GIS COTS Integration in a SDI Software Architecture, a Study Case in the Galicia Region SDI

Rubén Béjar[1], M.Á. Latre[1], Michael Gould[2], P.R. Muro-Medrano[1]

[1]Computer Science and Systems Engineering Dep.
University of Zaragoza (Spain)
{rbejar, latre, prmuro}@unizar.es

[2]Department of Information Systems
University Jaume I (Spain)
gould@lsi.uji.es

ABSTRACT

Public administrations have been facing problems with geographical information use and management for years. This was the case of the Galicia Department of Environment (CMA), in charge of managing environmental resources in that Spanish NUTS-2 region. On one hand, INSPIRE work in progress toward a European Spatial Data Infrastructure (ESDI), currently focused on environmental issues, suggested a solution based on INSPIRE recommended standards. On the other hand, the availability of commercial software licenses in the CMA, Oracle Spatial, ESRI products, SAFE software products, and other software, suggested an approach based as much as possible on commercial off the shelf (COTS) software. The developed system architecture integrates heterogeneous COTS products, but is perceived as a homogeneous and standards-based infrastructure.

The main responsible for the perceived homogeneity of this infrastructure are metadata. Metadata catalogs are the core of a SDI, and all available data should have metadata. Many SDI components share these data and metadata: i.e. a map server portrays some data that is already catalogued, and also served by a feature server, but all these components need to have their own metadata (or capabilities). Being able to make profit of the, already created, metadata in the catalog to facilitate creation and management of the SDI services and components would make its administration easier and would prevent from metadata inconsistencies.

This paper presents the solution developed for the CMA, a standards-based infrastructure built upon a COTS architecture and with an emphasis on metadata, which is both useful for solving the CMA current needs and for becoming the core of a future Galician SDI.

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. 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 (INSPIRE), which include specifications and guidelines as the ones issued by the ISO 19100 series of standards for geographic information, the OpenGIS Consortium or the Dublin Core Metadata Initiative (ISO, OGC, DCMI). The most relevant standards and architectures that will be adopted (Brox, Bishr et al. 2002) have already been developed, proposed and tested, and there already exist implementations for the different components needed.

This suggests that a COTS approach for the implementation of Spatial Data Infrastructures can be truly interesting (Tu, Xu et al. 2002): GIS COTS offer complex software capabilities that have already been tested by both the vendor and the market, there have technical support, and the functionality they offer is maintained and may be improved (Albert and Brownsword 2002). In addition to this, the fact that commercial components that are compliant with widely adopted standards, such as the OpenGIS' ones, makes the COTS (that may be called SCOTS, standards-based commercials off-the-shelf (OGC)) approach even more appealing.

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, while using the available commercial software licences in the CMA in a COTS approach. This infrastructure has been designed to become the core of a future Galician SDI (GSDI, Nebert 2001).

This paper is structured as follows: a brief description of the Galicia Department of Environment (CMA) and the problems they face is given in the next section. Then, two views of the SDI Architecture are presented: the service architecture view and the COTS-based architecture; to conclude showing how system homogeneity can be achieved by means of metadata, the use of standards and the user applications.

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. The decentralization of the CMA makes the usual problems with geographic information in big organizations and public administrations much worse. It is 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.

CMA SPATIAL DATA INFRASTRUCTURE ARCHITECTURE

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.

There are also several reasons that leaded us to consider the Galicia Region SDI development as a COTS-based infrastructure:

- Most GIS software offers capabilities that have already been tested by both the vendor and the market. A COTS architecture allows for making profit of these complex capabilities by integrating this software as a component (Albert and Brownsword 2002).
- The availability of commercial software licences in the CMA (products such as ESRI's *ArcIMS*, *ArcMap*, *ArcInfo* and *ArcSDE*, *Oracle 9i* and SAFE's *FME* and *Spatial Direct*). These are complex, and expensive,

products that fulfil, at least partially, requirements of this infrastructure, and cold thus be used for a faster development.

- The fact that, at present, there exist not only standard interfaces an SDI should be based on (the available ArcIMS version is compliant with the OpenGIS WMS 1.0 specification and Oracle spatial format is built around OpenGIS Simple Features specification). Even though not all the components that are in use in the CMA are based on standards, we can still take full advantage of them. If a standard behaviour is needed, they can be adequately wrapped to conform to it; if not, they can still be useful and be used inside the infrastructure to perform some required task.

