Thesauri Design to improve access to Cartographic Heritage in the context of the Spatial Data Infrastructures

Luis Manuel Vilches-Blázquez
Ontology Engineering Group. Technical University of Madrid
National Geographic Institute of Spain
Spain
lmvilches@fomento.es
Alberto Fernández-Wyttenbach[1], Mabel Álvarez[2], Miguel Ángel Bernabé-Poveda[3]
Laboratory of Geographical Information Technologies. Technical University of Madrid
Spain
[1] a.fernandez@topografia.upm.es, [2] mablop@speedy.com.ar, [3] ma.bernabe@upm.es

Abstract

The access of historians and document management specialists to documentary funds and old cartography is at times arduous due to the scattering of maps throughout different map libraries. Thus, it would be of interest to be able to access all available information remotely, following the Spatial Data Infrastructures (SDI) guidelines.

DIGMAP is a cooperative project between six European countries that proposes to develop a solution for indexing, searching and browsing, through a thesaurus, in the European collections of digitized historical maps.

It will be possible to match them with the geographic areas covered by each map in the collection and to find it on them, based on standard and open data models. These results will be useful for local digital map libraries, especially if it is of historical – cultural importance, or as interoperable components for wider and distributed systems.

This project takes advantage of thesauri technologies to develop three major subsystems: On one hand, the gazetteer for managing and providing access to multilingual information relating to geographic features like geographic coordinates, names of places or areas and historical events related with the geographic points, places or areas (with dates or time intervals); The authority file to maintain and provide easy access to the author's information and identification and disambiguating from similar and duplicate authors; and on the other hand, the map services for providing access to map information in raster and vector formats.

The paper presents an early overview on the DIGMAP project, particularly focusing on the multilingual Thesauri aspects with the aim of providing solutions for the cartographic heritage framework, applying the main Spatial Data Infrastructures guidelines. ISO 19112 (Spatial Reference by Geographic Identifiers) will also be considered and its geographic identifiers issues.

1. Introduction

Europe is a continent with a very old and heterogeneous political history. Its large diversity of cultures has been a reason for the definition of multiple borders, but the closed space has been also promoting the inevitable cultural exchange. As a result, the European legacy in historical maps is enormous, very rich and diversified. This cultural and scientific genre of artifacts represents therefore a special perspective of our history.

Assuming that historical maps have important details that make them very difficult to describe, classify and index in "hand-made" descriptive metadata structures, them we must conclude that the actual related discovering services might be not giving us the best of the worlds (Borbinha 2007)

So far, several technological initiatives and scientific forums are aiming at the widespread distribution of historical cartographic funds in the Web through geographic localization tools. The objective is to supply historians and documentary experts with the arduous task of access to the documentary contents stored in different map libraries. It would seem useful to have a tool allowing remote access to a Geographic Information (GI) repository facilitating the access to information and the comparison of some documents located in different servers. All available information would be remotely accessible through a Digital Map Library (DML) abiding by the Spatial Data Infrastructure (SDI) guidelines.

In this sense, some cooperative projects have been set off by a group of map libraries to simultaneously and remotely access several historical cartographic funds. DIGMAP proposes to develop a solution for indexing, searching and browsing in the European collections of digitized historical maps with the support of multilingual geographic gazetteers. It stands for "Discovering our Past World with Digitized Historical Maps", but it could stand also for "digging on maps"! (DIGMAP 2007). It is a project co-funded by the European Community program eContentplus[1], which case overall aim of this program is to make digital content in Europe more accessible, usable and exploitable, facilitating the creation and diffusion of information, in areas of public interest, at Community level (Watelet 2007).

On the other hand, DIGMAP echoes necessity to access, retrieve and sharing geographical information in our current society. For that, this project will develop different tools which allow remote access to a historical cartographic repository. It will facilitate access to information and the comparison of some documents located in different servers, providing access to funds through the creation of cataloguing and visualization Web tools. It follows Spatial Data Infrastructures (SDIs) guidelines.

The paper is organized as follows. In section 2, the current situation of digital map libraries is described. In section 3, some of the main characteristics and components of DIGMAP architecture are explained. The Thesauri and its three major subsystems (Authority file, Gazetteer and Map services) are commented in section 4. In section 5, some prototype features are described. Finally, in section 5, some conclusions are indicated.

2. Digital Map Libraries (DMLs)

There are useful and basic similarities between the global reality of a Spatial Data Infrastructure (SDI) in its well known traditional concept and a Digital Map Library (DML); according with standards (e.g. geographic metadata: ISO 19115 (ISO 2003), Dublin Core (Dublin 2004), agreements (e.g. local, regional, national or international level; INSPIRE Directive (INSPIRE 2007)) and services (e.g. Open GeoSpatial Consortium[2]: Web Map Service[3], Web Feature Service[4], Web Coverage Service[5], etc).

