(Draft) Proceedings of AGILE 2003: 6th AGILE Conference on Geographic Information Science. 2003, p. 267-277

Proceedings of the 6th AGILE
Dates – Lieu

SPATIAL DATA INFRASTRUCTURE IN THE GALICIA REGION: AN INSTANTIATION OF THE INSPIRE FRAMEWORK

R. Béjar[1], J. Aboal[3], M. Gould[2], P.R. Muro-Medrano[1], P. Vila[3]

[1]Department of Computer Science and System Engineering, University of Zaragoza

María de Luna 1, 50018 Zaragoza (Spain)

{rbejar, prmuro}@unizar.es

[2]Department of Information Systems, University Jaume I

Campus de Riu Sec, 12071 Castellón (Spain)

gould@lsi.uji.es

[3]Consellería de Medio Ambiente - Xunta de Galicia

Rúa San Lázaro, s/n Santiago de Compostela 15.781 A Coruña

pedro.vila@xunta.es

Abstract. Public administrations have been facing problems with geographical information use and management for years: difficult software interoperability, incompatible data formats and little knowledge about the available geodata are among the most common ones. This was the case of the Galicia Department of Environment, in charge of managing environmental resources in that NUTS-2 region. INSPIRE work in progress toward a European Spatial Data Infrastructure (ESDI), currently focused on environmental issues, led to a solution based on this European Commission initiative recommendations on architecture and standards. This paper will show that this solution has proven to be both adequate to solve this department's needs, and able to establish itself as the core of a future Galician SDI. This successful example may be a pattern to be followed by other regional governments.

1. INTRODUCTION

The Infrastructure of Spatial Information in Europe (INSPIRE) is a European Commission initiative that concerns the preparation of Community legislation that aims at making available quality geographic information for the purpose of environmental policy-making and for the citizen, in Europe. It also aims, in the longer term, to provide geographic information for the purpose of Community policy-making in a broad range of sectors. These objectives will be achieved through the establishment of integrated spatial information services, based upon a network of geographic databases linked by common standards and protocols [6].

In order to implement the infrastructures proposed by this initiative, the INSPIRE Architecture and Standards Working Group is recommending to define requirements for a European spatial data infrastructure (SDI), in order to provide a reference for national and regional SDIs, allowing thus their future integration in the European one [9]. These

(Draft) Proceedings of AGILE 2003: 6th AGILE Conference on Geographic Information Science. 2003, p. 267-277

requirements include specifications and guidelines as issued by the ISO 19100 series of standards for geographic information, the OpenGIS Consortium or the Dublin Core Metadata Initiative [7][10][2].

Plans for the adoption of INSPIRE framework legislation as early as in 2004, are encouraging EU national and regional governments to start giving effective steps towards the creation of the geographic data and services infrastructures this legislation is establishing (and not only because of INSPIRE, see [1]). This is possible now because the relevant standards and architectures that will be adopted have already been developed, proposed and tested, and there already exist implementations for the different components needed.

This is the case of the Spanish Region of Galicia Department of the Environment (Consellería de Medio Ambiente, Xunta de Galicia, its Galician acronym is CMA). This department had found the same kind of problems with geographic information that INSPIRE addresses: incompatible data formats and information systems, difficulties disseminating data among their users (it is a very decentralized department), difficulties to find relevant information, etc. The solution adopted to overcome these problems has been to develop a geographic information system for this department, following INSPIRE principles and recommendations in architecture and standards, thus effectively building an SDI. This infrastructure has been designed to become the core of a future Galician SDI.

This paper will describe this infrastructure, the established architecture, the components used and the processes involved, such as collecting and storing the data, creating metadata, and defining web map services. It will also describe several problems found and the solutions adopted.

2. GALICIA DEPARTMENT OF THE ENVIRONMENT (CMA)

Galicia is located at the northwest corner of the Iberian Peninsula. The climate is warm and wet so its land is covered with many forests (69% of its surface). This fact makes forests the main concern of the CMA, with water use, disposal of waste and protected natural environments among its other responsibilities.