Service Architecture

According to the architecture reference model provided by INSPIRE, several major groups of components can be distinguished. The instantiation of this model for the CMA SDI is shown in Fig., where dotted lines show elements not yet finished or integrated.

- The main components of this SDI are chainable web services with the main purpose of providing different standard ways of accessing the data. Map visualization of environmental, core and raster data is needed, along with access to the environmental and core geographic features and raster coverage data. Catalog searches of both metadata and services have to be supported too. Through the set-up of standard OGC web services, syntactic interoperability is easily achieved.
- These chainable services are built on top of geodata and metadata. This allows for a better, though ad hoc and very limited, semantic interoperability, as geodata are described by its associated metadata. Including metadata along with reference and thematic data is also a INSPIRE recommendation. In this case, all the data (both vector and raster) and metadata are stored in a shared database with spatial capabilities.
- User applications are built on top of distributed services, both chainable standard services and integration services. They include a catalog search client, a map viewer client and both vector and raster data access clients, customized and linked to allow the users of the infrastructure to exploit it the best the geographic information available while performing they duties. Integrated services would be those provided to support extended functionalities not covered by the standard ones or not logically related to individual end-user applications, such as user registration or authorization.

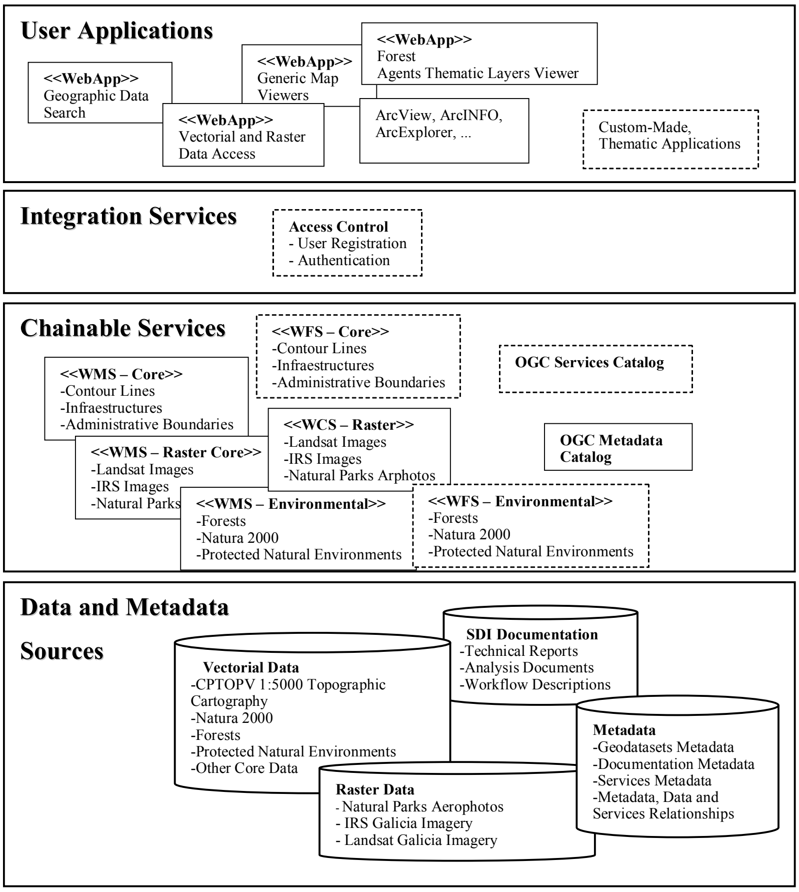

Fig. 1: Service-oriented architecture

COTS Component Architecture

In Fig., the deployment diagram is shown. Those shaded components are commercial software products that have been included into the infrastructure, some of them as commercial off-the-shelf components. How each product has been used as a COTS component is explained below.