However, there is a number of technological and policy considerations to be taken into account

apart from the characteristics common to all SDIs in any thematic field. Cartographic heritage contained in the DMLs stands out as an exceptional case within the generic frame of an SDI.

At first glance, INSPIRE Directive (INSPIRE 2007) is a nice example of high level community policies among environmental and GI authorities to follow and carry out cooperation agreements between libraries and archives. However, the cartographic heritage is not to be included within the thematic information layers of the European INSPIRE initiative. The European legacy of historical maps is mainly in the hands of many national libraries, archives, and related public institutions. Thus, the policies operating in the environmental– GI scope among the various European and world agencies are clearly different from the ones operating in the cultural– heritage scope. In addition, looking into the different hierarchical levels of the traditional SDI umbrella, the impossibility of harmonization with the cartographic heritage appears evident. (Fernández-Wyttenbach 2007)
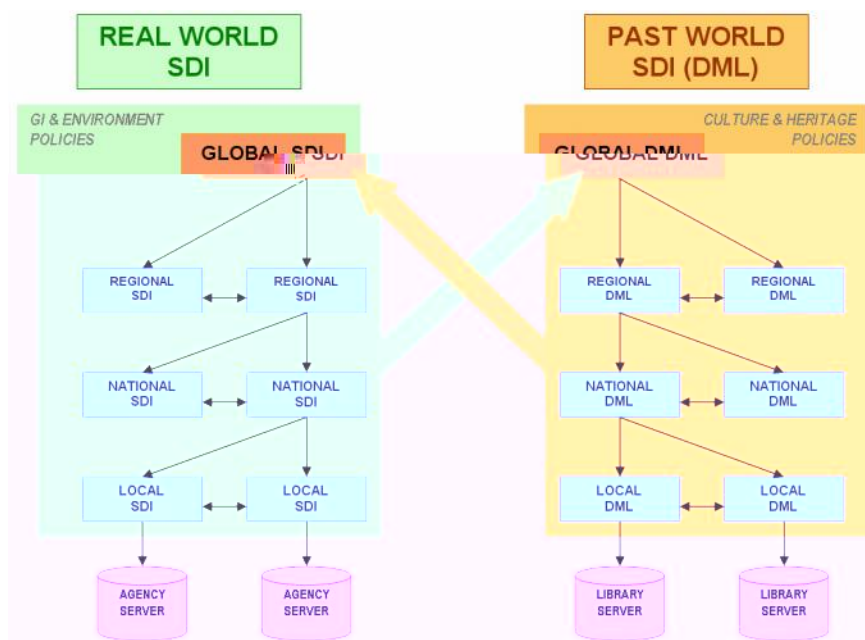

Figure 1. SDI & DML policies and data sharing schema

In the case of descriptive metadata, maps catalogued in libraries have been usually described according to generic bibliographic metadata schemas such as the MARC standards[6] (UNIMARC, MARC21, IBERMARC, etc.) which are in continuous debate in the librarian's community. Therefore it is necessary to define the appropriate gateways for these descriptive profiles to be understandable among themselves, which is a rather complex undertaking.

At the same time some special peculiarities of historical maps should be considered during metadata creation. Graphic scales interpretation (the analytic value it is not always given), archaic measuring units and linguistic peculiarities should be preserved to avoid deleting some data information (Fernández-Wyttenbach 2006).

The risk of deformation of documents of historical interest against the advantages of their publication together with other data in the Web should be also taken into account during the

georeference process, dealing with the image distortion that could not be funny for some scientific users.

