Sharing geodata through university libraries: the case of Politecnico di Milano

Ilaria Muratori[1], Marcella Samakovlija[1]

[1] Politecnico di Milano. TeDOC - Area servizi bibliotecari di Ateneo (University Library Services)

Abstract

University libraries often have huge collections of cartographic documentation, both in traditional paper format and in complex digital format.

It's not only a matter of images or vectors, but also a problem of dealing with complete databases with thematic attributes and details.

In Italy, the libraries' systems of classification and research of these resources are still far to be well developed.

TeDOC, that has a large collection of historical and actual maps, both on paper and on digital format, and is responsible of the distribution of these resources to professors and students within teaching and research institutional activities of Politecnico di Milano, is developing a new approach to make these resources more accessible and easier available through the implementation of a geoportal.

The goal is to build a geolibrary, a library that puts at users disposal geoinformation and documents for which the primary search mechanism is place.

The issues faced within this project are related to metadata creation for different geospatial information, technical policy to assure protected access to resources with particular use constraints (e.g. resources available only for specific categories of users such as students), the opportunity to share resources within federated catalog both Italian and foreigners, and lastly to develop monitoring techniques to get statistical data about users.

Keywords

Metadata standard, sharing geodata, university geolibraries, geoportal

1 Which was the problem?

At TeDOC we have a large collection of digital and paper maps. These documents were searchable through the library catalogue and our web site[1].

In the recent years we became more and more aware that we needed to change the approach to our cartography conservation, organization, storage and accessibility.

The tools available to manage library documents at Politecnico, such as the discovery tool PRIMO and the catalog software Sebina, are not suitable to deal with cartographic documents. This problem is not due to a particular lack of those tools, but it rather depends on the specific characteristics of cartography, for which the best search method is place.

The search through the library catalogue has always been very difficult mostly because each cartographic series consists of numerous map sheets and the

---

[1] www.tedoc.polimi.it

user, without an appropriate geographic reference, is not able to identify which is the sheet that represents the portion of land of his interest.

On the other hand our old web site was organised into sections provided with map tiles of each cartographic series:

- For paper maps it worked with interactive tiles that brought to the catalogue record of the map containing the information with the borrowing details;
- For digital maps it provided static tiles containing the map coordinates and unique codes to make the request of the files.

Figure 1 - TeDOC old web site: searching paper cartography through interactive tiles linked to library records.

Because map tiles are always different for each map series, the cartographic documents were divided into sections based on the owner/publisher, that in Italy usually corresponds to local administrations or regional agency, and in sub-sections based on the scale of the maps.

As you can imagine this system was quite difficult and put the user at the risk not to find what he was looking for, just because resources were too difficult to be found.

2 The technical choice - Why OGP?

After a deep analysis through several software and spatial search engines, our choice fell on the software open source OpenGeoportal[2]. The Open Geoportal (OGP) is a collaboratively developed, open source, federated web application to

---

[2] http://opengeoportal.org/

rapidly discover, preview, and retrieve geospatial data from multiple organizations. The project is lead by Tufts University along with Harvard and MIT. Several other partner organizations are assisting with the development[3].