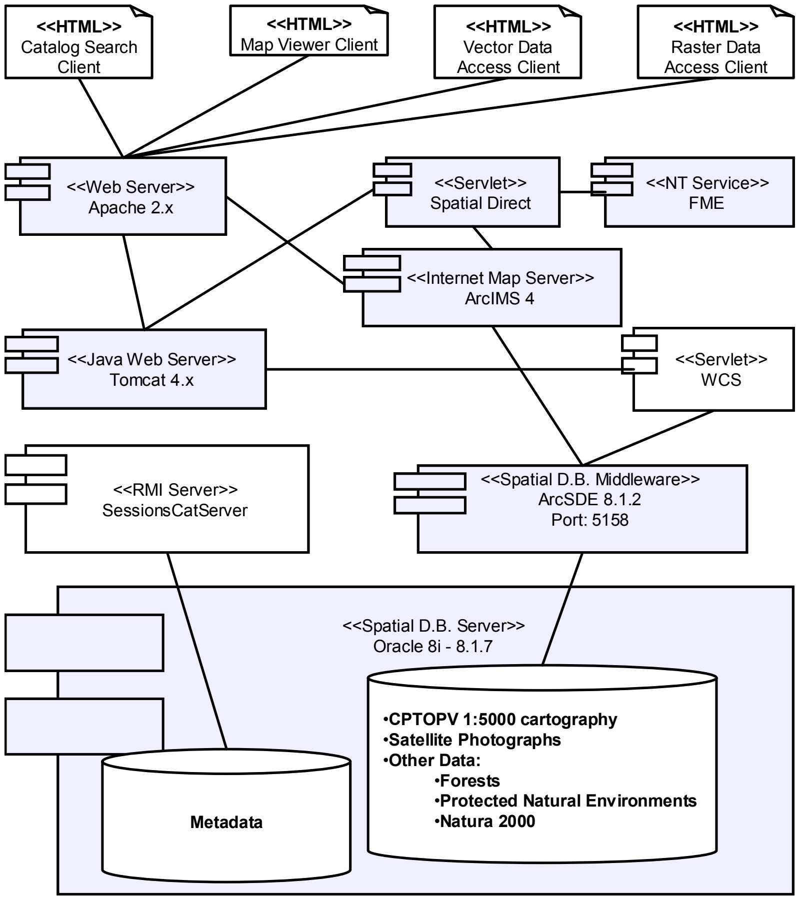

Fig. 2: Deployment diagram

- All the geodata (both vector and raster data) and metadata used in this project are stored in an *Oracle 8i* object-relational database with the spatial cartridge (*Oracle Spatial*) to provide geographic information support (spatial storage format built around OpenGIS Simple Features specification, spatial queries, spatial indexes, …).
- *ArcSDE* is an ESRI gateway that facilitates managing spatial data in different database management systems. It has not been used here to access different databases or to take advantage of the spatial capabilities it

provides, since all the data is stored in *Oracle*, but as a middleware component to provide good compatibility with ESRI products, including of course *ArcIMS* and all the other ESRI applications in use in the CMA, while keeping all the data stored in *Oracle*. *ArcSDE* is used also to give an entry point for other non-ESRI components that need to access the data, such as *Spatial Direct* or the developed WCS wrapper.

Although it has powerful spatial capabilities, in this architecture *ArcSDE* makes use of the spatial management facilities provided by *Oracle* in order to facilitate data access by other software products or components that may be incorporated into the infrastructure in the future. While ESRI applications can access directly the data through *ArcSDE* (and that is the case of the ESRI applications that are installed inside the CMA, such as *ArcMap* or its previous versions, *ArcInfo*, *ArcView* and *ArcCatalog*), other products, such as Intergraph's *GeoMedia*, could access much more easily to the data directly through *Oracle*.

- *ArcIMS 4*, ESRI's Internet Map Server, is a software product that is able to produce representations (images) and to deliver content (vector data) of maps through the Web. Since it is compatible with the OpenGIS WMS 1.0 specification and with the WFS 1.0 specification by the use of connectors, it is a standards-based commercial off-the-shelf (SCOTS) that has been used to provide the needed OpenGIS standard Web Map Service (WMS) and, in a close future, OpenGIS standard Web Feature Service.
- A metadata catalog implements a set of service interfaces that support management, discovery, and access of geospatial information, following the OGC Catalog Services specification. This component has been developed from scratch by our research laboratory at University of Zaragoza (Zarazaga, Lópes et al. 2000, Muro-Medrano, Nogueras et al. 2003), but not specifically for this project, so, in spite of not being a commercial product, it follows the SCOTS philosophy.
- Vector data could be provided in a standard way by the use of a Web Feature Server. While the WFS is not integrated into the infrastructure yet, and because the user will usually need the data in some specific format, *Spatial Direct 2002*, a *SAFE* software component for the internet download of vector data is used here in combination with *FME*, a component for geodata format transformation to allow the download of vector data in a variety of formats and spatial reference systems. Although *Spatial Direct* is not based on standard interfaces, its functionality can be used to perform this task and the component can be easily integrated in the system.