3. The DIGMAP framework

In DIGMAP, software solutions produced are based on open standards and released as open-source. The results of the project will be useful for local digital libraries of maps, as either a standalone system or as interoperable components for wider and distributed systems. The overview architecture is shown in Figure 2.
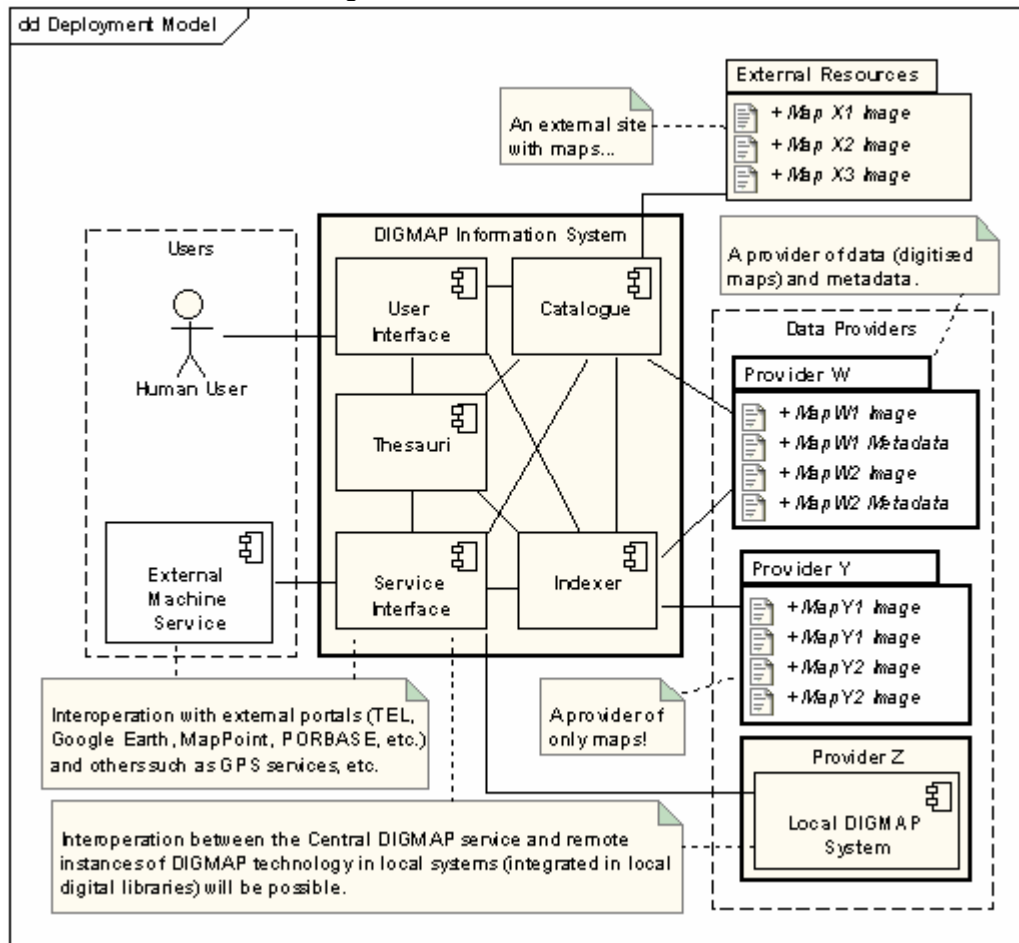

Figure 2. An overview of DIGMAP Architecture

The central component of the DIGMAP Information System will handle service configuration and interconnection of the different modules. Moreover, this central component will also be responsible for storing User data. Next, we describe briefly different components of this Information System (Martins 2007)

The Catalogue is the service responsible for the submission, editing and storage of the records describing the resources. The catalogue must support the definition of collections, i.e. groupings of the resources according to some criteria. It will be possible to register resources in the catalogue through a local user interface, or importing records (e.g. through Z39.50) and SRW/SRU, although the preferred interface for metadata interoperability will be OAI-PMH). The local user interface of the catalogue must make it possible to edit any existing record. This catalogue must be able to maintain the descriptions of authorities and of the maps in multiple metadata formats, especially in UNIMARC, MARC21 and Dublin Core.

The Indexer is the component that supports to the automatic indexing of resources. It will be composed of three major modules: the Image Analysis module, addressing the automatic enriching of metadata records through text mining techniques; the Retrieval module, addressing the generation of appropriate indexing structures for supporting the required retrieval operations.

The Thesauri is the component of the system that registers auxiliary information for the purpose of indexing, searching and browsing resources. A survey specifically focusing on historical and geographical thesaurus is available at (Vilches-Blázquez 2007). This component will be described in great detail in next section.

The User Interface will offer a browsing environment for humans. It will explore paradigms inspired by Google Maps [7], Virtual Earth [8], TimeMap (Johnson 2004) and other similar systems and previous research efforts. Some of the specific functions that will be provided by the User Interface include timeline visualizations, clustering of related resources and advanced geographical information retrieval mechanisms. This interface will also include a module for handling user questions, which will likely be developed by extending an existing forum system.

The Service Interface will provide access for external services. DIGMAP will explore different solutions for interoperability with other systems and services (e.g. Web portals, Mappings clients and applications such as Google Earth). For this purpose, the applications and services to be developed will be based on open standards (e.g. XML Web Services), sharing features with other applications.

These components will manage multilingual information and for that reason they will have multilingual interfaces for searching and browsing, as also machine interfaces for interoperability with other systems.

To sum up, DIGMAP Information System will make possible to match different resources with the geographic areas covered by each map in the collection and to find it on them, based on standard and open data models. It will produce useful results for local digital map libraries, especially if it is of historical – cultural importance or as interoperable components for wider and distributed systems. Moreover, from a technical viewpoint, this project involves many challenges related to the area of geographical information retrieval (Martins 2007). As such, it will build on previous research efforts such as the Alexandria Digital Library project (Hill 1999), SPIRIT (Jones 2004), GREASE (Martins 2006 and Silva 2006) and others (Reid 2003 and Kornai 2003). On the other hand, this architecture is also consistent with the INSPIRE (INSPIRE 2007) general model for an European SDI, which is direct consequence of the DIGMAP project philosophy follows interoperability way applied to georeferenced historical and/or cultural resources in digital libraries.

