Opening up the cartographic heritage of the Spanish Geographical Institute by means of publishing standardized, Inspire compatible metadata

Joan Capdevila Subirana

Servicio Regional en Cataluña,
Instituto Geográfico Nacional, Barcelona, Spain,
joan.capdevila@seap.minhap.es

Alejandra Sánchez Maganto

Centro Nacional de Información Geográfica,
Instituto Geográfico Nacional, Madrid, Spain,
asmaganto@fomento.es

Elena Camacho Arranz

Servicio de Documentación Geográfica y Biblioteca,
Instituto Geográfico Nacional, Madrid, Spain,
melena@fomento.es

Andrés Arístegui Cortijo

Servicio de Documentación Geográfica y Biblioteca,
Instituto Geográfico Nacional, Madrid, Spain,
aaristegui@fomento.es

Abstract

By making the most of the versatility that the Internet provides, and following the latest guidelines developed by the European Union through the Inspire and PSI Directives, the Spanish Geographical Institute has been undertaking in the last few years the release of geographic information in a free and interoperable way. Essential factors for this task are both creating metadata to describe and better use that information, and applying standards, both in the data model and in the web services. In this piece of work we will explain the application of these principles on the historical information held in the Technical Archive of the Spanish Geographical Institute. We will present the recent publication of 120,000 metadata records described as ISO (NEM model), which can be accessed through an Inspire compliant discovery service. This service is interrogated by an open source software catalogue client developed by the Spanish Geographical Institute.

Key Words: Interoperability; metadata; SDI; cartographic heritage; IGN

1. Introduction

A policy to promote the release of all information produced by the Public Administration and making this information accessible for all users was set in motion ten years ago and continues at present both in the European Union in general, and within the Spanish Public Administration in particular. As an example of this policy, we could mention various regulations, agreements and statements that have been approved with the aim of achieving this task, such as the Directive on the re-use of Public Sector Information (PSI)[1] and the Aarhus Convention[2].

Regarding geographic information, this policy has come to fruition in the Inspire Directive[3] across the European Union, and in the Law 14/2010, known as Lisige[4], within Spain. Both regulations recommend the use of Spatial Data Infrastructures (SDI) – which will be defined in the next chapter – as an instrument for publishing all the geographic information held by the Public Administrations.

Within this sphere, the Spanish Geographical Institute (IGN) is firmly upholding the development of SDI; on the one hand, by supporting activities carried out by the Spanish Geographical High Council intending to promote

and divulge the use of SDI in all Spanish domains[5]; on the other hand, by releasing its own information by this means. At present, the majority of the geographic information created by the IGN can be accessed over standard web services and is well described by metadata. In the third chapter of this article we will give an account on the CSW catalogue service and on the client application included in our website, which allows searching and looking up metadata related to geographic information produced by the IGN.

Among the data that we wish to publish, we shall mention all the information gathered since the IGN was founded in 1870. In the last years, the Documentation and Library Service has been making an effort to digitize and catalogue the most distinctive part of the collections held by the Technical Archive of the IGN. In the fourth chapter of this piece of work we will describe the release of metadata referring to 120,000 documents over the CSW catalogue service.

2. SDI, standards and interoperability

The term SDI is used for designating the group of policies, organizational structures and technologies which help locating, accessing and using geographic information over the Internet. In what concerns the Spanish Public Administration, the development of SDI tries to achieve four major targets:

- promoting the sharing of geographic information among the Spanish Administration in an effective way, and thus improving investment efficiency and guaranteeing the use of common reference data.
- contributing to the development of e-Government.
- placing all geographic information at the citizen's disposal.
- encouraging both the commercial field and the academic domain to use this means for their data.

One basic concept on which the SDI construction relies is interoperability, which could be defined as "the capacity to communicate, to process or to transfer information among various data sources, so that the user does not necessarily need a vast knowledge of the particular characteristics of each of

these sources." In practical terms, this means that the exchanged information shall not depend on computer architecture or specific formats, and that applications shall be able to exchange information without human intervention. The client-server architecture is planned in SDI in such a way that the person who releases the data, i.e. the server, exchanges the information on interoperable web services which have very precise functionalities and which anyone, i.e. the client, can use.

Interoperability can be achieved by the use of standards that clearly define the behaviour of web services, data and metadata structures, information embedding methods, etc. The Open Geospatial Consortium (OGC)[6] and other international organizations (ISO, W3C, CEN, etc) have created standards used for geographic information.

