Cataloguing and recovering distributed geospatial data, a Java approach to build the OpenGIS Catalog Services

F.J. Zarazaga-Soria[1], R. López, J. Nogueras, O. Cantán, P. Álvarez, P.R. Muro-Medrano
Department of Computer Science and System Engineering
University of Zaragoza
María de Luna 3, 50015 Zaragoza (Spain)
[1]javy@posta.unizar.es
http://iaaa.cps.unizar.es

Abstract
Much geospatial data is available on the web and in off-line archives, but it is complex, heterogeneous, and incompatible. Common interfaces are the only way to enable overlays and combinations of complex and essentially different kinds of geographic information to happen automatically over the Internet. The OpenGIS Consortium (OGC) uses the term "Catalog" to describe the set of service interfaces that supports organization, discovery, and access of geospatial information. Catalog services help users or application software to find information located anywhere within a distributed computing environment. This paper presents our experience in a Java approach to build Catalog services guided by the OGC Catalog Services specification, and shows the advantages provided by the use of Java to support Catalog services on the Internet.

Introduction
Geospatial data are being collecting during more than 35 years. This collection speed increases quickly with new technologies in high resolution satellites, GPS, data bases, new software technologies for processing geospatial data and the increase of people and organizations which are collecting and using this kind of data [1]. About 80% of data bases used by public administrations have geospatial references (addresses, city distributions, cartographic coordinates, etc) [2]. From other point of view, the geographic information production market in Europe is about 10 billions of Euros per year [3]. Information is being collected, but the market does not exploit it correctly. In many cases, geospatial data-consuming companies or people do not find the data they need and they usually pay data suppliers for custom-made products.

Most of advanced countries around the world, including USA and most of the European countries, are building their own national geospatial information infrastructures to centralize and co-ordinate efforts for optimizing the geospatial data production process, trying to avoid the duplicity of works in this way. According to [4], to do this it is necessary to have some tools to classify and catalogue geospatial data from different sources, and for publishing this information (Internet is this tool). Once catalogues have been created, end users can look for the data they need into them. These catalogues must offer services and tools for cataloguing and searching geospatial data. Furthermore, they must have the ability to connect to other catalogues in order to response as well as it could be to search services. It is necessary they can inter-operate among them. This way, the standardisation of services and information exchange is necessary to give the opportunity of communication among different catalogues. The OpenGIS Consortium [5] (OGC) uses the term "Catalog" to describe the set of service interfaces which support organization, discovery, and access of geospatial information. These services belong to the Domain Access services which are inside the Open Exchange services. All these kinds of services are managed as Common Applications by the OGC Technical Reference Model (see [6]).

On other hand, the Java programming language has gained quick and large acceptance as a prominent programming language in recent years [7]. Its programming capabilities, multi-platform implementations, facilities for developing distributed applications with powerful built in high level mechanisms of communication, its powerful standard libraries and the extensive availability of public domain software have converted Java into one of the most suitable tools for software development. In our days Internet and the web are, with no doubt, the preferred medium for sharing information and one of the most powerful mechanisms to provide access to public information where user interactivity is required in some way. In addition to the features explained above, one of the reasons for the Java impact in the programming community has been its ability to integrate in the web.

Efforts of standard organizations have been displaced to the Java and Internet worlds (the OGC has prioritized Java Implementation and Web services in its latest standards). This paper presents our experience in a Java approach to build Catalog services guided by the OGC Catalog Services specification, and shows the advantages provided by the use of Java to support Catalog services on the Internet.

The rest of the paper has been structured as follows. Next section presents the catalogue architecture and its relation with the overall project. Section three shows the design of the components which represents the geospatial information metadata managed by the Catalog. In section four some services and tools out of the scope of the OpenGIS Catalog are presented. This paper ends with a conclusions and future work section.

Catalogue services software architecture

A set of Internet services is being built to allow cataloguing and searching of geospatial information through their metadata. Figure 1 presents the architecture of these services. The OpenGIS compliant catalogue has the responsibility of offering components to satisfy search services (to discover information and to present results) and catalogue management services (to register and maintain the metadata into the database). The search services access to the Discovery Services which have the ability to access to other remote OpenGIS Discovery Services.