Just because it has been developed by GIS librarians, this software is perfectly able to meet the need of university libraries in cartography management, catalogue issues, authentication and login needs and to dialog with the different tools in use in the libraries.
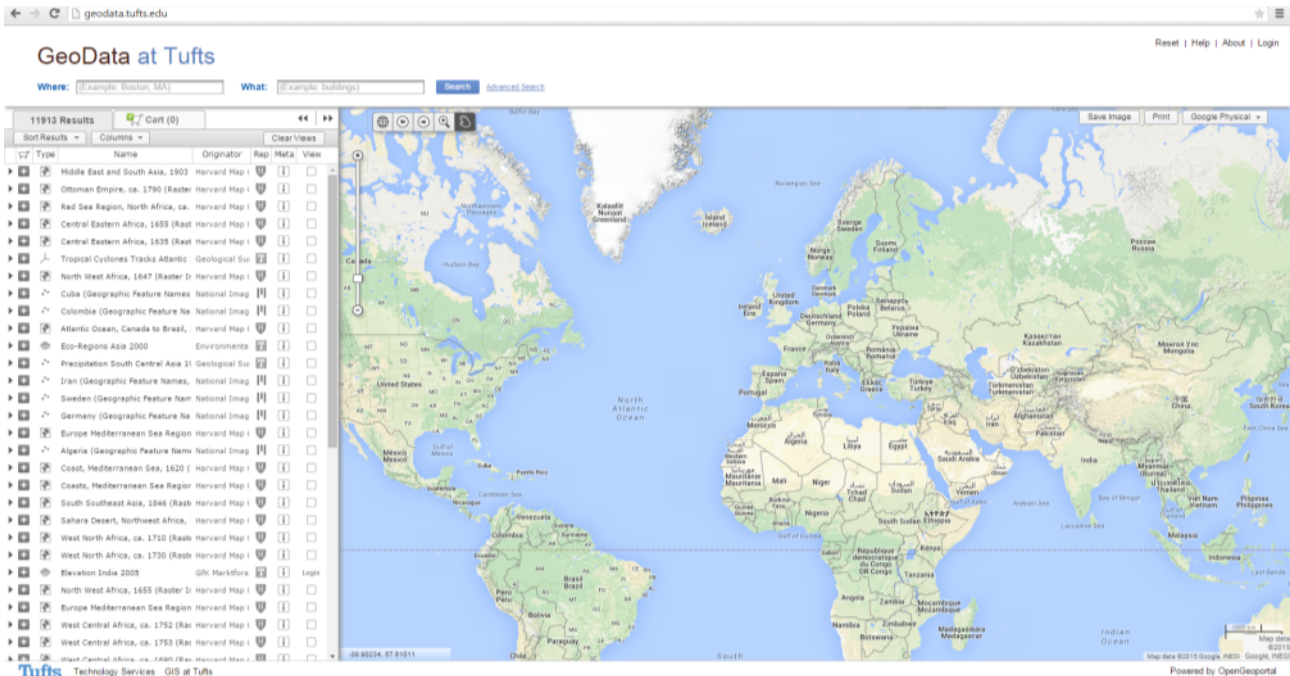

Figure 2 - GeoData@Tufts: Opengeoportal in production - http://geodata.tufts.edu/

3 The project

The project developed by TeDOC is supposed to solve several critical points, both of data search and of data delivery:

1. Difficulty by users in finding the cartographic resources needed;
2. Difficulty in delivery of digital geodata: an automatic download was not possible and the user had to come to our office to get the files;
3. Difficulty in producing statistical evaluations about resources utilization: any use survey had to be made by hand;
4. Need to find a unique method to catalogue and manage all kind of cartographic resources both digital and paper, raster and vector.

Thanks to the implementation of our geoportal based on the OGP web application the user can move and search into the different geographical areas in order to identify what are the maps available, regardless of the owner/publisher organization, or the data format, the scale of representation and the year of the survey. The geoportal provides results of the geographical search as a sort of map stratification or a core sampling of all the cartographic resources available for that land portion (e.g. for the territory of Milan there are municipal base maps at a very large scale, regional and national scale maps, cadastral maps, urban plans, historical maps, etc.). Furthermore it provides the

---

[3] Examples of OpenGeoportal in production: Harvard (http://calvert.hul.harvard.edu:8080/opengeoportal/), University of Wisconsin-Madison (http://maps.sco.wisc.edu/opengeoportal/), University of California, Berkeley (http://gis.lib.berkeley.edu:9081/)

download function both for selected metadata and files.

It is important to note that the implementation of this project involves, in addition to TeDOC, also the ICT Area, with informatics tasks such as the installation and customization of the software, and the Geodesy and Geomatics office of the Department of Civil and Environmental Engineering (DICA) for some technical-methodological support.

3.1 Analysis and re-organization of TeDOC cartographic resources

First of all, since the TeDOC cartographic collections are very different, we started with deep analysis of the digital collections to check the quality of metadata (if present) and, on the other hand, the methods of distribution to the users of each collections.