- A Web Coverage Server wrapper had to be built on top of *ArcSDE* to provide raster data. At the moment the infrastructure was built there were no commercial products compliant with the OpenGIS WCS specification and given the internal capabilities of *ArcSDE* for managing raster coverages, a Java servlet accesses the *ArcSDE* functionality through its C language API (since the Java API is not completely implemented) and offers access to raster data through a subset of the too much complex interfaces of the OGC WCS standard specification. This is the only component, apart from the final user applications, that has been built from scratch and specifically for this project.

GIVING HOMOGENEOUS VIEWS TO THE INFRASTRUCTURE

Homogeneous view for users

Creating a standards-based infrastructure is a good practice form the point of view of the administrators or advanced users, and creating it from a COTS approach, useful from the point of view of the developers of the infrastructure. However, users at the CMA will not care about the standards used and should not have to notice the heterogeneity of the software components that are included into the system.

The perceived homogeneity by the users of this infrastructure is provided by the applications that constitute its upper level. They take the form of HTML clients with custom-made graphic user interfaces and are built on top of both chainable standard services and non-standard functionalities provided by certain software products. The users of the SDI perceive an integrated web application that allows them to search, visualize and download both the raster and vector data that is held within the infrastructure.

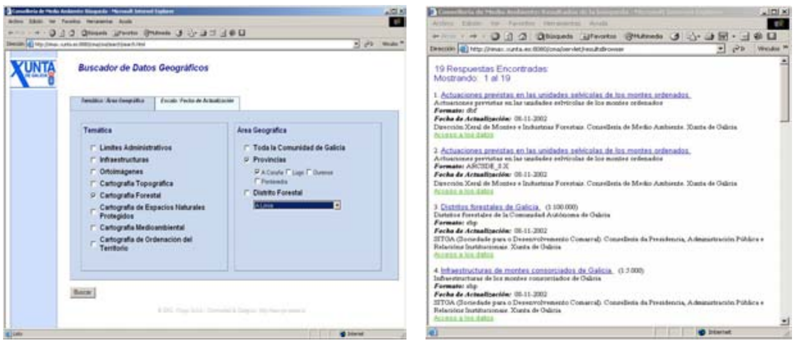

Fig. 3: Thematic data search and found results

In order to provide data searches in this catalog, a thematic search engine was developed (Fig. , left). It provides an interface that combines themes, areas, scales and dates to allow for customized data searches. The list results produced by a search (Fig. , right) shows some metadata for each dataset found (title, abstract, scale, format, date and producer), which can be expanded in a new window to the complete metadata.

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 map services that showed it. This was implemented by providing a link for each item in the search results list (Fig. , right) that opened a window where the available map services for that item where shown.

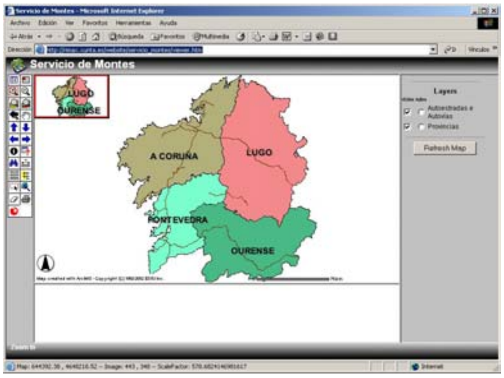

Fig. 4: Web map viewer

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 (Fig.). This is the *ArcIMS* HTML map viewer, with a little customization to fulfil 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, in the case of vector data, or the client that access the Web Coverage Server in the case of raster data, both of them 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 (linked), giving the users an integrated view of all the SDI elements.

This client linking could not have been accomplished without the availability metadata, so the final responsible for the apparent homogeneity perceived by users is the metadata. Many SDI components share these data and metadata, but, what is more, all these components need to have their own metadata (or capabilities). The metadata about the services and user applications is also stored in the database, providing the core of what will be a service catalog. But, the key role of metadata in the homogeneity of this infrastructure is achieved by considering also the relationships between the services and user applications and the geodata they use. This is what allows for better service chaining and even for a certain semantic interoperability on the user applications, allowing them to be linked to each other.

