Qualitative and Quantitative Methods in Libraries (QQML) 2: 199 – 204, 2012

On handling geographic data of paper and digital forms in academic libraries: the role of ontologies

Lila Theodoridou[1], Dimitris Kotzinos[1 & 2] and Zoe Sotiriou[1]

[1]Department of Geoinformatics and Surveying TEI of Serres, Library TEI of Serres Greece,
[2]Institute of Computer Science Foundation for Research and Technology – Hellas Greece.

Abstract: The last few years the availability of geographic data in various formats in academic libraries is increasing, bringing back in the spotlight issues that might not have been the focus of library related research for the past few years. Moreover the diversity of the available formats those data can be in, ranging from old-fashioned paper maps to digital maps and from satellite images to digital cartographic data, increases the complexity of the problem at hand. All these data cannot be considered anymore out of scope for the libraries since they are tightly related with the rest of information maintained by them, especially in academic settings where they can have an impact both on teaching and research. On the other hand, the so called digital libraries initiatives have brought into the picture the wide use of ontologies and semantic models in order to facilitate the better understanding among the librarians, the users and the expanded possibilities of using the material itself. The focus of the work discussed in this paper is to analyze and present ontology based solutions that would allow academic libraries to combine current or future semantic based catalogues with ontologies that describe the spatial characteristics of such items. The tools to support such implementations become slowly but increasingly available and this makes its implementation more apparent for the academic libraries of the (near) future.
Keywords: digital libraries, ontologies, semantics, map, geographic information

1. INTRODUCTION

Geolibrary is a library that contains geographically referenced information, i.e. information that has a geographic extension and relates to a specific "area". The development of the geolibraries is shaped in the framework of the global economy and the quest for "fast and easy access to the information" and since this information gets more and more tied to the space and the time it takes place. Actually nowadays, more and more data with a geographic extension or reference are available but they do not necessarily have a uniform nature. Thus we can include in these data of interest items like paper maps and old

Received: 3 July 2012 / Accepted: 26 September 2012 ISSN 2241-1925

(architectural or topographic) sketches but also digital data in various formats and shapes that can constitute or be included in a map. But why are geo-data becoming increasingly important for our libraries? Mainly because more and more information that is catalogued in a library is based on or refers to them. Additionally more and more geographic related data are becoming available and need to be stored and be accessible in the same area where other pieces of information can be found and retrieved in the same way that the rest of the library information is retrieved.

So a library can posses, register, manipulate and curate such kind of data; then the follow-up question should be how it would do it. Using standard cataloguing techniques is always one of the ways to go. But this cannot be enough. One aspect under consideration here can be the fact that geolibraries are by nature distributed since a lot of information has a topical aspect; thus it is much easier to find information about a place in the "local" library than in remote libraries. In that sense another interesting challenge could be to unify this kind of disperse information in a way that would be transparent to the users.

Recent advances to information science usually found under the auspices of "digital libraries" and semantic web initiatives can give us the necessary conceptual and informational tools in order to respond to the challenges described above. In that sense we can use ontologies and other conceptual schemas to integrate the information space and semantic web query languages and tools to retrieve and manipulate this information.

This paper is structured as follows: next section provides a brief overview of efforts in the area and the one after that discusses metadata and the semantics around geolibraries. Finally we draw some conclusions and provide some pointers for future work.

2. RELATED WORK

There have been various efforts throughout the years in order to support integration of geodata into academic libraries both at the conceptual and the implementation level. Various issues have been raised since 1994 (Coxe and Fitzpatrick, 1994). The most notable effort – and one of the pioneering ones in the field – has been the Alexandria Digital Library (ADL) project[1], started in 1994. The project is hosted by the Map and Imagery Laboratory of the Davidson Library of the University of California, Santa Barbara, USA and has greatly contributed to the enhancement of the field since it can show contributions in various areas like the conceptual modelling of the domain, the wide use of a gazetteer service, which contains more than 6.5 million records and even the introduction of the Alexandria Digital Earth Prototype (ADEPT)[2]. The Gazetteer is of special interest to us since it includes a complex feature-

[1] http://www.alexandria.ucsb.edu/

[2] http://www.alexandria.ucsb.edu/research/learning/index.htm

type thesaurus, which includes the notion of the Geographic Namespace (i.e. a spatial partition of a region into uniquely named sub-regions) and is highly related to the modern ontology based digital libraries (see Figure 1). The digital library has been built according to MARC[3] (Machine Readable Cataloguing) and the FGDC (U.S. Federal Geographic Data Committee's) Content Standard for Digital Geospatial Metadata[4].