Galicia is divided into four provinces, but the CMA divides it also into nineteen forest districts, in order to address the necessities (reforestation, forest fires, cleaning…) of such a big forest surface. This is the main reason behind the decentralization of this department, with only 10% if its workers at the central building in Santiago de Compostela, and the rest of them at the provincial delegations and the forest districts. Currently this department is finishing the wiring of all provincial delegations and forest districts. Once the wiring is complete, all delegations and districts will be connected to the department Intranet.

The decentralization of the CMA makes the usual problems with geographic information in big organizations and public administrations much worse. It’s difficult for users to find the data they need or even to find out if that information exists. In some cases, i.e. people in forest districts away from the central building, users just didn’t have any access to geographic information that would make their work much easier. Another problem is the delay in the building of an integrated solution that has led some districts to adopt different GIS software solutions, or no solutions at all in some of them. This situation has added problems of data formats and interoperability.

3. CMA SPATIAL DATA INFRASTRUCTURE

Given the problems related to geographic information management and use in the CMA, and the development of the INSPIRE initiative with the requirements it will impose to EU members in some years, building an SDI following this initiative principles was the best option to address both issues simultaneously. This would solve the CMA geographic information users’ needs while giving some effective steps in order to fulfil the future INSPIRE legislation, making profit of the recent networking of all delegations and districts.

(Draft) Proceedings of AGILE 2003: 6th AGILE Conference on Geographic Information Science. 2003, p. 267-277

3.1 Requirements

The first work was, of course, collecting users' requirements. Here is an overview of the most important ones:

- all geographic information, raster and vectorial coverages, must be stored in a spatial database;
- it will be possible to find relevant geographic information available,
- users will be given tools to view the geographical information they need over the CMA Intranet. They will also be given the possibility to make spatial and non-spatial queries to this information;
- users will also be given the possibility to download the geographical data (vectorial and raster coverages) in the format they are used to work in;
- advanced users will be able to access simple map services, designed to be combinable with their local data, and to query the information viewed in this map services;
- software already owned by the CMA will be used where possible. This includes: ESRI ArcSDE 8.x and ArcIMS 4.x, Oracle 8i/9i with the spatial cartridge, the Safe Software universal translator FME, and its web version Spatial Direct.

Other requirements where specified and recommended to the CMA, after studying their necessities and the guidelines offered by INSPIRE and GSDI [4]; briefly:

- standards must be followed, if possible, where available; specifically those standards recommended by INSPIRE and GSDI ([7][10][2]);
- architecture will follow that recommended by INSPIRE. There will be data, a metadata catalog, and chainable services offering at least visualization, access and data searches.

All this needed to be accomplished in a short period of time (half a year), and there had to be visible results in half that time, in order to convince the decision makers (politicians) to approve the needed funding for improving and expanding this infrastructure. This short deadline has obviously conditioned some parts of this work.

3.2 Web Services Design

Fig. 1 shows a service-centered view of the CMA SDI. This helps to emphasize several important things that are the core of any SDI:

- the main components of an SDI are chainable web services. Through the set-up of standard OGC web services (see [5] for a good overview of them), syntactic interoperability is easily achieved. This allows for an easier integration of different services, both coming from these infrastructure and from any other place, to quickly develop new custom-made or thematic applications in answer to the evolving users' needs;
- it also emphasizes that chainable services are built on top of geodata and metadata. This allows for a better semantic interoperability, as geodata are described by its associated metadata. Including metadata along with reference and thematic data is also a INSPIRE recommendation;
- user applications are built on top of distributed services, both chainable standard services and integration services. Integrated services would be those provided to support extended functionalities, i.e. not covered by the standard ones, not logically related to individual end-user applications.

