

Using GIS for Searching Medieval Texts – Beginnings of a Project

Correlating Digitized Manuscripts and Historical Maps on the Example of Manuscriptorium Digital Library

Tomáš Klimek

National Library of the Czech Republic
Manuscriptorium Digital Library Department
Prague, Czech Republic
Tomas.klimek@nkp.cz

Abstract— This paper provides a theoretical introduction presenting possibilities and challenges of correlation between medieval texts and historical maps using GIS on the example of Manuscriptorium Digital Library and historical mappings from the area of Central Europe. It discusses methodological questions connected to this task, as well as organizational aspects of concrete suggested solution.

Keywords- GIS; historical maps; medieval manuscripts; digital libraries of historical documents; information science theory.

I. INTRODUCTION

Setting up connections between digitized manuscripts and historical maps can become an important functionality of digital libraries providing historical written documents. It can help users not only to study spatial connections and consequences of historical documents and make easier their representation but it can also provide totally new possibilities for searching and sorting information included in historical texts. The most natural way how to build connections between texts and maps in digital environment is to use GIS.

The aim of this paper is not to describe or project a complete technical solution for connecting digitized manuscripts and historical maps using GIS. It should be rather seen as an introductory theoretical input discussing some methodological and organizational aspects of this task.

The paper will discuss the topic on the concrete example of Manuscriptorium Digital Library and suggest theoretical but realizable solutions. The rest of this paper is organized as follows: Section II introduces with documents we can correlate, Section III introduces with difficulties and challenges of this correlation, Sections IV and V illustrate next practical proceedings and outline of possible final solution.

II. BUILDING MATERIAL

If we want to correlate digitized historical documents and historical maps, the first question must be – what are our sources? Do we know what to correlate?

A. Textual Documents – Content of Manuscriptorium Digital Library

This paper will discuss the topic using the example of Manuscriptorium Digital Library [1], an international digital library for historical written documents provided and coordinated by the Czech National Library. This is a digital library based upon TEI P5 metadata standards fulfilling the role of an integrator on this field. It provides data of more than 120 institutions across Europe, nowadays metadata for more than 330,000 documents, and more than 25,000 fully digitized items.

The core of the library is created by medieval manuscripts (therefore also its name), but there are also many incunabula, early printed books (printed before 1800) and maps.

B. Maps – First Detailed Mapping and Their Digital Representation

For the purposes of this paper we will need rather to think about the type of available material than to provide its complete overview. Therefore, I will briefly sum up the situation for Central Europe area only.

Concerning the oldest maps detailed enough for our purposes, for countries from Habsburg Empire, we have copies of the “Ist Military Survey“ for our disposal covering the whole area of today’s Czech Republic [2], Austria, Hungary, Slovak Republic, Slovenia and some parts of Rumania, Croatia, Ukraine and Poland [3] in year 1763 and 1787, then we can use Dritte Kursächsische Landesaufnahme for Saxony (1780–1806) [4], and e.g., Dufour Map (1809) for Switzerland [5]. All these mappings have been digitized (some of them several times) and georeferenced, therefore it is a task of cultural policy to set up a cooperation with institutions providing or possessing this data.

III. DIFFICULTIES AND CHALLENGES

A. Metodological Problems – Different Times of Origin

The first problem of correlation medieval manuscripts and historical maps is obvious. There are no maps in the modern sense of this term coming from medieval period. Although maps as such were known, we cannot speak about

modern type of maps – so called “world maps” were being used as a specific representation of theological and cosmological concepts, therefore they were not used for orientation in the landscape and their aim was not to depict and represent the terrain. We know also “itinerary maps” and later also “portulan maps” serving for special purposes only (planning timeline of the journey, navigating along the seashore), and not usable as a modern map at least in inland. It is clear that it would make no sense to use these maps for our correlation – still they do not reflect shapes of terrain in a way which would be sufficient to enable their georeferencing.

The oldest modern maps corresponding enough to the real terrain are not earlier than from the 17th century, and the oldest mappings detailed enough to enable analyses of depicted landscape not earlier than from the second half of the 18th century. This is the time much closer to ours than to the period we are used to call *the Middle Ages* (500-1500). Why to use these documents rather than today’s much more elaborated and detailed maps and map servers?

If you open any book dealing with the history of landscape, settlement, mining, warfare or any other topics demanding the analyses of the form of historical landscape, you can almost always meet with reproductions of mappings from the 18th century although the book deals with much more older history. The reason lies in the fact that these historical mappings depict the landscape before huge changes during the industrial revolution in the 19th century and even bigger changes due to large urbanization and growth of settlements as well as changing agricultural management we experienced during the 20th century.

Nobody would like to claim that there were no big changes in the landscape before the 18th century. We know several processes affecting the form of landscape not less than the processes which followed. For the area of Central Europe I could mention here at least the big colonization in the 13th century, period of destabilization and colder climate leading to abandoning one third of settlements in the first half of the 15th century, and the period of huge depopulation during the Thirty Years War in the first half of the 17th century. However mappings from the 18th century show many aspects, features and facts which have been destroyed and are invisible or hidden in our today’s landscape. Therefore, supposing the knowledge of basic changes, they can be efficiently used by scholars to study older situations.