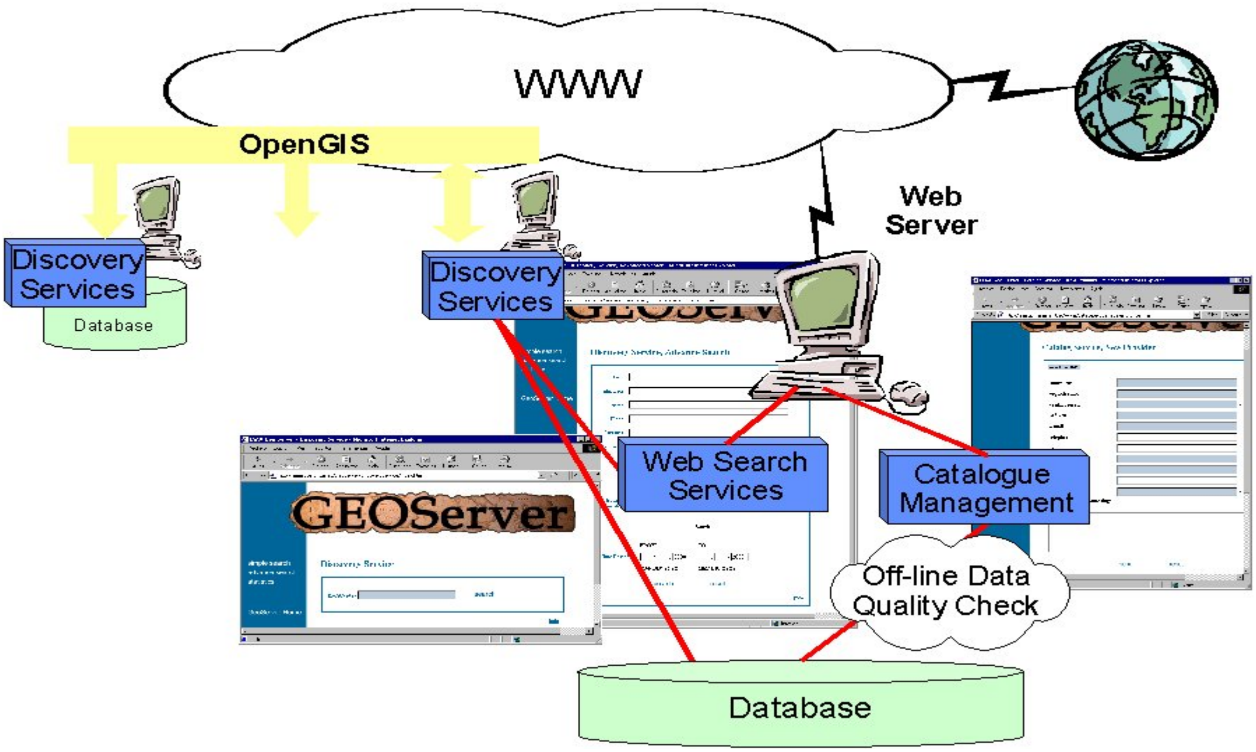

Figure 1: Internet services architecture

Several components and tools provide these services. Maybe, the most relevant one is the Catalog (Figure 2 describes its architecture). Basically, the *Catalog kernel* implements the services described by the OpenGIS Catalog Services specification proposed in the Coarse-Grain Structural Model [8]. It has three kinds of functions: one for Catalog management (named Catalog Manager), another for recovering the stored information through their metadata description (named Access Service) and the third one to search within the catalogue and present the metadata found (named Discovery Service). These services access to three different databases: metadata, Controlled keywords and Knowledge representation. The first one is the most important because it stores the metadata which describes the geospatial information according with different standards. The *Catalog Manager Service* has the responsibility of create and maintenance this information, while *Discovery Service* searches on it using restrictions fixed by other application or the user. The controlled keywords database stores sets of terms representing concepts in a subject domain and relations among them. They are used to facilitate the mapping between a selected vocabulary and a large collection of datasets, offering an indexing capability which increases the performance of catalogue queries based on controlled sections. This process of classification is assisted by recognised authorities and organisations which provide essential indices and lists. Some examples of these lists are: "Discipline" and "Parameter Type" controlled lists provided by CEO [11]; "NASA Master Directory" (see [13]) proposed by FGDC as a standard thematic thesaurus; or "ADL Feature Type Thesaurus" (see [14]) to indicate the nature of a geographic place. Anyway, it is also possible to develop a special Controlled List for more specific subject domains not considered yet in previous lists. One example could be the development we have made to classify the geospatial data generated by the Ebro River Basin Organisation in Spain (http://oph.chebro.es).

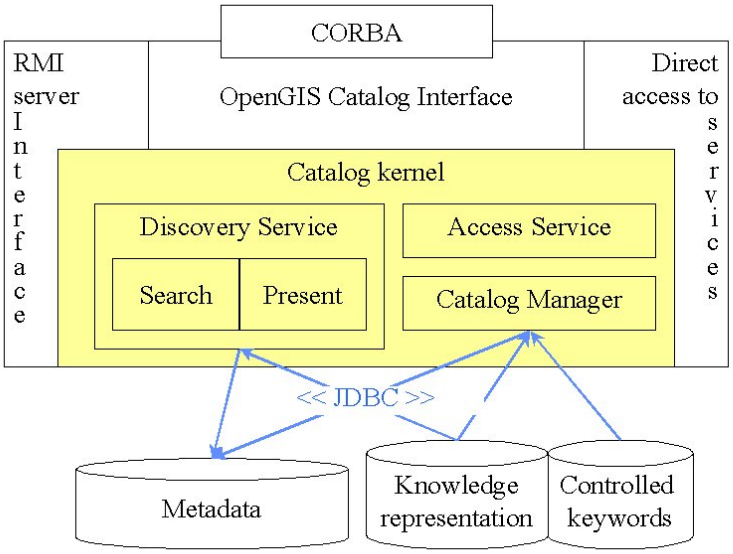

Figure 2: Catalogue architecture

Knowledge representation database stores information about special relations among terms which depends on the context where the geospatial information is related to. One example of these special relations could be the case of Political Units Organisation of Spain. This way, it is possible to classify datasets according to Political Unit they cover. This Political Units Organisation could also be used to increase the capability of the *Discovery Service* because it could be used to expand the queries. For example, in order to find spatial references of a political unit, such as a village, the query may be expanded with the knowledge of political subdivisions such as the subdivision of a country in regions, a region in villages, etc. Data at the level of village may be not available, but data of superior levels which include the village may be sufficient to answer the question.