(Draft) Proceedings of AGILE 2003: 6th AGILE Conference on Geographic Information Science. 2003, p. 267-277

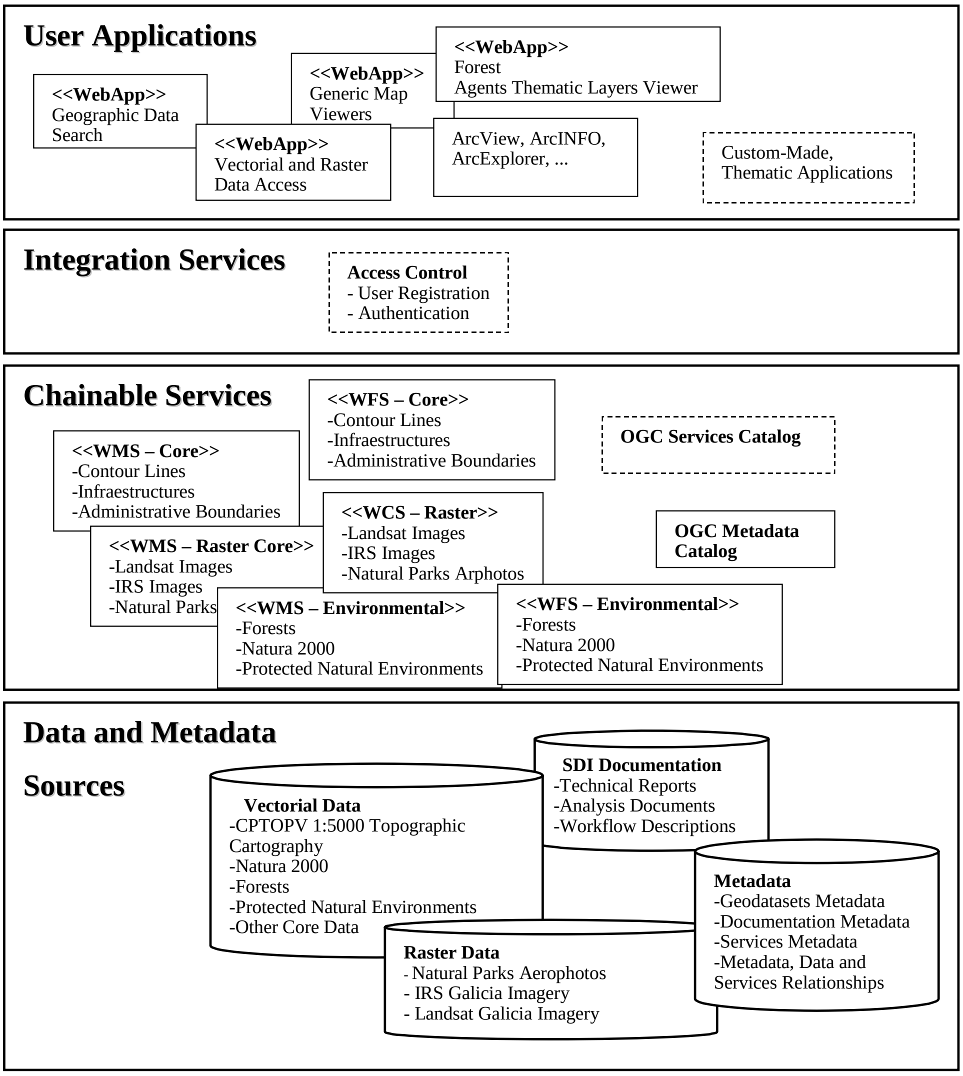

Fig. 1 Service-oriented view. Dotted lines show elements not yet finished or integrated

After studying the users' needs and the available geodata (more on data in the next section), a number of visualization and access web services were planned (OGC standards: Web Map Service (WMS), Web Feature Service (WFS) and Web Coverage Service (WCS) [10]), trying to:

- allow for both visualization and access of all vectorial and raster data,
- allow for an easily customizable integration of different data,
- keep a "controllable" number of web services, aiming at an easier maintenance,
- group related things together, in order to make it more intuitive.

Three simple conceptual categories were decided for the data (core vectorial (reference data, basic topographic information), environmental vectorial (thematic data related to the CMA) and raster) and web services were designed consequently. The planned visualization and access services are shown in Table 1 .

(Draft) Proceedings of AGILE 2003: 6th AGILE Conference on Geographic Information Science. 2003, p. 267-277

	WMS	WFS	WCS
Core Vectorial Data	X	X	
Environmental Vectorial Data	X	X	
Raster Data	X		X

Table 1 Overview of the map, feature and coverage services planned

Besides visualization and access, the other main required functionalities were data searches and metadata management. The OGC catalog service covers these areas, so one metadata catalog was included (a services catalog could also be included in the future).