Geographic information needs specific standards that consider various aspects such as data georeferencing, information viewer, topological capacities, etc. As regards the interests of metadata, the Spanish Geographical High Council accepted in the year 2004 the "Spanish Metadata Core (NEM)"[7] as a profile of the ISO 19115 *Geographic Information - Metadata* regulation, which includes the minimum number of metadata recommended for describing geographic information. A working group headed by the IGN has defined recently a translation protocol that enables a transformation from MARC21 catalogue data into NEM/ISO 19115 data and vice versa[8]. Metadata held by the IGN can be accessed over a catalogue web service (CSW). This will be described in the next chapter.

Some historical information has already been published on web services. At present, the IGN releases a Web Map Service (WMS) which includes both the first edition of the 1:50,000[9] scale National Topographical Map and a mid-19th Century land registry map comprising all its graphic and text information[10].

3. Web service catalogue and search tools offered by the IGN

The Inspire Directive states that Member States of the European Union shall provide descriptions of spatial datasets and services within the scope of the National SDI and shall establish and operate a network for those spatial data sets and services for which metadata have been created.

To ensure that the spatial data infrastructures of the Member States are compatible and usable in a Community and trans-boundary context, this Directive states that Implementing Rules (IR) about Network Services are adopted and required for different web services. One of them is the "Discovery Service" that allows users and computer programs to search for spatial datasets and services based on their metadata records.

The Discovery service defines a common interface to describe, capture and query data, services and geographical resources with metadata. There is a specification of the Open Geospatial Consortium that establishes what should be a standard and interoperable catalogue service, i.e. a catalogue service web (CSW)[11], which works as shown in figure 1.

The specifications of CSW describe the abstract information model (the language of the query), the general model (the sets of service interfaces) and the protocols (Z 39.50, HTTP...). The actual version of this specification is version 2.0.2.

Taking into account the principles of the Inspire Directive about services, and due to the necessity to ensure that Inspire Discovery Services are

Fig. 1: Use of catalogue service web.
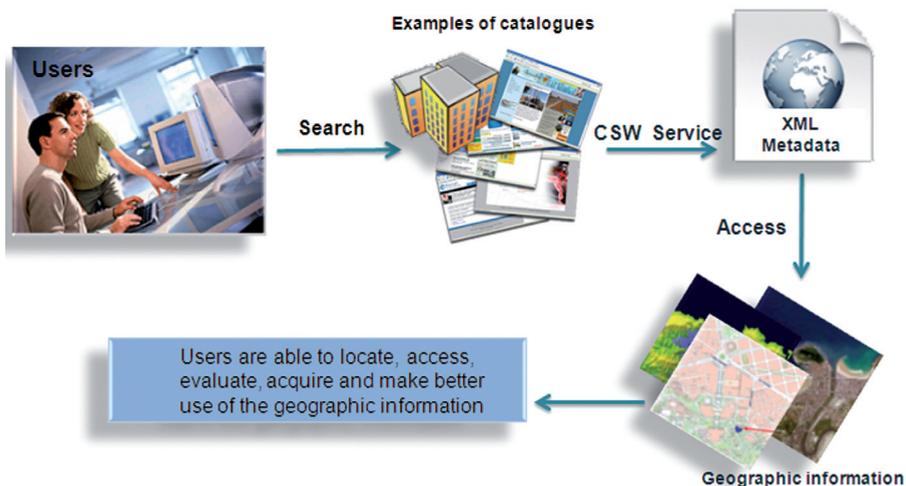
implemented in a consistent and compatible way across Europe, the IGN has taken the initiative to create the Inspire Discovery Service and a catalogue client that gives support to data and service metadata of the resources held by this organization.

The CSW service of the IGN[12] is a discovery service that allows search and retrieval of descriptive information (metadata) about data and services and meets the Inspire Profile of OGC™ *Catalogue Services Specification 2.0.2 - ISO Metadata Application Profile for CSW 2.0* (CSW ISO AP) and the *Technical Guidance for the implementation of Inspire Discovery Services*. The creation of this service has been made with open source software (Geonetwork 2.6.4).

This service implements the required behaviour of a CSW service and all the extensions required by the Inspire Directive.

In order to facilitate its handling, the user may check the geographical products of the IGN through an extensive collection of metadata. The IGN has created a web catalogue[13] (figure 2) where the user finds descriptive information (metadata) of map sheets, digital terrain models, topographical databases, orthoimagery and web map services. If the user wants to buy these products after having consulted this information, he can enter the Download Center [14] or the Virtual Store[15].

