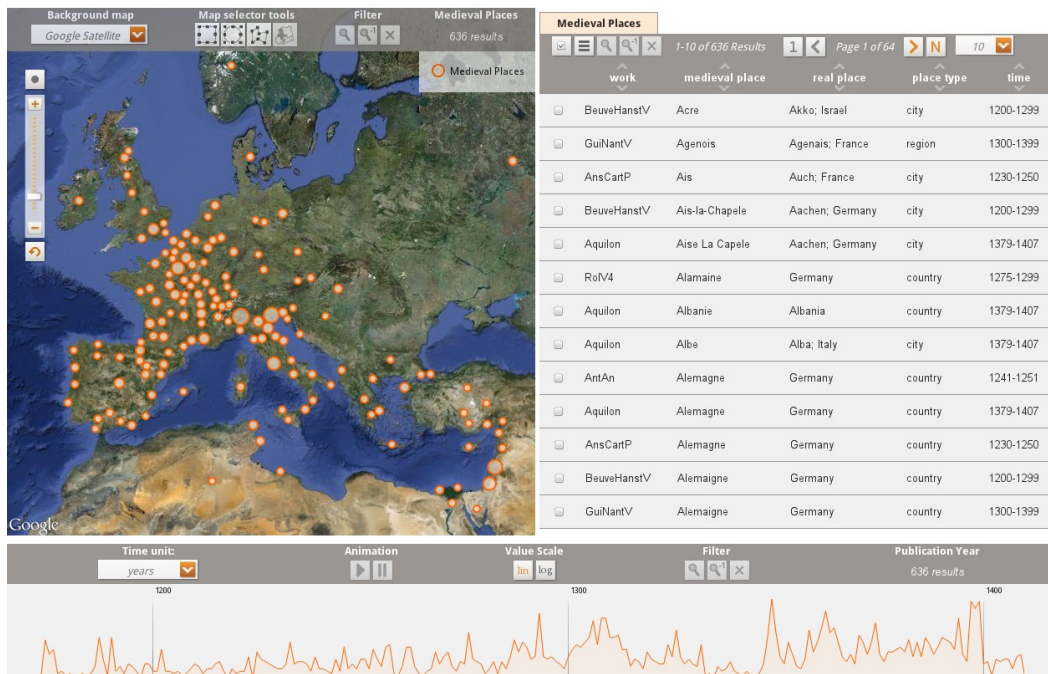


Figure 1: The current GeoTemCo user interface, showing 636 data points from the Franco-Italian dataset. NB: The time line below does not represent the real temporal data.

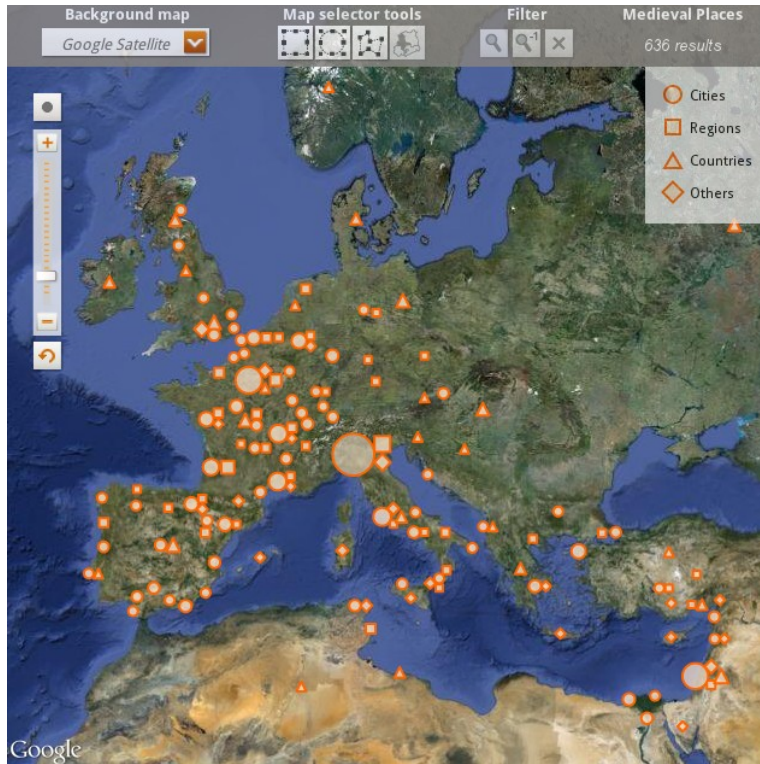


Figure 2: A visualization of 636 points from the Franco-Italian dataset, with non-overlapping shapes denoting different toponym types.

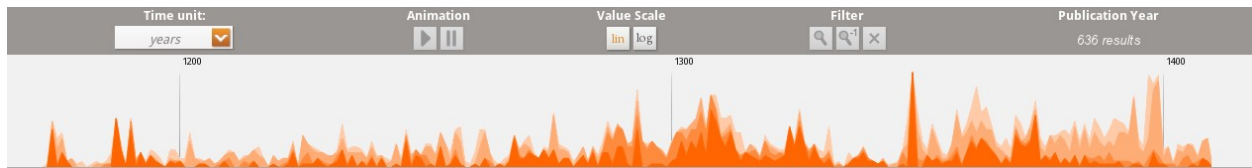


Figure 3: A timeline represented as a stacked graph with multiple aggregated uncertain temporal values (increased saturated designates increased certainty in lineage value). NB: The time line below does not represent the real temporal data of this project.

References:

Brun, Laurent. (2012) *Archives de littérature du moyen âge*. University of Ottawa <http://www.arlima.net> [accessed 28 October 2012].

Flutre, L.-F. (1962). *Table des noms propres avec toutes leurs variantes figurant dans les romans du moyen âge écrits en français ou en provençal et actuellement publiés ou analysés.*

Gregory, I. (2012). *Spatial Humanities: Texts, Geographic Information Systems and Places.* <http://www.lancs.ac.uk/spatialhum/> [accessed 15 February 2013]

Griethe, H., Schumann, H. (2006). *The Visualization of Uncertain Data: Methods and Problems.* In *Proceedings of SimVis'06*, pp. 143-156.

Harris, R. L. (1999). *Information Graphics: A Comprehensive Illustrated Reference.*

Hurni, L., Piatti, B. et al. (2013). *A Literary Atlas of Europe - Ein Literarischer Atlas Europas.* <http://www.literaturatlas.eu/> [accessed 1 March 2013].

Jänicke, S. (2012) *GeoTemCo: Comparative Visualization of Geospatial-Temporal Data* <http://www.informatik.uni-leipzig.de/geotemco> [accessed 29 October 2012].

MacEachren, A., Robinson, A., Hopper, S., Gardner, R.M., Gahegan, M., Hetzler, E. (2005). *Visualizing Geospatial Information Uncertainty: What We Know and What We Need to Know.* *Cartography and Geographic Information Science* 32, 3, pp. 139-160.

Moretti, F. (2005). *Graphs Maps Trees: Alternative Models for a Literary History.*

Rees, G. P. (2012). *Uncertain Date, Uncertain Place: Interpreting the History of Jewish Communities in the Byzantine Empire using GIS.*

Zuk, T., Carpendale, S., and Glanzman, W. (2005). *Visualizing Temporal Uncertainty in 3D Virtual Reconstructions.* In *Proceedings of the 6th International Symposium on Virtual Reality, Archaeology and Cultural Heritage (VAST'05)*, pp. 99-106.