Thoughts on a Web Based Co-productive Spatio-Temporal Information System

Leif Scheuermann

Abstract With the new technologies—summarized under the buzzword "Web 2.0"—the options of web based collaborations have increased strongly. In this context large projects on a voluntary basis like Open Street maps or Wikipedia have formed platforms to accumulate enormous quantities of data. Also in e.Humanities there are trends and plans towards such crowd sourced information systems and, like their big brothers, they face some basic problems: first how to bring together existing but very diverse distributed data in one system and second, how to guarantee the quality of the data—respectively, how to prevent the misuse of a system. The following paper provides some basic ideas for this discussion, without having the claim of being unique or in any way complete. The first part will cover the topic of how to implement a co-productive historical spatio temporal information system using webble technology, and the second part focuses on the question of quality management in such a system.

1 Introduction/Requirements

Since the introduction of the term "Web 2.0" by di Nucci in 1999, the idea of the collaboration of large groups of users has determined the development of web based applications. Today social media platforms such as "Twitter" or "Facebook" are as omnipresent as the different Google services like "Google maps" or "Google earth". Therefore, it is no wonder that historians (professionals as well as laymen) are trying to jump on the bandwagon, either by using the big platforms or by creating their own. In this dynamic, the question of what is already existing and actually needed is often neglected. During the last few decades numberless

L. Scheuermann (✉)
Interdisciplinary Center of E-Humanitites in History and Social Sciences, Max Weber Center for Advanced Cultural and Social Studies, University of Erfurt, Erfurt, Germany
e-mail: leif.scheuermann@uni-erfurt.de

S. Rau and E. Schönherr (eds.), *Mapping Spatial Relations, Their Perceptions and Dynamics*, Lecture Notes in Geoinformation and Cartography,
DOI: 10.1007/978-3-319-00993-3_2,

projects digitized huge amounts of data and often provided them via the internet. A central challenge for the future will be to combine these insular sources and to enable users to add their own sources as well as to ask their own questions to them. The first requirement on a comprehensive historical information system is to provide interfaces for all kinds of distributed data as well as tools for joint analysis and visualization.

Besides this more technical challenge, a major problem of these new co-productive information systems is the quality management of the data. Especially in using big data, the researchers have to rely on the accuracy of the sources. There have to be strategies to ensure the quality and to document the origin of the data. Furthermore, the objective of the data acquisition and the data model has to be apparent to ensure a correct understanding of the data. Thus a set of metadata and a metadata model has to be provided to enable the user to handle the co-productive historical spatio-temporal-information system properly.

Due to these now specified requirements, two topics shall be discussed below: on the technical side, webble technology as a middleware that is able to combine different distributed tools and data and on the methodical side, metadata and the Conceptual Reference Model of the International Committee for Documentation of the International Council of Museums (ICOM—CIDOC CRM).

2 Webble Technology

One of the main features of the "Web 2.0" technology is the combination of different web based applications in so called "mashups". Via web application programming interfaces (API), distributed media can be (re-)combined to create new content. Usually these APIs are libraries containing specifications for the data structures, methods, classes and the routines strictly defining the coverage of the interchange between the applications, which limits the compatibility. The exchange itself is only content based and the applications such as "Google Maps" etc. remain unaffected. So users can integrate their data, perform some analyses provided by the basic application and its APIs, and display them.

Here the meme-media approach[1] takes a step further, by wrapping the applications with reactive media-component representations, which act as interfaces. With this object oriented approach not only the content, but also the whole application is integrated into the meme media framework; therefore every user can adapt it for his/her own purpose and extend it with new functionality. At the same time most diverse and widely distributed wrapped data can be included without changing the original. The web based form of the meme-media approach—webble-technology (web based lifelike entities)[2]—allows the combination of wrapped objects—called

1 A basic introduction to meme-media: Tanaka (2003).

2 Basic introductions to webbles: Arnold et al. (2013), Tanaka and Kuwahara (2012).

webbles—simply via drag and drop. These combinations create new compound webbles, which can be combined with others and so on. The communication between objects is controlled by slots, hierarchically ordered, configurable bilateral stream channels, which can transport any types of data from one to one or more objects. Via these slots the receiving webbles can be directly manipulated by the sender. So the change of one webble influences all the connected webbles. To give a very basic example: the size of one webble with the shape of a rectangle can be determined through the incoming data from a second webble, which is a slide control.