For example in the case of the Regional Base Map produced by Regione Lombardia, the geodata (in shp format) are organized into thematic groups and distributed to users in a compressed zip format, not in single layers. This is because each theme, such as for example the land use theme, is divided into many layers.

ente	DESCRIZIONE	codice carta	scala	formato	note
Comune di Milano	Carta Tecnica Comunale (2006)	0101	1:1.000	jpg	Vedi quadro d'unione
	Carta Tecnica Comunale (2006)	0102	1:1.000	dwg	Vedi quadro d'unione
	Carta Tecnica Comunale (2004) Altimetrie	0103	1:1.000	tiff	Vedi quadro d'unione
	Carta Tecnica Comunale (2004) Pertinenze	0104	1:1.000	tiff	Vedi quadro d'unione
	Carta Tecnica Comunale (1990)	0105	1:2.000	tiff	Vedi quadro d'unione
Comune di Cesano Maderno	Carta Tecnica Comunale (2005)	1701	1:2.000	dwg	Vedi quadro d'unione
	Carta Tecnica Comunale (2005)	1702	1:5.000	dwg	Intero territorio comunale
Comune di Cremona	Carta Tecnica Comunale (2003)	0201	1:1.000	shp	Intero territorio comunale
	Carta Tecnica Comunale (2003)	0202	1:1.000	dwg	Intero territorio comunale
Comune di Gessate	Carta Tecnica Comunale (2008)	0301	1:1.000	shp	Intero territorio comunale
Comune di Giussano	Carta Tecnica Comunale (1982)	1501	1:5.000	tiff	Intero territorio comunale
	Carta Tecnica Comunale (2012)	1502	1:1.000	dwg	Intero territorio comunale
	Carta Tecnica Comunale (2012)	1503	1:1.000	shp	Intero territorio comunale
Comune di Monza	Carta Tecnica Comunale (1993)	0401	1:2.000	dwg	Vedi quadro d'unione
	Carta Tecnica Comunale (2005)	0402	1:2.000	shp	Intero territorio comunale
	Carta Tecnica Comunale (2005)	0403	1:2.000	dwg	Vedi quadro d'unione
	Ortofoto (2005)	0404	1:2.000	tiff	Vedi quadro d'unione
Comune di Novate Milanese	Carta Tecnica Comunale (2010)	1601	1:1.000	dwg	Intero territorio comunale
	Carta Tecnica Comunale (2008)	1602	1:1.000	shp	Intero territorio comunale
Comune di Pavia	Carta Tecnica Comunale (2001)	0501	1:2.000	dwg	Vedi quadro d'unione
Comune di Rho	Carta Tecnica Comunale (1996)	0601	1:2.000	dwg	Vedi quadro d'unione
Comune di Sesto San Giovanni	Carta Tecnica Comunale (2001)	0701	1:2.000	dwg	Vedi quadro d'unione
	Aerofoto (2001)	0702	1:2.000	jpg	Vedi quadro d'unione
	Carta Tecnica Regionale, volo 1980-91	0801	1:10.000	tiff	Vedi quadro d'unione
	Carta Tecnica Regionale, volo 1994	0802	1:10.000	tiff	Vedi quadro d'unione
	Carta Tecnica Regionale, volo 1994	0803	1:25.000	tiff	Vedi quadro d'unione
	Carta Tecnica Regionale, volo 1994	0804	1:50.000	tiff	Vedi quadro d'unione
	Base Dati Geografica CT50 (a colori), volo 1994	0805	1:50.000	tiff	Vedi quadro d'unione
	Modello digitale del terreno DTM20, volo 1981-83	0806	1:10.000	grid-ascii	Vedi quadro d'unione
	Base Dati Geografica - CT10	0807	1:10.000	shp	Scegliere tematismi
	Basi Informative Ambientali della Pianura	0808	1:25.000	shp	Scegliere tematismi
	Base Dati Geografica di Sintesi	0809	1:250.000	shp	Scegliere tematismi

Figure 3 - Re-organization of TeDOC cartographic resources: extract of cartographic digital collections

The single layer is not significant in an analysis of the territory and, for the user, finding all the layers that belong to a theme would be quite difficult. That's why we decided to maintain the same kind of aggregation even within the geoportal: actually the various collections, or cartographic series, are

organized into general themes (e.g. land use, transportation infrastructure, administrative units, etc. ...) in order to make it easier to select and download the data.

Instead, the raster data, that are usually divided into several tiles, are ingested into the geoportal one by one so that the user can identify exactly which is the area represented by the map that he is selecting and that he may download.

Making our choices we have considered our experience gained through years and years of observation of our users who often have no expertise in GIS.