3.3 Spatial Data and Metadata

Before building a spatial data infrastructure, some spatial data are needed! In order to provide their users with some fundamental core geographic information, the CMA reached an agreement with the Department of Public Works, Land Administration and Housing (its Galician acronym is CPTOPV) to use the 1:5000 Galician cartography they produce, which includes topographic information, administrative boundaries, communication infrastructures, public service infrastructures and some other layers. This cartography is distributed as approx. 4000 dgn files. These files contain nine thematic layers, and are more CAD-oriented than GIS-oriented (i.e. contour lines are broken to accommodate labels!), what makes it quite difficult to extract features from them.

The CMA negotiated with other public entities in Galicia (as well as with private companies) to get Landsat and IRS images covering all their territory, in order to provide their users with raster data, at different resolutions.

In the INSPIRE spirit, buying this information should not have been necessary, because providing access to this information should have been the responsibility of other departments in Galicia government, but this was not the case. The CMA is the first SDI initiative in Galicia, and needing to give their users a useful, and as complete as possible, service, there has been to take some decisions of this kind.

Besides the information already mentioned, the environmental information owned or created by the CMA and made available through this infrastructure includes coverages, mainly shapefiles, of forest management (most of the data), natura 2000, and protected natural environments (including some high resolution ortophotographs of several natural parks).

One of the user requirements was storing all these data in a spatial database. The next two sections describe the process of storing both vectorial and raster data in the database. Finally, there is a section describing the metadata creation process, metadata needed in order to allow for searching data (another user requirement) in a geodata catalog.

3.3.1 Vectorial Data D.B. Storage

The objective was to use the universal translator FME to store all vectorial coverages in Oracle Spatial format, though through SDE in order to allow for an easy access both from ESRI software and for other software directly compatible with Oracle Spatial.

This process, briefly described, consists of designing FME filters (semantic and syntactic mapping between data in one format and data in other format) and applying them to all the data files (coverages) to be translated.

Carrying out this has taken much more time than initially planned, and this fact should be taken into consideration in any similar project: almost 12 man-months have been needed, when initial plans where of 4 to 6 man months, at most, for vectorial data storage. The main problems found, and the solutions applied, were:

- poorly specified data: The CPTOPV 1:5000 cartography, the bulk of the data to be stored in Oracle, was very poorly specified. The aim was splitting the original thematic layers into features, in order to allow for a better manipulation of them. In

(Draft) Proceedings of AGILE 2003: 6th AGILE Conference on Geographic Information Science. 2003, p. 267-277

order to achieve this, a good knowledge of this cartography was needed. After recovering all information (i.e. some metadata) available on these coverages, and interviewing people from the CPTOPV, it was evident there was little control over the data: different companies had made it on different years, and there had not been an as good as possible quality assurance process. As a result there was a lack of reliable information that led to a trial-and-error, slow, process of extracting features from these coverages. Extracted features were then visually matched against the original files to achieve a minimum quality check;

- geometry errors: "False" line segments (segments shorter than a small threshold), self-intersecting polygon boundaries, self-intersecting contour lines and other similar problems were found in the CPTOPV cartography, and in other data files. These errors prevented FME from storing these coverages correctly in Oracle and/or prevented Oracle from creating spatial indexes on them (very important for an efficient access). These kinds of problems made it difficult to create reliable batch processes to store coverages massively. The solution was making more robust FME filters, by including more geometry tests, to cope with these problems before starting the translations. Some other errors were reported to the creators / maintainers of the coverages in order to get them fixed;

3.3.2 Raster Data Storage

Storing the available raster data, satellite and aerial photographs at different resolutions, in SDE / Oracle needed a different approach to because FME doesn't support raster data. This was not really a problem because their translation didn't have semantic problems; nevertheless other issues arise with raster data:

- mosaicing: Satellite and aerial images of large areas come typically in tiles. These tiles may be regular or not, and may present overlapping, white frames around them and a few other problems. Available images, GeoTIFF files, to be included in the CMA SDI presented a variety of these issues:
  - the Landsat image covering the entire region was only one file;
  - the available IRS imagery came in regular tiles, perfectly orthorectified and georeferenced;
  - aerial 1 meter resolution ortophotographs of natural parks came in irregular tiles, with overlapping and contrast / brightness / shadows differences but without frames or other major problems.