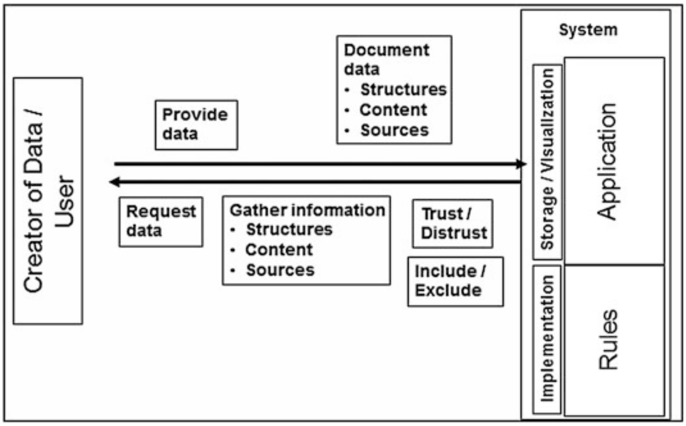
2.1 First Ideas on a Webble Based Spatio-Temporal Information System

This on first sight fairly simple working principle can now be used for the construction of highly complex compound structures like a spatio-temporal information system. But what would such a system look like? First of all, there has to be some input. To bring them into the system, input webbles are needed. Basically it doesn't matter if the data comes from local sources, from the internet, or is directly typed into a text field and it is of no relevance how complex its structures are, as long as the webble embeds it in the right slots. Next the framework needs visualization tools. For a spatio-temporal information system this means some kind of a map and a timeline, but also text fields, tables, charts and graphical representations such as network diagrams are conceivable. Because of the openness of the system, the users can even include their own custom-made visualization webble or build a compound webble from the existing ones. An interesting feature about webbles is that they are not only for the mere display of the data, but also can be used as selection tools. For example, the map can be used for the geographical, or the timeline for the temporal localization. This manipulation of one of the visualization webbles can then have a direct effect on all the others and even on the input webbles, which form new requests to the data sources. Thus the boundaries between visualization and analysis tools become permeable. But of course there should also be specific analysis tools such as spatial analysis tools and agent-based simulation models.

3 Quality Management

After this brief outline of a co-productive spatio-temporal information system, the following paragraphs will broach the issue of the quality management of the data used in the system. This topic is one of the most important problems especially for co-productive voluntary based projects, but also for frameworks using various distributed data, as, for example, when embedding web pages. At the first sight, the large public based projects seem to be self-organized and the data public property. At a second glance the self-organization is gradual and the data is always related to actors, even if these are not necessarily related to an individual. The responsibilities for the quality of the data therefore lie in three hands:

- The creator of the data, who has not only to provide the data, but also to document their origin and structure and to classify them.
- The user, who is responsible to inform himself about the origin of the data, decides to trust the data and has the possibility to include, exclude and combine data according to his personal view/choice.
- The framework of the project, with its rules and structures that coordinates the asynchronous communication.



In the end, it's the duty of the system to provide a framework for a responsible research process. It is not a question of whether or not self-organized collections of geographic data are still maps, but of how a framework can ensure this and how the individual responsibilities of creators as well as users can be warranted. Therefore, as the basis of a web based co-productive spatio-temporal framework, two particular topics have to be discussed more intensely: meta data and data models.

3.1 Meta Data

To guarantee quality standards, there has to be a catalog of supplemental information for each data set. This metadata has to be visible to the users and enable each user to trust or distrust the data set.

To be able to understand the data, a brief description of the objects, their classification and the exact circumstances under which they were collected, is needed. The reason for this is the often neglected fact, that every digitalization project has a basic purpose and aim, which determines its design. Not knowing this purpose will lead to a misinterpretation. For example, the data for one of the big databases on Roman epigraphy—the "ubi erat lupa"[3]—was mostly provided by museums. Hence all the objects preserved in the museums are included—even those not yet published elsewhere. On the other hand, because of the nature of the project, some only literarily documented monuments are not included. Mapping this data gives us an image of the spatial distribution of the inventories of the museums, but not necessarily the distribution of all the monuments found or of the Roman settlements in an area. This is explicitly not the fault of the database, but a restriction by its scope, which has to be represented in the metadata.

Next, every data set needs legal notes containing the producer of the data and the person in charge, a contact, the terms of a license for the use of the data (e.g., Creative Commons Licenses) and a list of all the sources that were used for the creation of the data sets. It must be clear, especially for the use of distributed web based data, as to who owns them and who is responsible for their validity in order to respect the copyright.