For what paper maps are concerned, they are registered in the University Library Catalogue and we adopt as metadata the library record of each map.

3.2 Metadata standards. A solution for cataloguing issues

The implementation of metadata is critical, as metadata provide valuable information about the data itself. In fact collecting, creating and organizing the metadata for our resources has been the most time-consuming and labor-intensive process we had to face.

We chose to follow international standard to guarantee time preservation and exchange opportunities.

The largest part of our digital cartographic resources had original metadata with very different levels of quality. For example "historical" datasets are given (by the owner/publisher institution) with incomplete or not precise metadata that need to be re-compiled almost completely. Instead more actual datasets have more comprehensive metadata that are filled according to European standards.

This fact can be explained by the delay, typical of Italy, in conforming to the international regulations, and mainly by the real difficulties encountered by the institutions to go back in time to create metadata for old geodata that were published when still the concern about metadata was not yet a critical issue.

And since our idea was to create a geoportal that could be shared with other universities, including foreigners, and at the same time to build a web application that can keep together (read, search) resources held even by other universities, it has been necessary to make a comparison of the different standards in use. We of course decided to provide bilingual metadata for all our resources.

It was therefore carried out an analysis of the metadata standards in use, of the information that they must include and the specific ways in which this information should be expressed. The starting point for this analysis was the INSPIRE Directive, not only for its prescriptive nature, but mainly because it introduces the idea of building a transnational structure for geospatial information and thus the need to catalog this information according to specific criteria that allow to go back over the path of the data history and that make the search process the easiest possible.

As known, the standard in force for Europe is ISO19139, while in the US it is in use the standard FGDC (but debates are ongoing on whether to move to ISO19139 which allows a more detailed description of the geographic data).

To follow the standard, we listed the following elements to be reviewed:
- keywords
- abstract
- lineage
- purposes
- data format

- use constraints

Our job was carried out following the OpenGeoPortal Metadata Working Group, a multi-institutional partnership of data and metadata experts, who are seeking to develop a set of common practices for creating and exchanging geospatial metadata.

Included in these best practices there are recommendations for the use of keyword and thesaurus, outlining encoding standards for the construction of free text fields, and methodologies about describing and relating metadata of collections to layers' metadata.

This is very important since our future target is to federate our geoportal with the others existing university geoportals to share geodata.

In a first step we decided to create the metadata template for some cartographic series. In particular, we intervened on the maps of Milano Municipality and Regione Lombardia, which are the institutions with whom we have the more frequent and consolidated relationships. For a better result in the template creation we consulted the cartographic departments of the two institutions, either to ensure the correctness of the information or to give a feedback information to the institutions themselves that may use our support in their processes of metadata review and creation.

Our target for this step was to define a complete procedure to be applied for the creation of the metadata for the wholeness of our cartographic resources.

The metadata template is filled with the basic info that are the same for all the elements of a specific cartographic collection.

The fields that must filled up creating the metadata template are:

1. contact information: Responsible Party (owner, publisher, resource provider, point of contact...)
2. series name
3. collective title
4. abstract
5. keywords and thesaurus name
6. lineage
7. use limitations
8. other constraints
9. distribution format

The field "title" has a part in common with all the maps of a specific cartographic series, but it usually contains a specification related to the single map tile or theme.

To create the metadata template we used ArcCatalog 10.3, then editing it with notepad++ if necessary. Afterwards, we passed through the metadata validation using the translator ARCGIS2ISO19139.xml

3.3 Procedures for Automatic Metadata creation

After having developed the metadata template we proceeded with metadata import for each element of each cartographic series.

First of all, it is necessary to project the element in its own coordinate system. A pyton script, designed at this purpose, run the projection command for all files of a particular series.

Afterwards, using another pyton script, we run the import metadata command for every file. At this point it is necessary to make some manual, or semi-automatic, changes to modify the fields "title" and "file identifier". We underline the importance of choosing a file identifier that can assure the

unique identification of the resource now and in the future.

At the end of this procedure we perform the validation (in batch) of all the metadata in the standard ISO 19139.

Finally the validated metadata can be ingested in OGP with the OGPingest tool that perform another check of the conformity of metadata to the standard.