The main features of this tool are:

- The cartographic products shown are classified as dataset, series and web services.
- The metadata records for datasets and series are described as ISO 19115 (Metadata Spanish Core model).
- It was made with open source software (Geonetwork 2.6.4).
- It is a bilingual client (accessible both in Spanish and in English).
- Customization of the interface has been made.
- The client includes a map viewer to view the position of the resources.
- It includes a GEORSS channel to show their latest products.

Nowadays, there are a lot of CSW and catalogue clients that integrate metadata, but the amount of Inspire-compliant CSW and client is minimal. Therefore, the IGN, by means of this service and client, has contributed to

Fig. 2: IGN Catalogue web client.

create an example of implementation of a technical guidance for service within Spain, which is a meeting point for people cataloguing geographic information with the aim of stimulating the development of discovery services and therefore for helping develop SDI.

The aim of this catalogue is to include all metadata records of geographic products held by the IGN and to put them at the user´s disposal.

4. Cataloguing and digitizing historical documents at the Technical Archive of the IGN

The Spanish Geographical Institute (IGN) was founded in 1870 following the previous Statistics General Board. Therefore, its Technical Archive, also known as the Topographical Archive, stores both some of the works carried out by the Statistics General Board and the works related to the information-gathering process that was undertaken by the IGN for developing the 1:50,000 scale National Topographical Map (MTN50).

On the one hand, the Cadastral Topography of Spain is the most significant work carried out by the Statistics General Board. It was accomplished around the year 1860, and even though it was not completed throughout Spain, it was indeed finished in many municipalities within the province of Madrid. The documents worth mentioning in this work are the so-called Kilometrical Sheets (Capdevila & Bonilla, 2009).

On the other hand, the works related to the MTN50 were carried out by the IGN according to the *Topographical triangulation and map drawing General Plan*, which was designed by the first Chairman of the IGN, General Ibáñez de Ibero, and which was passed by Parliament on the 30th of September 1870. In order to complete this work, field surveys were based on municipalities and were accomplished in different stages. First, municipalities were outlined by means of drafting boundary line statements between different towns. Later on, triangulations were carried out covering complete municipalities. However, since the geodesic network had not yet been completed, triangulations were often based on astronomic observations and base measuring. Finally, mapping was undertaken following classical topography and comprised 1:25,000 scale planimetric maps, 1:25,000 scale altimetric maps with contour lines every 20 meters, and 1:2,000 or 1:5,000 scale built-up area maps. Due to their scales, all these documents included more precise information and place names than the final map MTN50, which was printed on scale 1:50,000.

The Technical Archive of the IGN followed the same patterns of work from its foundation in 1870 until 1996. During this period, every search in its collection was done by means of looking up in the book in which the entry of the document had been registered. These books were arranged by provinces and each province was organized by municipalities. Furthermore, each search and copy had to be made on the original file. This process meant using a long time for each search as well as a gradual damage of documents. Therefore, in the year 1996 the decision to update and improve the Topographical Archive was made.

In 1996 began the process of cataloguing and digitizing fieldwork notebooks and boundary line statements. These works were prepared in the late 19th Century and early 20th Century for releasing the above mentioned MTN50, and are still nowadays the legal and geometrical definition for most

administrative divisions throughout Spain. On the one hand, the cataloguing was carried out on Invesdoc software, which was based on Oracle-7. On the other hand, the digitization was done in gray scale, 200 dots per inch quality, one jpg file per sheet. An application for handling documents called SID-DAE (Document Information System for Administrative Divisions in Spain) was developed using Visual Basic. This application controlled over 70,000 documents and over 1,000,000 jpg pictures that were stored in approximately 350 CDs. This work was finished in the year 2003, yet it is still updated nowadays with other documents that later have been found, produced or modified.

Between the years 2003 and 2004 and making the most of a removal, all maps, cadastral statements and owner listings held in the Archive were listed in an Access database. On the one hand, maps included over 50,000 mid-19th Century to mid-20th Century cartographic documents with precise geometrical measurements comprising the above-mentioned planimetric, altimetric and built-up area maps and kilometrical sheets. An application called SID-Carto (Document Information System for the Cartography held in the Topographical Archive) was developed for handling these documents using Visual Basic. The digitization of these maps was concluded between the years 2005 and 2006 in a colour scale using a 400 dots per inch resolution. On the other hand, 100,000 cadastral statements and owner listings completed for the "Cadastral Topography of Spain" carried out in the 1860s were also counted and digitized in the year 2003 in an Access database by means of selecting the most important information of each plot. An application called SID-CECA (Document Information System for Cadastral Statements) was developed for handling these documents using Visual Basic. The digitization was accomplished in gray scale, jpg format, 200 dots per inch resolution.