4. The Thesauri

This Thesauri approach is really different to traditional approaches of this concept so this one is a proposal out of ISO standards (ISO 1985 and ISO 1986) or different historical and geographical thesaurus collect at (Vilches-Blázquez 2007). This thesaurus is the subsystem of the DIGMAP system that registers auxiliary information for the purpose of indexing and

searching of resources and browsing for resources (Manguinhas 2007). In DIGMAP two distinct space are exited for this purpose; the first one manages geographic feature, like geographic coordinates, names of places, names of areas and historical events related with the geographic points, places or areas (with dates or times intervals). The other one manages person names or organizations (authors, publisher, editor, etc.)

In order to support the processes of managing these two spaces of information, three major subsystems were designed (see figure 3): The Authority file to maintain and provide easy access to descriptive information of an authority entity; the Gazetteer for managing and providing access to information related to current and historical geographical features and time periods; and the Map Services for providing access to map information in raster and vector formats.
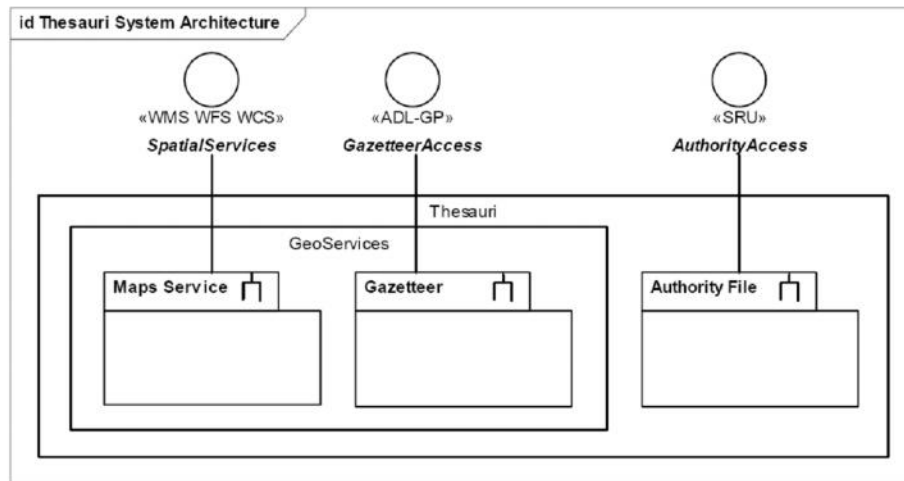
Figure 3. The system architecture for Thesauri system

Authority file. Authority control is a term in library and information science to refer to the practice of creating and maintaining headings for bibliographic material in a catalog. It enables cataloguers to disambiguate items with similar or identical headings and second function. Besides, authority control is used by cataloguers to collocate materials that logically belong together, although they present themselves differently.

The most common way of enforcing authority control in a bibliographic catalog is to set up a separate index of authority records, which relates to and governs the headings used in the main catalog. This separate index is often referred to as an authority file. It contains an indexable record of all the authority information in a given library, which cataloguers consult when making, or revising, decisions about headings.

The DIGMAP authority file will be restrained to control over author information with the purpose of providing a consolidated view over the authors. For indexing, the authority file provides the means for aggregating resources from the same author together in the same index which would be otherwise dispersed in several different indexes. The authority file is also used within the User Interface, to provide an accurate index of author's resources for the purpose of searching and browsing.

The data model used for the authority file is based on the Metadata Authority Description Schema (MADS) specification[9]. It is a XML Schema[10] that may be used to provide metadata

about agents (people, organizations) events and terms (topics, genres, etc.). MADS has a relationship to the MARC 21 Authority Format, carrying selected data from MARC 21.
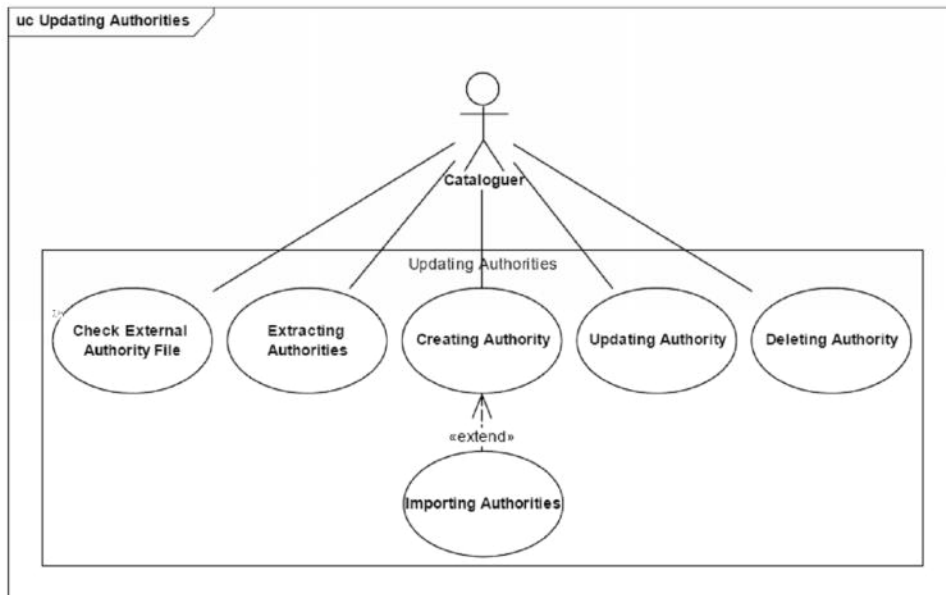