3.4 Paper maps and geoportal: which solutions?

The TeDOC, as already mentioned, has a large collection of paper maps that have been cataloged according to the standard UNIMARC.

Our project wants to connect the catalog with the TeDOC geoportal to allow the search of paper maps together with the digital ones.

To do this, it is necessary to update all the library records by entering geographic coordinates in degrees and the Prime Meridian.

Our work is not yet complete. For the moment we started working on historical maps produced by IGM (Istituto Geografico Militare). We studied the procedures to derive automatically the coordinates of each map sheets from the tilesheet in shp format. Subsequently, we created a DB containing, for each library record, the coordinates of the bounding box and the permalink to the resource.

In the near future it should be possible, from that DB, to perform an automatic update of the catalog by entering the coordinates of the bounding box of each element in the library record. Afterward we will perform an extraction of library records in MARC format (compatible with our system) and, thanks to customization of the OGP made by MIT staff, it should be possible for us to ingest our MARC records into the SOLr OGP index.

4 Conclusion

The TeDOC geoportal, currently in test version, does not yet provide preview and download functions. We still have a lot of geodata to ingest and we are waiting to be able to ingest also all the paper maps. Anyway, since January 2015 the geoportal can be used by Politecnico students to make their search among the cartographic collections already ingested.
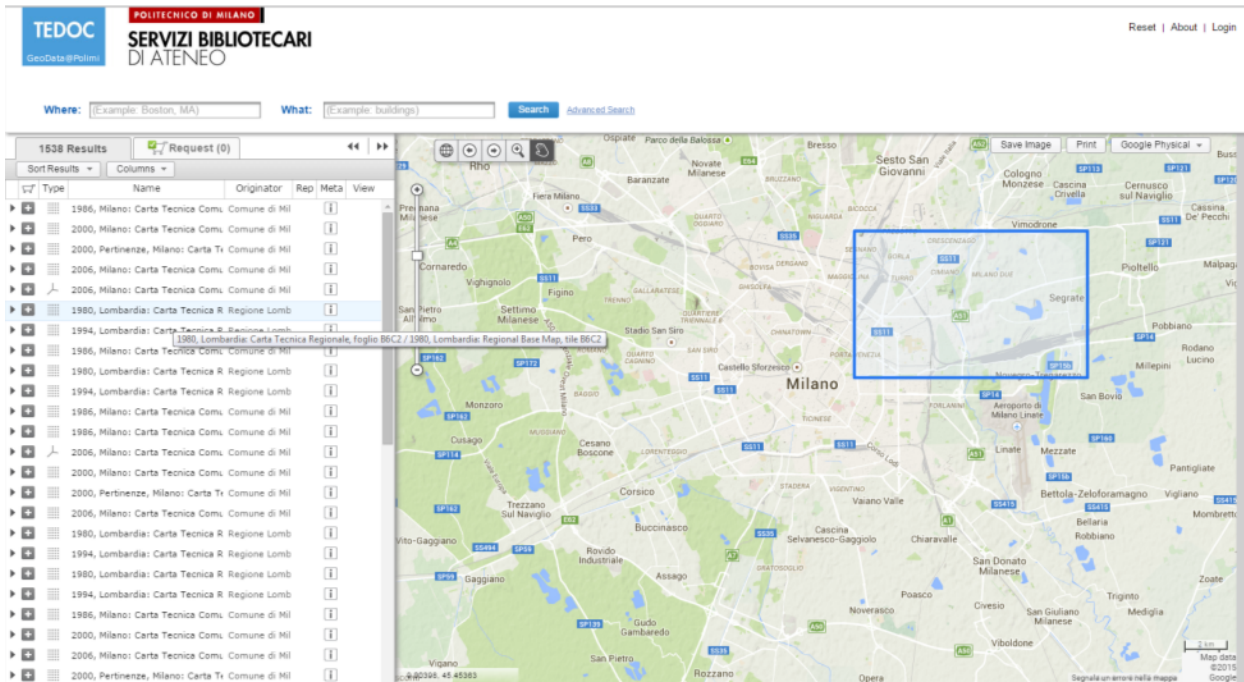
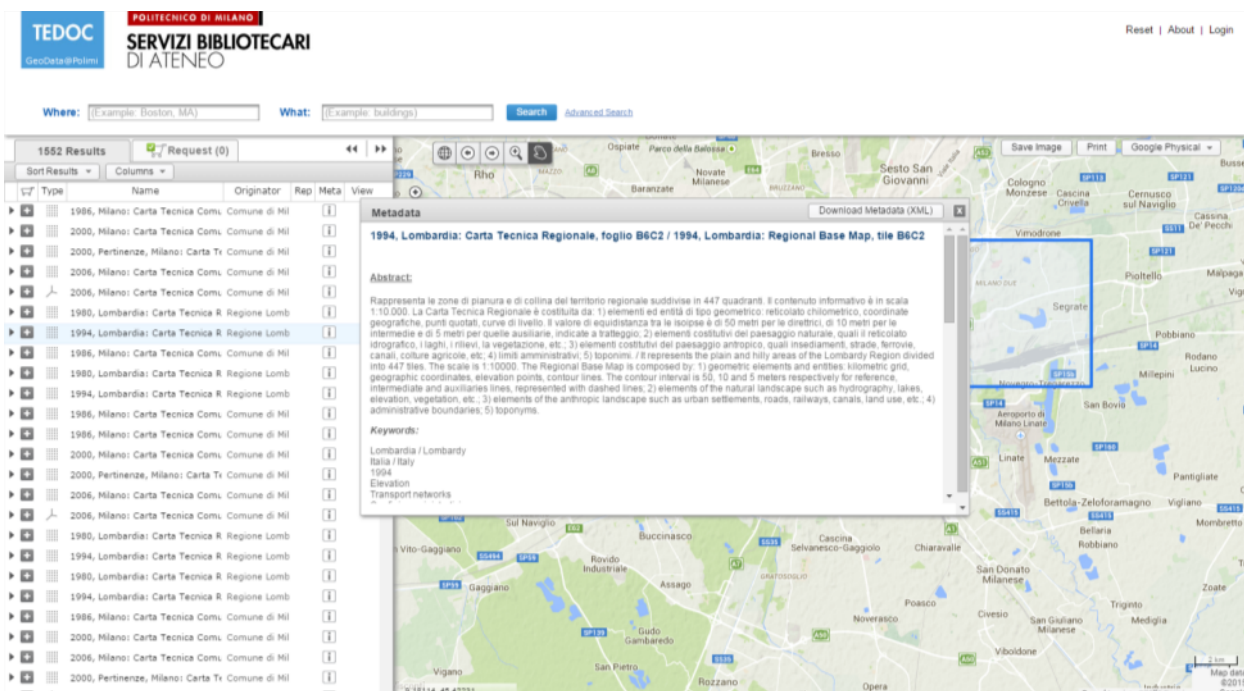*Figure 4 - Searching cartographic resources in GeoData@polimi*