Thus, document conservation and searching was available if the user could visit the Archive. However, since this information was of great use to other sections of the IGN, i.e. Demarcation, Cartography, Geographic Names…, working patterns had to be changed. We could not persist in working with Access databases installed in each PC and with CDs in order to view different maps. Therefore, in the year 2003 we decided to dispense with Invesdoc, move the database to Oracle-9, and install it in a net server together with all images, so that all users within the IGN could access this database. Some months later all data concerning SID-DAE and SID-CECA were also added to the same Oracle-9 database.

Other improvements that were also introduced between the years 2004 and 2008 were the following:

- converting jpg files into a single pdf file per document.
- .net developing of SID-DAE, SID-Carto and SID-CECA.
- georeferencing some maps and cadastral statements held in the Archive.
- developing GIS applications after this georeferenced information (Geodocat, figure 3) was available.
- moving the database from Oracle to PostgreSQL, with all the advantages that an open source software entails.

Thus, the following information, including the three applications that were designed for the different purposes for which documents are acquired, is accessible for all users within the IGN since the year 2009:

Fig. 3: View of GEODOCAT Geographic Information System.
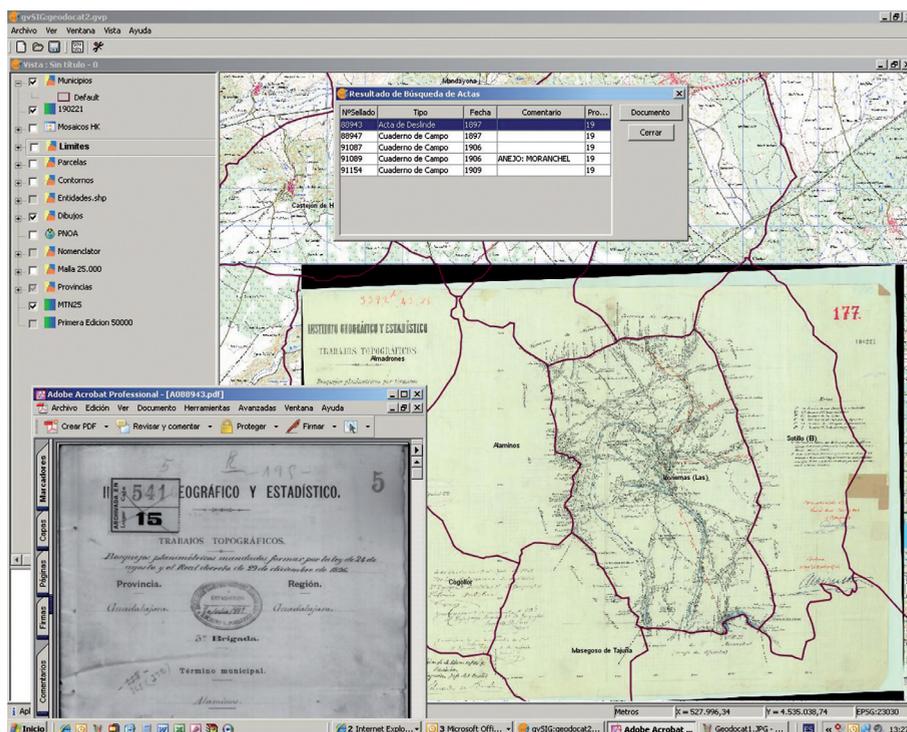
- 80,000 jpg digitized and listed fieldwork notebooks and boundary line statements.
- 50,000 listed maps held in the Topographical Archive, which are also accessible in jpg and ecw georeferenced format.
- 100,000 jpg listed cadastral statements with their coordinates for their GIS location.

However, over 40,000 documents are annually requested from outside the IGN concerning boundary lines between different municipalities, territorial changes, geographic names, etc. Therefore, the next step that had to be faced at the Archive included the possibility to allow all these users to access all this information from their computers. To develop this stage we had to create all metadata for this information and release it in the IGN Metadata Catalogue. Ten patterns were developed for creating metadata, depending on the kind of document considered. An open source application called CatMDEdit was used to conceive these patterns with an Inspire profile. This implied that by means of some patterns and an application developed on .net, we had to be able to publish in a first step over 80,000 metadata concerning fieldwork notebooks and boundary line statements as well as 50,000 maps, and in a second stage we had to divulge over 100,000 cadastral statements with all their owner listings. At present, this work is being gradually published on the IGN Download Center[16], so that it can be unloaded for non-commercial purposes.