Figure 4. Detailed view for the Updating Authorities use case

Gazetteer. Per (ISO 19112, 2003) is a directory of instances of a class or classes of features containing some information regarding position. On the other hand, a Gazetteer Service (Fitzke and Atkinson 2006b) is defined as a network-accessible service that retrieves one or more features, given a query (filter) request. This filter request must support selection by well-known feature attribute values, and especially by published or context-unique identifiers. The queryable feature attributes are any properties that describe features, including but not limited to feature type, feature name, authority, or identification code.

In DIGMAP project, the gazetteer is used for indexing purposes, providing the means for converting current and historic place names into locations on the Earth's surface or vice versa. The gazetteer is also used within the User Interface, in order to provide geographical searching and browsing. By translating between geographic names and locations, resources can be better discovered trough matching the footprint of a geographic name to the footprints of the resources.

Open GeoSpatial Consortium (OGC) proposes as gazetteer a refactored ISO 19112 (ISO 19112, 2003) content model published through a Web Feature Service (Fitzke 2006a), that is, Web Feature Service Gazetteer (WFS-G). However, no major developments have been made since the original proposal and there are still no available implementations of this standard. By way of proof some reference SDI such as United States geodata.gov[11] offer a non WFS gazetteer service[12]. Other SDIs publish their own gazetteer content model through WFS (e.g. the Canadian Geospatial Data Infrastructure (CGDI)[13] (GeoConnections 2004) and the Spanish

[11] Itto://www.w3c.org/XML/Schema
[12] Itto://gos2.geodata.gov/wps/portal
[13] Itto://geodiscover.cgdi.ca/web/guest/home
[14] Itto://geodiscover.cgdi.ca/gdp/search?action=entryForm&typeForm=service&language=en&resultsType=Web Services

Spatial Data Infrastructure[14] (IDE-E)). Even in those cases the content model is far from ISO 19112. For example, the Spanish SDI Working Group is developing the Spanish Gazetteer Profile (Modelo Español de Nomenclátor, MEN) based on the Alexandria Digital Library Gazetteer (Alexandria 2004) (ADL) content model (Rodríguez-Pascual 2005) that would become the recommended content model in Spain for both gazetteer services and sharing standard gazetteer data between administrations. (López-Pellicer 2007). For these reason, the overall design process is aligned with the approaches taken by the Alexandria Digital Library (ADL) Gazetteer service, which is perhaps the most important research effort addressing the modelling, prototyping and evaluation of gazetteer applications for georeferenced digital library architectures.
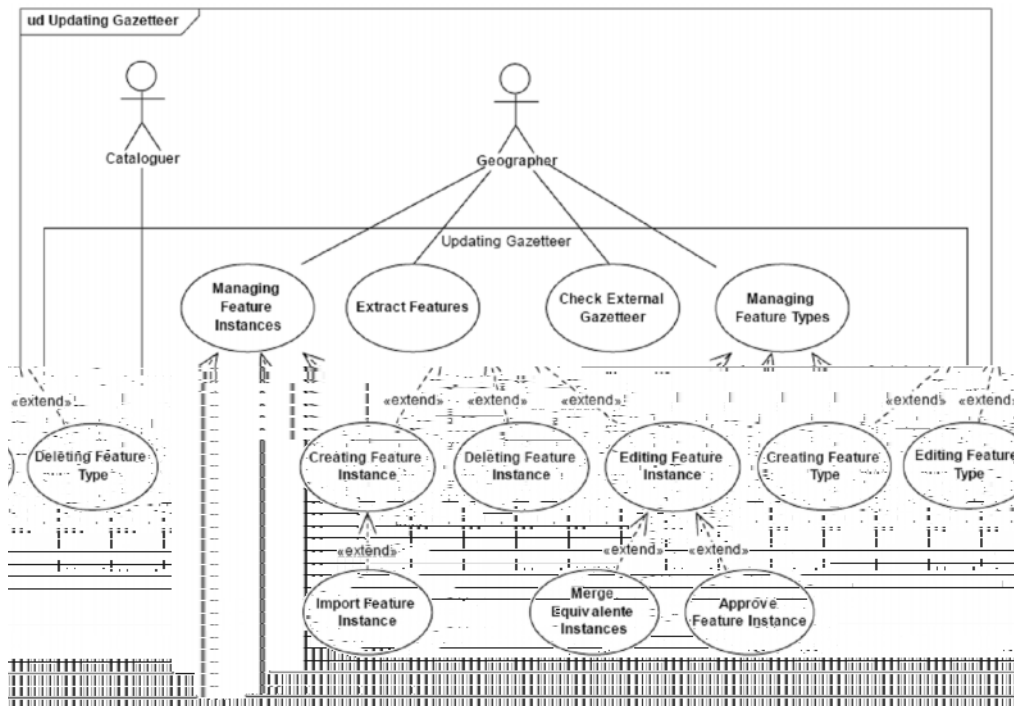
Figure 5. Detailed view for the Updating Gazetter use case