B. Practical Problems

Dealing with medieval texts and geographical locations we have to challenge several specific difficulties. Apart from modern texts we do not have complete indexes for localities due to two facts. Authors and scribes of medieval texts used many various forms of local names for concrete settlements. Therefore we need special dictionaries providing synonyms and antonyms for older local names.[3] For many cases we will even need to look into specialized monographs dealing with some concrete situations.

At the same time we have here many locations which do not already exist. Sometimes we even do not know its original position in the landscape. Sometimes we will be able

to use localizations of such places from historiographical or archeological literature, sometimes we will be forced to exclude some localities from the correlation.

Another difficulty will be caused by variability of information types in Manuscriptorium Digital Library. We must imagine that only a small part of included information is in a text form. Images of digitized documents are of course not searchable. Then we have here so called metadata including technical metadata and descriptions, and then other types of representation - full text editions correlated to images of original documents. We cannot distinguish strictly between descriptions and full text editions parts of descriptions used for identifying the texts are taken directly from texts from original documents (so called incipits, explicits, etc.). These parts are rather data than metadata. On the other hand, other parts of descriptions refer to holders of physical documents, history of concrete items, and their physical features like origin of the writing material, type of bindings, etc. These both totally different types of information have their own value and purpose speaking about their correlation with maps, and must be sorted separately.

Another issue is caused by various coordinate systems of historical maps. However for the case of Manuscriptorium’s texts we can use maps digitized by other institutions (e.g., The University of Jan Evangelista Purkyně in Ústí nad Labem – Faculty of Environment; Research Institute of Geodesy, Topography and Cartography in Zdíby) which have already converted the old coordinate systems of some maps and georeferenced them to modern maps.

Concerning belongings of concrete places to various countries, considering the changing borders throughout the time, to avoid confusions, we should come out from the current situation.

IV. PRACTICAL STEPS DO BE DONE

Apart from starting negotiations with institutions holding data concerning historical mappings, we would have to enrich existing data of the digital library of historical written documents by XML tagging providing geographical coordinates. The reason why to use XML and e.g., not to start with a special ontology is following – surely we can use existing thesauri or dictionaries like CERL (gathering local and personal names from early printed books) for this work and try to semi automatize the process but medieval texts are unique concerning the issue of local names. [6] Many of them contain many mentions about abandoned and forgotten localities, we can meet with many variants of names and many names used for more localities, etc. therefore each text demands special care and linking to universal dictionary becomes extremely problematic. Of course separate internal ontology using RDF can be created as a usable output – maybe it can appear as necessary because of the need of representing relations between localities and their hierarchical levels; nevertheless also this internal ontology would have to be linked to concrete texts by all means and probably won’t be transferable.

For this activity there is a good chance to use crowdsourcing. As other project show, work with maps is

attractive enough to arouse interest. In this case it would be essential to develop a user-friendly editorial tool for XML tagging providing a workspace with a modern map open able with tagged text. Users would then just click on a word from the text and to the map to add necessary information; the tool would prepare all the technical work automatically.

V. OUTLINE OF POSSIBLE SOLUTION

The final result of our effort would enhance facilities of digital research environment of Manuscriptorium. Users could easily work with geographical data included in provided documents; they would be able to study or represent this data on various time layers maps which would enable them to study some mass phenomena like spreading information across the space, setting up events to the landscape in the literature (representation could reflect the sequence of localities mentioned in the text), etc., etc. At the same time – a map could be incorporated into searching of the digital library – the user could choose one or more localities or determine an area and the searching tool linked to the map environment would narrow the next user query (i.e. the system would look for documents with metadata containing at least one of localities included in the query and possible in defined fields – tags only). Processing texts in this way would also increase the searchability of maps at the same time because maps would be connected to metadata of digitized written documents. All this would again substantially increase the usability of the digital library for research.

REFERENCES

- [1] Manuscriptorium. [Online]. Available from: <http://www.manuscriptorium.com/> 2014.10.09
- [2] Presentation of old maps covering the area of Czechia, Moravia and Silesia. [Online]. Available from: <http://oldmaps.geolab.cz/> 2014.10.09.
- [3] Historical Maps of the Habsburg Empire. [Online]. Available from: <http://mapire.eu/en/> 2014.10.09.
- [4] Geoportal of Hochschule für Technik und Wirtschaft Dresden - Geoportal of Faculty of Spatial Information. [Online]. Available from: <http://geoinformatik.htw-dresden.de/> 2014.10.09.
- [5] Dufour Map of Switzerland. [Online]. Available from: <http://map.geo.admin.ch/> 2014.10.09.
- [6] CERL Thesaurus. [Online]. Available from: <http://www.cerl.org/> 2014.10.09.