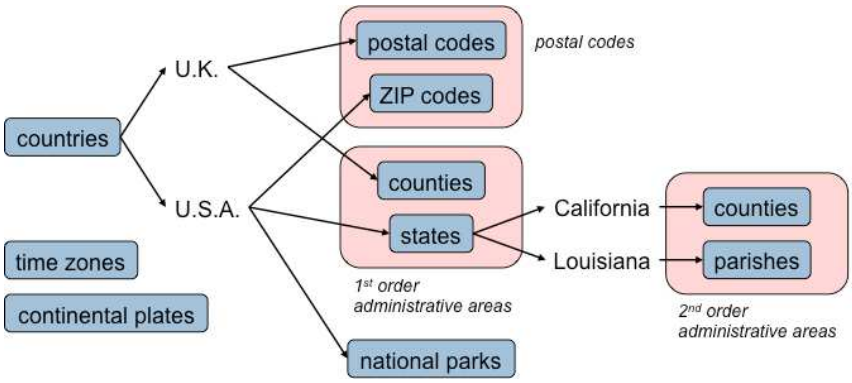

Figure 1. The conceptual structure of the ADL gazetteer[5]

Besides this there have been other notable efforts around the world with the most notable being in the US the efforts by the University of Washington and North Carolina State University (Abresch et al., 2008). In Greece the most notable effort is initiated by the University of the Aegean. Other efforts have been undertaken by WAML (WAML, 2009) and some are also described here (Shawa, 2001), mostly from a software perspective.

3. METADATA, GEOLIBRARIES AND THE SEMANTIC WEB

Metadata play a crucial role in digital geolibraries; they are the main means of information integration and are used as the way to record and subsequently identify properties of the data that can be of interest and can also be used to provide a common ground for data coming from different sources. In academic geolibraries metadata standards that are coming both from the libraries' scientific communities and from the geospatial scientific community should be applied since the effort is to bring those worlds together and make data usable in all environments.

[3] http://en.wikipedia.org/wiki/MARC_standards
[4] http://www.fgdc.gov/metadata/csdgm/
[5] http://www.alexandria.ucsb.edu/gazetteer/

Various efforts through the years have dealt with the issue of managing metadata for digital geolibraries, thus resulting to the proposal of a handful of related standards.
One of the earliest ones is the "Content Standards for Digital Geospatial Metadata", created from "Federal Geographic Data Committee" (FGDC). Another one is the ISO-TC 211[6], which is a generic ISO standard on metadata Also metadata standards on spatial metadata are highly relevant. The most interesting from them include the FIPS 173: Spatial Data Transfer Standard (SDTS)[7], DIGEST[8], the TIFF and GeoTIFF format and the more generic file formats HDF[9] and netCDF[10].
A special note should be made here at the interoperability standards from OGC, which has greatly enhanced the ability to exchange data in geolibraries since they have provided a common layer that allows data and metadata to be exchanged seamlessly.
Since the efforts of integrating geolibraries into current digital academic libraries are based on handling properly data descriptions (i.e. metadata), linking them together and providing a common metadata infrastructure and since we need geolibraries to also be available online then there a very close relation between geolibraries and current efforts of the semantic web. Semantic web can be seen as an effort to provide unified descriptions for any kind of data based on semantic structures with either soft or stricter rules that we usually call ontologies. Especially when we want to integrate metadata based on descriptions referring to different schemas then semantic web provides a unified way to store, update and query these metadata. Thus using semantic web advances seems like a well-suited choice, giving the geolibraries the ability to utilize an available (at various degrees) infrastructure and conceptual modelling primitives.
The Semantic Web gives us the ability to provide rich semantic organization for our geodata, allows the use and integration of different description schemas and different conceptual models and provides resource descriptions. But mainly it is valuable for the geolibraries since allows integration of different conceptual models based on the use of RDF/S[11] at its base model for all. The RDF Schema defines the basic common concepts like classes, properties, sub- and super- classes and their relationships, including some simple rules. RDF is a framework, mostly a description language that allows for describing any kind of resources (Figure 2) like objects, literals, relationships, etc. All these are also based on our capability to provide unique URIs (Uniform Resource

[6] http://www.isotc211.org/
[7] http://mcmcweb.er.usgs.gov/sdts/standard.html
[8] https://www.dgiwg.org/digest/
[9] http://en.wikipedia.org/wiki/Hierarchical_Data_Format
[10] http://www.unidata.ucar.edu/software/netcdf/
[11] http://www.w3.org/TR/rdf-concepts/

Identifier) that would allow distinguishing one object from another. Luckily this is not a problem in geolibraries since libraries in general provide unique ways to identify different items – one thing that is needed would probably be to extend that to cover geodata. Thus we can provide unique URIs for each item that can help us collect all the related information for this item to what we call a *resource description* (Figure 2).