In order to build the DIGMAP Gazetteer, we propose to reuse the software and the data model (ADL Gazetteer Content Standard – GCS)[15] from the ADL Gazetteer. The ADL Gazetteer server uses a relational database system as the back-end implementation for the GCS. The ADL Gazetteer Protocol is essentially a web services interface allowing users to query the back-end with basis on place names, spatial locations, feature types and identifiers. The Web-based client interface interacts with the web service, providing the support for browsing the gazetteer contents.

Concerning the gazetteer data, DIGMAP studies the integration of place name information coming form multiple sources. However, as a starting point, this project focuses on the use of GeoNames[16] dataset. It contains over eight million geographical names and consists of 6.5 million unique geographical features. Another source of gazetteer related information is the one contained within bibliographic catalogues and authority files.

An important limitation of the ADL gazetteer server is related to the fact that it has no transactional capabilities (adding and editing gazetteer entries). DIGMAP gazetteer defines ways for extending the ADL Gazetteer Protocol in order to consider requests for creating, deleting and updating features. Also, the gazetteer service considers authorization, as the transactional capabilities will be restricted to specific classes of DIGMAP users. The addition of transactional capabilities is taken as inspiration the Transactional Web Feature Service (WFS-T) specification from the OGC.

Map Services. The idea is compare map resources with reference map data, in order to "match" the resources to specific regions of the globe. Building the DIGMAP Map Services, we propose to use the open-source software form the GeoServer project[17]. It supports the WFS-T, Web Coverage Service (WCS) and Web Map Service (WMS) open protocols from the OGC, in order to provide map data in JPEG, PNG, SVG, KML, GML and many other formats.

By building on OGC standards to manage the access to geographical information, the DIGMAP architecture provides a high degree of flexibility, as well as many possibilities for deployment in a distributed setting. Moreover, as we said previously, the architecture is also consistent with the INSPIRE (INSPIRE 2007), which is direct consequence of project philosophy follows interoperability way applied to georeferenced historical and/or cultural resources in digital libraries.

5. The Prototype: cat.on.map

This prototype makes possible to register and/or import references to maps, as also structured description of the geographic information of the maps (geographic boundaries, scales, details, etc.). It is available on DIGMAP prototype website[18].
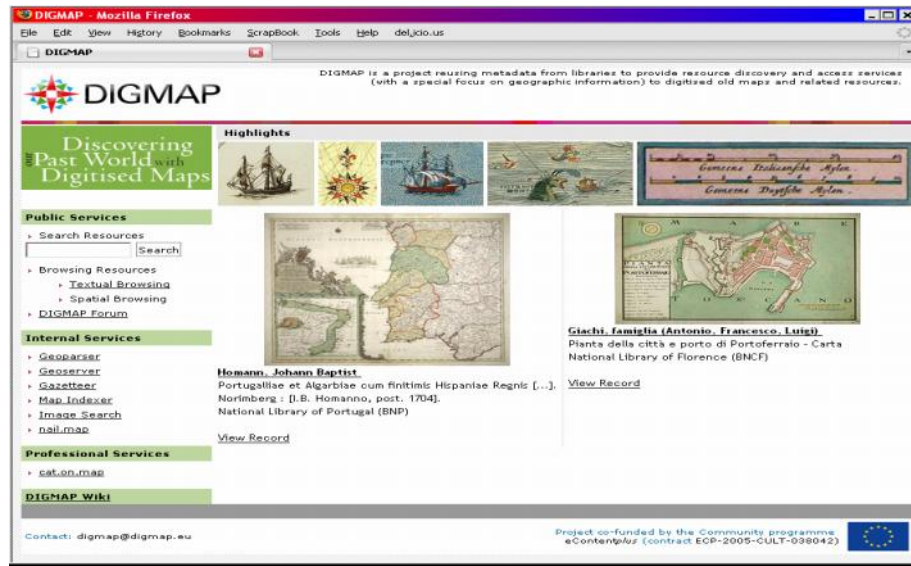
Figure 6. A prototype of the portal