Not having adequate software for airphoto mosaicing, a more creative solution had to be found in order to cope with the natural parks ortophotographs mainly. An approach based on ArcObjects, the software components offered by ESRI Arc products, was finally decided. A small visual basic script was written in order to make profit from ArcObjects capabilities to create mosaics from different georeferenced images, even those presenting overlapping, and to store them in a D.B. through SDE. With a higher budget and more time, a more sophisticated approach could have been taken in order to color balance the aerial photographs etc., but the solution adopted was enough to cover the users' needs.

- fast access: One of the main concerns with large images is allowing for a fast visualization of them at all resolutions. The typical approach to this problem consists of creating pyramids of images composed by the original one and several versions of it, all adequately partitioned and at different resolutions. This way, only the pieces of the images at the resolution actually needed by the user are loaded, vastly improving the efficiency. SDE can automatically create these pyramids for stored raster data, so this issue was quite easily solved.

3.3.3 Metadata Creation Process

Creating metadata about, or cataloguing, all this spatial information had two major objectives. The fist one was to organize and maintain the CMA investment in data, preventing data value being lost by losing the knowledge that makes data really useful, and

(Draft) Proceedings of AGILE 2003: 6th AGILE Conference on Geographic Information Science. 2003, p. 267-277

encouraging its appropriate use. The second objective was to provide information to data catalogs, in order to allow for data searches and, in the long term, to facilitate the integration of this SDI with others, following the INSPIRE recommendations on distributed interoperable services, which include distributed catalog services.

There are a number of geospatial metadata standards (some of them still drafts) as the American CSDGM of the Federal Geographic Data Committee [3] and ISO/TC 211 draft international standard DIS 19115 [8], and they had to be applied. These, and other, decisions and guidelines needed to be stated before starting the metadata creation process. The objective was creating metadata as meaningful and complete as possible, with a reasonable effort and to facilitate as much as possible data searches on this metadata. Other point to be taken into consideration was that CMA users were not used to managing or creating metadata, so they needed clear guidelines, education and a simple approach to metadata (it is expected than as they mature as metadata users, it will be easier for them to understand, complete and improve the already created metadata). Thus, the main decisions taken for this project were:

- trying to achieve at least the obligatory ISO metadata fields (more or less 40 from the around 500 fields of the complete standard) when possible;
- using Dublin Core (a generic metadata standard that provides a minimal description of resources for discovery purposes [2]) as a guideline for minimum metadata;
- giving CMA users seminars, and spreading adequate tools, in order to promote metadata usage. The chosen tool for metadata creation was a software developed in our research lab called CatMDEdit, that allows for creating metadata and exporting/importing to/from the different relevant standards (ISO, FGDC, Dublin Core), and working against XML files, or databases (MS. Access, Oracle) [11].

3.4 Design and Implementation

This chapter will describe the design of the CMA SDI from the point of view of the software components deployed and will give an overview of its implementation. Some design decisions will be commented and other interesting issues arisen during the development of this project will be stated.

3.4.1 Component Architecture

The deployment diagram of the CMA SDI is shown in Fig. 2 . This diagram shows the different components used and their interactions. As the figure shows, two Compaq ProLiant servers are currently dedicated to this infrastructure. One of them, carballo.xunta.es, has the database and closely related elements (data, metadata, and ArcSDE). The other one, rimax.xunta.es, contains the web server, the Internet map server (ArcIMS) and the software for downloading vectorial data (Spatial Direct). These servers are currently behind a firewall, and thus are only visible in the CMA Intranet.