5. Conclusions

The IGN is firmly contributing to the development of SDI within Spain by supporting its organizational structures, releasing information using interoperable web services, and upholding the spreading of the SDI paradigm (Rodríguez et al., 2009; Vandenbroucke & Biliouris, 2011). One of the main purposes is to be in compliance with the Inspire Directive. However, we also wish to promote that the release of all sorts of geographic information is carried out using SDI, and especially all data included in the Cartographic Heritage, which is stored in archives, libraries and map collections.

Within this sphere, the first step to be faced is the release of metadata. We have explained in this article how the IGN has developed a CSW service that checks the geographical products through an extensive collection of metadata with over 120,000 metadata registers related to the collection held in the Technical Archive. This CSW service is open source and can be consulted in each user´s web page or geographical information system. In order to help using it, an open source software client has been created, so that all users can apply it for their own purposes.

With this first step, we have shown both the method followed for releasing data and the information included in the cartographic and text collection gathered since the IGN was founded in 1870. The next step, which is being taken at present, includes the massive disclosing of all this information on interoperable web services (WMS, WFS, etc.). By this means, this distinctive and poorly known collection is being placed at the citizen´s disposal in a simple and flexible way.

References

Capdevila, J., & Bonilla, R. (2009). Cartographic patrimony in the Spanish SDI. The cadastral series of nineteenth century: Hojas kilométricas (Kilometric Sheets). *e-Perimetron 4* (1), 38-44. Retrieved November 26, 2012, from http://www.e-perimetron.org/Vol_4_1/Capdevilla_Bonilla.pdf.

Crespo, M., Criado, M., Capdevila, J., Sánchez, A., Soteres, C., Juanatey, M., Bravo, M. J., & Rodriguez, C. (2010). El patrimonio cartográfico en las infraestructuras de datos espaciales. *Revista Catalana de Geografia 15* (41). Retrieved November 26, 2012, from http://www.rcg.cat/articles.php?id=199.

Rodriguez, A., Mas, S., Abad, P., Alonso, J.A., Sánchez, A., & González, C. (2009). Data and services availability in Spanish NSDI. In *GSDI 11 World Conference, Rotterdam, 15-19 June 2009*. Retrieved November 26, 2012, from http://www.gsdi.org/gsdiconf/gsdi11/papers/pdf/161.pdf.

Vandenbroucke, D., & Biliouris, D. (2011). Spatial data infrastructures in Spain: State of play 2011. In: *Spatial Data Infrastructures in Europe: State of play Spring 2011*. Leuven: KU Leuven. Retrieved November 26, 2012, from http://inspire.jrc.ec.europa.eu/reports/stateofplay2011/rcr11ESv120.pdf.

Notes

---

[1] Directive 2003/98/EC of 17 November 2003 on the re-use of public sector information.

[2] Convention on access to information, public participation in decision-making and access to justice in environmental matters, done at Aarhus (Denmark) on 25 June 1998 and ratified on 29 December 2004.

[3] Directive 2007/2/CE of 14 March 2007 on the establishment of a spatial information infrastructure within the European Union (INSPIRE).

[4] Law 14/2010 of 05 July 2010 on infrastructures and geographic information services in Spain.

[5] For further information, check the Spanish Spatial Data Infrastructures geoportal http://www.idee.es.

[6] http://www.opengeospatial.org/.

[7] The last NEM updating is the 1.1 November 2010 version, http://www.idee.es/resources/recomendacionesCSG/NEMv1.1.pdf.

[8] The definition of run can be searched on http://metadatos.ign.es/metadatos/Patrimonio_cartografico and the reference computer installation has been developed on CatMDEdit open source software program http://catmdedit.sourceforge.net/. For further information, check Crespo et al. (2010).

[9] It is a WMS 1.3 service. Url: http://www.idee.es/wms/IGN-1EdicionMTN50/IGN-1EdicionMTN50?.

[10] It is a WMS 1.3 service. Url: http://www.idee.es/wms/IGN-HK/IGN-HK? A description of this project and its content can be looked up in Capdevila and Bonilla (2009).

[11] http://www.opengeospatial.org/standards/specifications/catalog >.

[12] http://www.ign.es/csw-inspire/service/csw>.

[13] http://www.ign.es/csw-inspire.

[14] http://centrodedescargas.cnig.es/CentroDescargas/index.jsp.

[15] http://www.ign.es/ign/main/index.do.

[16] http://centrodedescargas.cnig.es/CentroDescargas/index.jsp.