The main tasks for this prototype development were; one first period was the development of the Catalogue Requirements and second one was the development of the Catalogue Design and implementation. The cat.on.map includes the following features:

Authority File: In this task the requirements and design of an information system for the

[17] Ltto://dkq urx qlk$

[18] Ltto://HeOler'#NMVWWRV-%%$OMIESLQQIREKI$
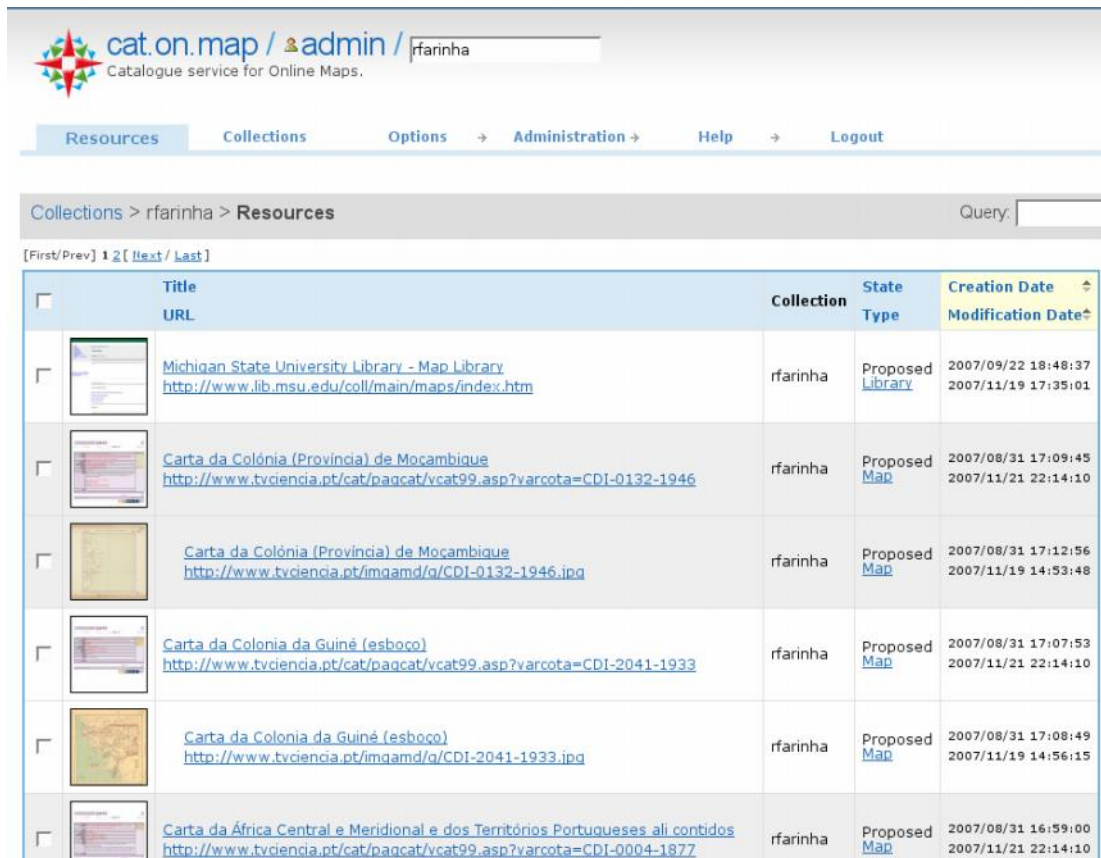
management of the Authority File subsystem were analyzed.

Map Harvester: This task focused in the development of   a subsystem to register sites in the Web with digitized maps. A prototype of this service was developed. The first version of the prototype was called cat.on. map.

Metadata cross mappings: This task developed functional cross mappings between the schemas to use in the project.

Catalogue: This task is in progress and produced a first version of the prototype.

Next, different examples of the prototype cat.on.map are shown in the following figures:

Figure 7. Resources

Figure 7 is an example of resources, indicating for each one: Title, URL, collection, state, type and creation and modification date.

Figure 8. Catalogue – cat.on.map / cataloguing and external resource.

Previous figure provides an example of resources information. For each resource, following data are included: URL, title description, date, geographic box, spatial coverage, scale, creator, publisher, part of resource, type, language and collection. This prototype is translated in several languages: English, French, German, Spanish, Portuguese, among others.

Finally, this prototype is only an example DIGMAP work. It is working towards the goal of becoming the main international information source and reference service for old maps and related bibliography. The targeted users of the services are citizens, students, scholars, teachers, professionals in libraries, archives and technology, as well as institutions intending to provide similar services, such as other national, university and research libraries, archives and so on (Fernández- Wyttenbach 2007).

6. Conclusion

- New agreements and cooperative projects of Digital Map Library services are to be build representing a new concept within the Spatial Data Infrastructure as applied to the cartographic heritage.