(Draft) Proceedings of AGILE 2003: 6th AGILE Conference on Geographic Information Science. 2003, p. 267-277

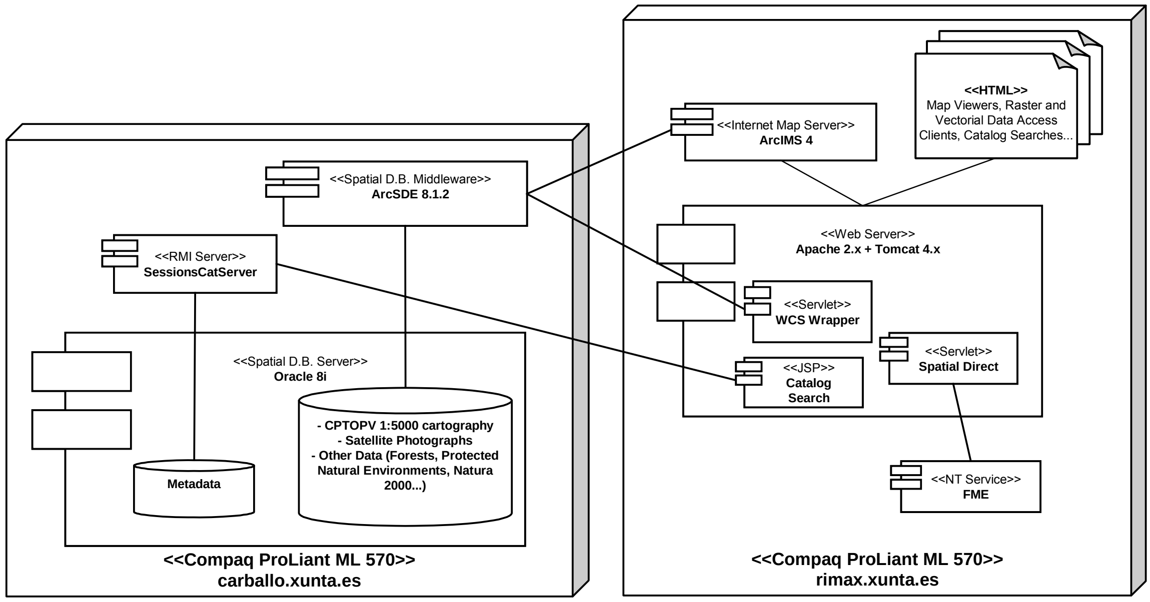

Fig. 2 Deployment diagram

3.4.2 Implementation

This section will give a significant overview of the implementation of CMA SDI, which follows the architecture and design established in the previous chapters.

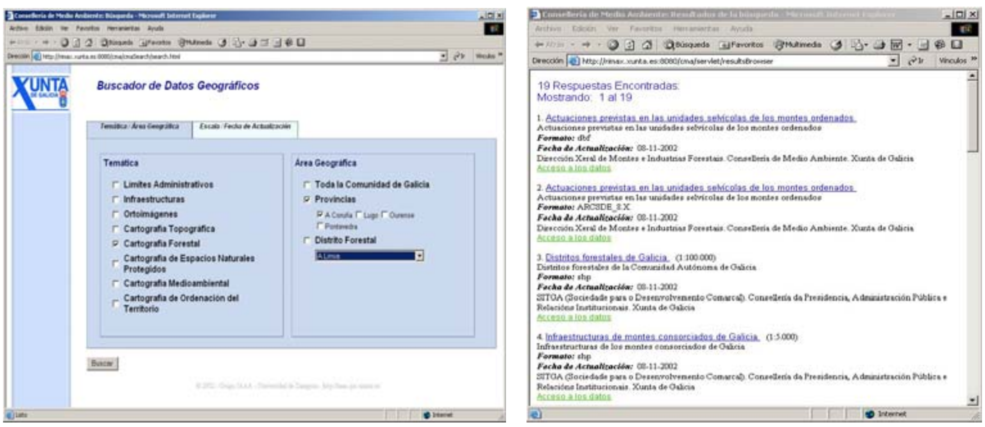

Fig. 3 Thematic data search and found results

As the metadata catalog can be seen as the central element of an SDI architecture, this will be the starting point to describe the implementation. In order to provide data searches in this catalog, a thematic search engine was developed, as can be seen on the left part of Fig. 3 . This search engine provides an interface that combines themes, areas, scales and dates to allow for customized data searches. An example list of the results produced by a search can be seen also in Fig. 3 , on the right. This list shows some metadata for each dataset found (title, abstract, scale, format, date and producer). Clicking

(Draft) Proceedings of AGILE 2003: 6th AGILE Conference on Geographic Information Science. 2003, p. 267-277

the title of an item opens a new window with the complete metadata for that element as shown in Fig. 4 (left side).