Figure 5 - Metadata in GeoData@polimi

In the next steps, the project will provide the opportunity to sharing the cartographic resources with the other university that are currently in the OGP group with different level of accessibility depending on the use constraints of the documents.

We finally want to note that during the last decade, while users demand of geodata was increasing, also the web availability of these information has greatly increased, that's why the use of a federated "geo-catalogue" will be a great opportunity for the diffusion and sharing information both directly owned by Politecnico and by other University in the world.

Politecnico users will be able to find everything about a geographic area just with a click of the mouse.

We hope that in the short term other Italian and European universities, in addition to those already operating overseas, will join this project to allow an ever wider sharing of knowledge and geospatial information.

References

✔Boxall, J. (2003). Geolibraries: geographers, librarians and spatial collaboration. The Canadian Geographer/Le Géographe canadien, 47(1), 18-27.

✔Donnelly, F. P. (2010). Evaluating open source GIS for libraries. Library Hi Tech, 28(1), 131-151.

✔Durante, K., & Wang, Z. (2012). Creating an actionable assessment framework for discovery services in academic libraries. College & Undergraduate Libraries, 19(2-4), 215-228.

✔Florance, P. (2006). GIS collection development within an academic library. Library trends, 55(2), 222-235.

✔Goodchild, M. F. (2000). Communicating geographic information in a digital age. Annals of the Association of American Geographers, 90(2), 344-355.

✔Hardy, D., & Durante, K. (2014). A Metadata Schema for Geospatial Resource Discovery Use Cases. Code4lib Journal, 25.