And of course there has to be a temporal and spatial localization, which provides the basic values for the requests as well as the accuracy of the data. The metadata has to include the notation of the geographic coordinate system, the encoding, the data format and a list of all attributes with a separate description and data types.

3.2 Metadata Models and Ontologies

Last but not least, one of the main problems for a spatio-temporal information system using distributed data sources lies in the different data models describing the databases. As mentioned before, if the topic and basic structures of a database are not clarified and expressed in form of metadata, there can be neither an exchange of data nor can the quality of the data be evaluated. But even with a proper documented structure, the system itself can't handle the different data structures. So the question is not only how can data be described properly but how can this description itself be understood by the framework beyond the borders of language and technologies? The answer to this is to be found in so called

[3] http://www.ubi-erat-lupa.org. Accessed 30 July 2013.

ontologies—meta data models containing rules of integrity, a specific standardized shared vocabulary and rules as to how these elements are related to each other.[4] Each ontology has a scope or domain of interest and is itself related to others via higher ranked ontologies. For cultural heritage information—mental as well as material—the most used domain ontology is CIDOC CRM (International Committee for Documentation Conceptual Reference Model)[5]—created and provided by ICOM (International Council of Museums). Since 9/12/2006 it has been official standard ISO 21127:2006 and currently includes 90 classes of entities and 152 properties[6] describing all aspects of a cultural heritage. In a first step, the data structure of the sources must be understood and formalized with the entities and their relationships. The result will be a diagram imaging the data structure of the source, which is to be encoded, for example, in a XML file or a RDF schema and connected to the data sets. Now the information of the sources is compatible with any other sources using the same ontology and the framework in which they are embedded has now the possibility to correlate them to each other properly. Concurrently the data of the sources remains untouched in its structures and form. It can even be left physically at its server of origin, as long as a wrapper combines the embedded data with the descriptive files.

4 Conclusion

In the previous pages a few ideas on how to design and develop a co-productive spatio-temporal information system have been presented. The text does not make any claim to be complete or comprehensive, but wants to broach the issue of some important requirements and possible solutions. On a methodological side, this is the question of quality management and compatibility of distributed data stored in different environments and on a more technical side, the design and composition of a webble based spatio-temporal information system.

The implementation of such a system is still to be done and the next few years will show in which frame it can be realized. The necessity to get away from the traditional monolithic insular applications—local or web based—is unmistakable. The next step in the development of a computer based historical science is to bring the amount of digitized sources which has already become immense and the various analysis tools together in a dynamic and open manner. The presented ideas on a spatio-temporal framework might be helpful on that way.

---

4 A more detailed introduction to ontologies in general and CIDOC CRM with further literature: Scheuermann (2006).

5 http://www.cidoc-crm.org/. Accessed 30 July 2013.

6 CIDOC CRM Special Interest Group (2013).

References

Arnold O, Spickermann W, Spyratos N, Tanaka Y (2013) Webble technology. In: Proceedings of first webble summit, WWS 2013, Erfurt, Germany, 3–5 June, 2013. Springer, Berlin (Communications in Computer and Information Science 372)

CIDOC CRM Special Interest Group (2013) Definition of the CIDOC Conceptual Reference Model Version 5.1 (May 2013). Downloadable at: http://www.cidoc-crm.org/docs/cidoc_crm_version_5.1-draft-2013May.pdf. Accessed 30 July 2013

DiNucci D (1999) Fragmented future. Print 53(4):32

Scheuermann L (2006) Ontologien in den historischen Wissenschaften. Hist Soc Res 32:308–316

Tanaka, Y (2003) Meme media market architectures: knowledge media for editing, distributing and managing intellectual resources. IEEE Press, Piscataway, Hoboken

Tanaka Y, Kuwahara MN (2012) Webble world user's and developer's guide. Downloadable at: http://cow.meme.hokudai.ac.jp/docs/WebbleWorldManual.pdf. Accessed 30 July 2013

Springer

http://www.springer.com/978-3-319-00992-6

Mapping Spatial Relations, Their Perceptions and Dynamics
The City Today and in the Past
Rau, S.; Schönherr, E. (Eds.)
2014, XII, 214 p. 65 illus., Hardcover
ISBN: 978-3-319-00992-6