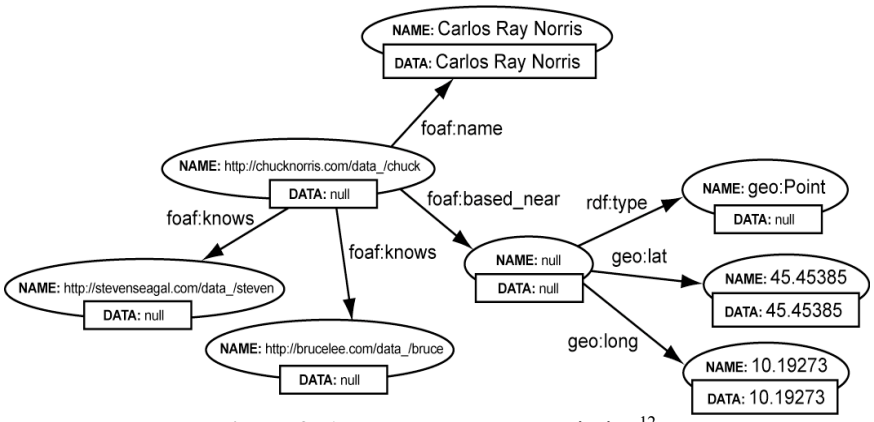

Figure 2. A RDF Resource Description[12]

Another interesting aspect and highly relevant in this discussion is the ability to navigate through the semantically described information regardless of its nature. In this area interesting efforts include the ability to provide semantic hyperlinks for moving from one area of the geolibrary to the other or use concept maps or topic maps to identify collections and relationships of items.

But what can one expect by integrating or at least bringing together geolibraries and the Semantic Web? One of the first gains would be the ability to extend information provided by the geolibrary by providing the capability to ask "smart" queries. This means that we can ask not only for keywords, authors or specific titles but we would be able to to connect information about the author, the place, the keywords, etc We would also be able to to extend or restrict information based on the meaning it carries. For example one could consider places that are parts of other places, if you are looking related info on e.g. Chania, maps of Crete might be another source of information. This procedure becomes now automated and the inference should be done by the system exploiting the georeferenced information.

But some more steps are needed if we want to achieve a successful integration of geolibraries and the semantic web:

- We need to prepare ontologies to describe georeferenced information

[12] http://milicicvuk.com/blog/page/2/

- We need to use semantic web tools and languages (SPARQL, RQL) to perform queries instead of databases and SQL
- And we need to provide prototype systems to evaluate actual deployment

4. CONCLUSIONS

In this work we tried to establish the need for Geolibraries, either as independent entities or as integral parts of the existing academic libraries. These efforts are highly connected to the Digital Libraries and the Semantic Web initiatives and provide contributions to the standards that define the metadata that are being used. Extensions on existing schemas are needed in order to fully respond to whatever the Geolibraries' world needs. Moreover we have discussed that current advances in the area of the semantic web can help provide an integrated and rich access to Geolibraries, adding information on both data and semantics and allowing the academic researcher to better understand where the georeferenced information refers.

These help us to better position the use of such data in the everyday library life by cataloguing them not just as items but also linking their spatial references to the rest of library resources providing not only thematic based catalogues but also spatially enabled catalogues. In that sense both concept based taxonomies of the Geoinformatics field but also geographic based ontologies (that define geographic entities) and library initiated cataloguing systems can be combined to catalogue, archive and retrieve the necessary items.

REFERENCES

Abresch John, Hanson Ardis, Heron Susan, Reehling Pete, (2008). *Integrating Geographic Information Systems Into Library Services: A Guide for Academic Libraries*. Idea Group Inc (IGI), 2008. ISBN 1599047268, 9781599047263

Coxe C., Fitzpatrick G., (1994). *Exploring the Future of Map Libraries*, http://www.loc.gov/loc/lcib/94/9407/map.html

IFLA, (2009). *http://www.ifla.org/VII/s6/index.htm*.

Shawa T. W. (2001). *Creative Ways to Make Geospatial Data Accessible in Academic Libraries*, ESRI 2001 User Conference Proceedings,http://gis.esri.com/library/userconf/proc01/professional/papers/pap427/p427.htm

WAML, (2009). *http://www.waml.org/wmlpubs.html*