Homogeneous view for developers and advanced users

Developers and advanced users homogeneity lies in both the use of standards and in metadata. A subset of the components included in the infrastructure can be access conforming to OpenGIS open standard interfaces, which allow syntactic homogeneity and hide the underlying software products heterogeneity. At the present moment, the infrastructure is not open to outside entities, but it will be in a close future, and all the standard elements will be fully accessible without any major additional effort. For instance, users of a Portuguese SDI could access the environmental border data of the CMA SDI in a standard way.

Also metadata plays a role in the homogeneous view of the infrastructure, provided that metadata has been created in a coherent manner and avoiding unnecessary duplicities. Consistency problems and update and version administration tasks of geodata are much easier when data is centrally stored and quality metadata available. Also, holding the relations between geodata and services and among services themselves allow a better control of the SDI, providing a way to make the integration and management of the interoperable services and the user applications much more systematic and thus, more automatic, facilitating a great deal their management and traceability.

CONCLUSIONS & FUTURE WORK

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 and a COTS-based architecture.

Following INSPIRE recommendations on architecture and standards has proven to be an adequate strategy, on one hand because the proposed 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.

The COTS approach has proved to be really useful, cutting down the time needed for developing and installing the system by relying in the tested functionality provided by the commercial components already available at the CMA

The time devoted to the project has been spend mainly in analysis and design tasks (the latter consisting in deciding to integrate the available software products into the infrastructure), loading data into the database, developing visualization styles for the data and configuring and adjusting the software while there was just the need of developing small ones needed and in integrating all of these together. In fact, data loading was the most time-consuming task (nearly 3 man months), due mainly to the data volume, disparity and inconsistency (Béjar, Aboal et al. 2003).

This approach has made it possible to develop the Galicia Region SDI, form scratch, in a six-month period. The COTS approach to the development of the SDI also had allowed the developers to take full advantage of all the previous knowledge they had about the products they were integrating.

The Spatial Data Infrastructure development is not yet complete. A second phase of this project is currently planned, and will provide the infrastructure with multilingual catalog capabilities; OpenGIS standard Web Features Servers and Gazetteer Server (to cope with the complex toponymy problems in the Galicia Region (García Pazos 2001)), the former is planned to built on top of *ArcIMS* and the latter on top of the Web Feature Server; and several specialized web users' applications.

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, Consellería de Medio Ambiente (CMA).

REFERENCES

Albert, C., L. Brownsword (2002): *Meeting the challenges of commercial-off-the shelf (COTS) products: the information technology solutions evolution process (ITSEP)*. Proceedings of the First International Conference on COTS-Based Software Systems (ICCBSS 2002). LNCS 2255, Springer-Verlag, Heidelberg, Germany.

Béjar, R., J. Aboal, M. Gould, P.R. Muro-Medrano, P. Vila (2003): *Spatial Data Infrastructure in the Galicia Region: An Instantiation of the INSPIRE Framework.* Proceedings of the 6th AGILE Conference on Geographic Information Science, Presses Polytechniques et universitaires Romandes, 267-277.

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

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

FGDC (1998): Metadata Ad Hoc Working Group. Document FGDC-STD-001-1998: *Content Standard for Digital Geospatial Metadata*. Federal Geographic Data Committee (USA).

García Pazos, F. (2001): *Proyecto de Toponimia de Galicia Thesaurus Toponímico y su Integración Cartográfica*. Mapping, June 2001.

GSDI. *Global Spatial Data Infrastructure homepage*. [Online] http://www.gsdi.org.

Harrison, J (2002): *OGC Web Services. Geoprocessing and the New Web Computing Paradigm*. GeoInformatics, October/November 2002, 18-21.

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

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

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

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

Muro-Medrano, P.R., J. Nogueras, M.P. Torres, F.J. Zarazaga (2003): *Web catalog services of geographic information, an OpenGIS based approach in benefit of Interoperability.* Proceedings of the 6th AGILE Conference on Geographic Information Science, Presses Polytechniques et universitaires Romandes, 169-177.

Nebert, D.D. (2001): *Developing Spatial Data Infrastructures: The SDI Cookbook.* [Online] http://www.gsdi.org/pubs/cookbook/.

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

Tu, S., L. Xu, M. Abdelguerfi, J. Ratcliff (2002): *Achieving Interoperability for integration of Heterogeneous COTS Geographic Information Systems*. ACM-GIS'02, November 8-9, 162-167.

Zarazaga, F.J., R. López, J. Nogueras, O. Cantán, P. Álvarez, P.R. Muro-Medrano (2000): *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.