The main advantage of the SDI architecture is the use of chainable services. This was proved true when users asked for a way to allow for a connection between data searches and map services, allowing thus to make a search in the catalog and, once found an appropriate dataset, directly access to the map services that showed it. This was implemented by providing a link for each item in the search results list (Fig. 3) that opened a window where the available map services for that item where shown (see Fig. 4 , right side). An extension of the catalog was included to support this functionality, but aiming at the implantation of a services catalog in a near future to support this and other functionalities.

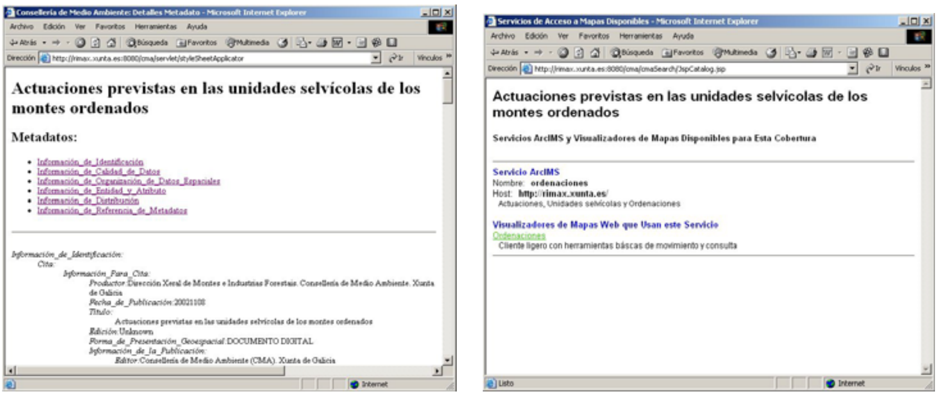

Fig. 4 Complete metadata and available map services for a dataset

After finding an adequate map service for his/her needs, a user can, of course, access to a map viewer showing this service by following a link. An example is shown in Fig. 5 (on the left), where a map service showing Galicia provinces and highways is shown. This is the ArcIMS HTML map viewer, with a little customization to fulfill CMA users' needs. The user can stop here, if geodata visualization is its only need, or he/she can download the data being showed. A link is provided in the map viewers to access Spatial Direct download form (right side of Fig. 5), already customized to provide the area and the data currently selected in the map viewer. This way all services (search, view and access) are connected (chained), giving the users an integrated view of all the SDI elements.

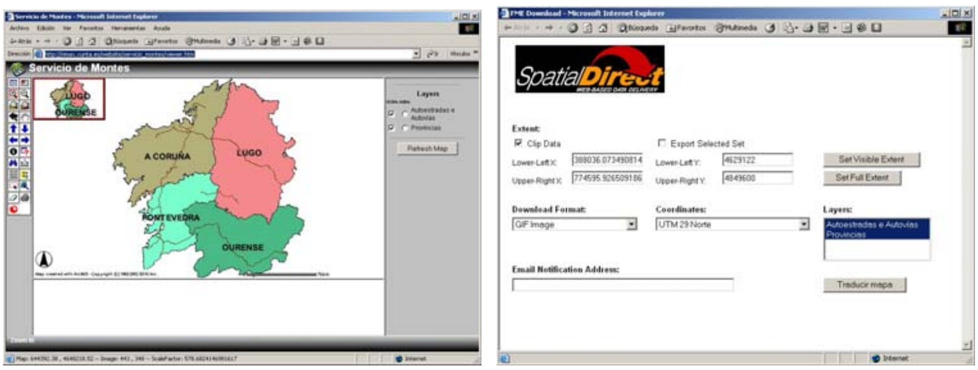

Fig. 5 Web map viewer and coverages download form

(Draft) Proceedings of AGILE 2003: 6th AGILE Conference on Geographic Information Science. 2003, p. 267-277

4. CONCLUSIONS

The Spatial Data Infrastructure developed for the Department of the Environment of Galicia (CMA) has been designed and implemented to address the typical geographic information management and use problems found in big, decentralized companies and public administrations, in particular those found in the CMA, while following INSPIRE recommendations. The requirements have been fulfilled, and the system is now being tested by a selected group of users that will suggest possible improvements and detect possible bugs.