- DIGMAP comes up as a case in point in the creation of a European Digital Map Library in parallel to INSPIRE, the European Spatial Data Infrastructure. It should be a case in point and an example to be disseminated throughout the world.

- The DIGMAP Project offers a new perspective in thesauri design and implementation within a geographical and historical context. Thus, it success taking advantage of thesauri technologies,

developing three particular subsystems to its framework: Authority File, Gazetteer and Map Services.

- The preliminary version of the DIGMAP Prototype has been presented in this paper. It has demonstrated that it can deal properly with the bibliographic records from multiple sources and it has validated the generic use cases. Spatial browsing will be provided in the second year.

7. References

Fritzke, J., Atkinson, R. (2006a). Gazetteer service profile of the web feature service implementation. OCG Draft Implementation Specification 05-035rl.

Watelet, Marcel (2007) eContentplus Work Programme. 1st DIGMAP Workshop, Lisboa.

Boyle, James (1996) "Sold Out," *New York Times*, OP-ED, March 31, p. E15.

Mechling, Jerry (1994). "The Hidden Data Pricing War." *Governing*, October, p. 82.

Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE). http://www.ec-gis.org/inspire/directive/l_10820070425en00010014.pdf

DIGMAP - Discovering our Past World with Digitised Maps (2007). Programme eContentplus – Project: ECP-2005-CULT-038042. http://www.digmap.eu/

Martins, B., Silva, M., Freitas, S. and Alfonso, A., (2006), "Handling Locations in Search Engine Queries", Workshop on Geographic Information Retrieval at SIGIR, August 2006. http://www.digmap.eu/doku.php?id=wiki:papers

Fernández-Wyttenbach, A., Ballari, D. and Manso, M. A. (2006), "Digital Map Library of the Canary Islands". e-Perimetron: International Web Journal on Sciences and Technologies affined to History of Cartography. Vol.1, No. 4. Autumn 2006. Pp.262-273. ISSN 1790-3769. http://www.e-perimetron.org/Vol_1_4/Wyttenbach.pdf

Fernández-Wyttenbach, A., Álvarez, M., Bernabé, M. and Borbinha, J., 2007, "Digital Map Libraries Services in the Spatial Data Infrastructure (SDI) Framework: The Digmap Project". 23rd. International Conference in Cartography (ICC2007). International Cartographic Association (ICA). Moscow, (Russia). http://www.digmap.eu/doku.php?id=wiki:papers

Borbinha, J., Pedrosa, G., Gil, J., Martins, B., Freire, N., Dobreva, M. and Fernández-Wyttenbach, A., 2007, "Digital Libraries and Digitised Maps: An Early Overview of the DIGMAP Project". 10th. International Conference on Asian Digital Libraries (ICADL). Hanoi, (Vietnam). Springer 2007. ISBN: 978-3-540-77093-0.

Manguinhas, H., Pedrosa, G., Martins, B., Borbinha, J., Bracke, W., Vilches-Blázquez, L. and Álvarez, M., 2007, "Thesauri Design and Implementation", September 2007. DIGMAP Project Deliverable (D.2.2). eContentplus Programme - European Commission (ECP-2005-CULT-038042)

International Standard Organization (ISO) (2003) ISO 19115 Geographic Information - Metadata

Dublin Core Metadata Initiative (2004) DCMI Metadata Terms. Available on website: http://dublincore.org/documents/dcmi-terms/)

Martins, B., Borbinha, J., Pedrosa, G., Gil, J., Freire, N (2007). Geographically-Aware Information Retrieval for Collections of Digitized Historical Maps. CIKM-07 conference – 2007 Workshop on Geographical Information Retrieval. Lisbon, (Portugal).

Vilches-Blázquez, L.M., Martins, B., Wyttenbach, A., Bernabé-Poveda, M.A., Álvarez, M., Luzio, J. and Borbinha. J. (2007) Geographical and historical thesauri: The state of the art. DIGMAP Project Deliverable (D.2-3(b)1(i)4(n)1(h)1(a)1(.)]TJETQq0 -2 612 792 reW n100 TzBT/F(L)1Einha.

Alexandria Digital Library Project (2004) Guide to the ADL Gazetteer Content Standard. University of California, Santa Barbara. 3.2 edn.

Rodríguez-Pascual, A., López-Romero, E., Abad-Power, P., Sánchez-Maganto, A. (2005) Modelo de Nomenclátor de España. Consejo Superior Geográfico, Ministerio de Fomento de España

F.J.López-Pellicer, F.J. Zarazaga-Soria, A. Mogollón-Díaz, J. Nogueras-Iso y P.R. Muro-Medrano (2007), The gazetteer content model issue:Could Spatial Data Infrastructures provide it?, Lecture Notes in Geoinformation and Cartography (LNG&C), vol. XVII, p.187-200. ISSN 1863-2246.