Following INSPIRE recommendations on architecture and standards has proven to be an adequate strategy, both from a technical and a strategic perspective:

- on one hand the chainable web services architecture has allowed for an easier integration of all elements in the infrastructure, while preparing the system for its future integration in bigger initiatives. On the other hand, the emphasis in putting a metadata catalog in the heart of the infrastructure, has shown is usefulness both allowing for a richer semantic description of data, thus encouraging its proper use, and giving a central component to organize the others around;
- CMA's is the first SDI initiative in Galicia. Having chosen INSPIRE recommendations as guidelines, it is now in an unbeatable strategic position to become the core of an SDI for that Region, when these recommendations evolve into Community legislation.

Although currently for internal use, this SDI has been designed to allow for its immediate opening when needed. All implemented map and access services can be accessed through standard OGC Web Map Server and Web Feature Server interfaces. The catalog also follows OGC standards so it will be able to interoperate with others when needed. The appropriate technical steps have thus been given to make this SDI the core of a future Galician one, compatible with INSPIRE legislation when established. Politic and economic factors, and not technical challenges, will now determine if this becomes a reality.

5. ACKNOWLEDGMENTS

The basic technology of this work has been partially supported by the Spanish Ministry of Science and Technology through the projects TIC2000-1568-C03-01 and FIT-150500-2002-347, and by the project P089/2001 of the Council of I+D of the Government of Aragón. The specific design and implementation reported here was sponsored by two R+D contracts to the authors, by the Xunta de Galicia, Conselleria de Medio Ambiente (CMA).

6. REFERENCES

[1] Brox, C., Bishr, Y., Senkler, K., Zens, K., Kuhn, W. (2002). "Toward a geospatial data infrastructure for Northrhine-Westphalia", *Computers, Environment and Urban Systems,* vol. 26, 19-37, 2002.

[2] DCMI. Dublin Core Metadata Initiative homepage. [Online] http://www.dublincore.org

[3] FGDC. (1998). Metadata Ad Hoc Working Group. Document FGDC-STD-001-1998 "Content Standard for Digital Geospatial Metadata". Federal Geographic Data Committee (USA), 1998.

[4] GSDI. Global Spatial Data Infraestructure homepage. [Online] http://www.gsdi.org

[5] Harrison, J. "OGC Web Services. Geoprocessing and the New Web Computing Paradigm". *GeoInformatics*, October/November 2002, 18-21.

[6] INSPIRE. The Infrastructure for Spatial Information in Europe (INSPIRE) initiative homepage. [Online] http://www.ec-gis.org/inspire

[7] ISO. ISO homepage. [Online] http://www.iso.org

[8] ISO. (2001). ISO/TC 211. "Draft International Standard ISO/DIS 19115, Geographic information — Metadata". [Online] http://www.isotc211.org/.

(Draft) Proceedings of AGILE 2003: 6th AGILE Conference on Geographic Information Science. 2003, p. 267-277

[9] JRC, 2002. JRC-Institute for Environment and Sustainability, Ispra (publisher). "INSPIRE Architecture and Standards Position Paper". [Online] http://inspire.jrc.it/reports/position_papers/inspire_ast_pp_v4_2_en.pdf

[10] OGC. OpenGIS Consortium homepage. [Online] http://www.opengis.org

[11] Zarazaga, F.J., López, R., Nogueras, J., Cantán, O., Álvarez, P., Muro-Medrano, P.R. "First Steps to Set Up Java Components for the OpenGIS Catalog Services and its Software Infrastructure". *Proceedings of the 3rd AGILE Conference on Geographic Information Science.* May, 2000.