There are three different types of interface to access to the *Catalog kernel* services. The first one is the *OpenGIS Catalog Interface* which allows others OpenGIS compliant catalogues to access to it. The second one is the *RMI server interface*. This interface offers the possibility of install our catalogue component as a RMI server in a particular installation, out of the scope of OpenGIS. Finally, the third one permits to include the component inside an application, not working like a server. These two interfaces offer a supra set of services, which permits to increase the functionality offered by the OpenGIS catalogues.

Metadata support

There are several standards or pre-standards in the market identifying the metadata elements that describe consistently a specific geospatial data resource. Some examples of these standards are American CSDGM of FGDC [9], ISO/TC 211 committee draft ISO CD 15046 (see [10]), CEO-Recommendations on Metadata [11], and European CEN/TC 287 prENV 12657 (see [12]). However, most of these standards share a common core of elements distributed in different ways.

The metadata object model we have designed (Figure 3 shows a subset of it) follows basically the FGDC standard (CSDGM), with some additional features such as language of metadata element descriptions. Although we must take in mind that there is still no commitment about what metadata standard is going to 100% succeed in the future. Therefore our focus is not to implement exactly a unique standard but to be compliant with different standards. To be compliant does not mean create a new standard consisting of an increasingly number of metadata elements covering all possibilities but to make efforts on providing translation services from/to different standards.

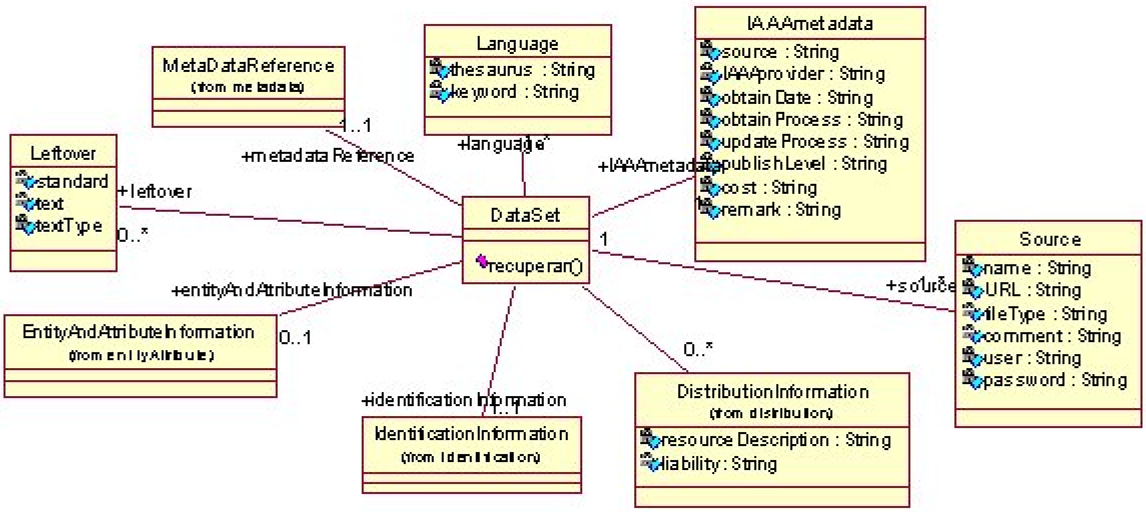

Figure 3: Metadata Object Model

The persistence of the metadata model in the database has been done using an Oracle 8i relational database server. This engine by means of its spatial package offers interesting utilities for geospatial data management. This object model has been completed with a text field containing all the sections not included directly in the previous schema but available in the cataloguing process (for example: importing an XML file with metadata according to FGDC standard). The Entity Relation schema gives persistence to the core elements which are considered to be the more likely to be searched, and other sections interesting for geospatial data identification and description. On the other hand, the text field allows us to give support to all the sections of the rest of the standard used to create the metadata without having to worry about creating a new column for every text-descriptive metadata element.

The implementation of the controlled keywords we have built is based on the concept of thesaurus, and allows the definition of relationships among terms such as hierarchy, equivalence (such as synonyms), association and national language support. Another reason to choose Oracle 8i database server has been its support for thesaurus in the database. Oracle 8i provides a thesaurus package, named CTX_THES, which enables controlled keywords management.

Additional Services and Tools

Out of the OpenGIS Consortium Catalog specification, a set of libraries and tools has been developed to enable internal catalogue management tasks. Maybe the most relevant of them are the next ones:

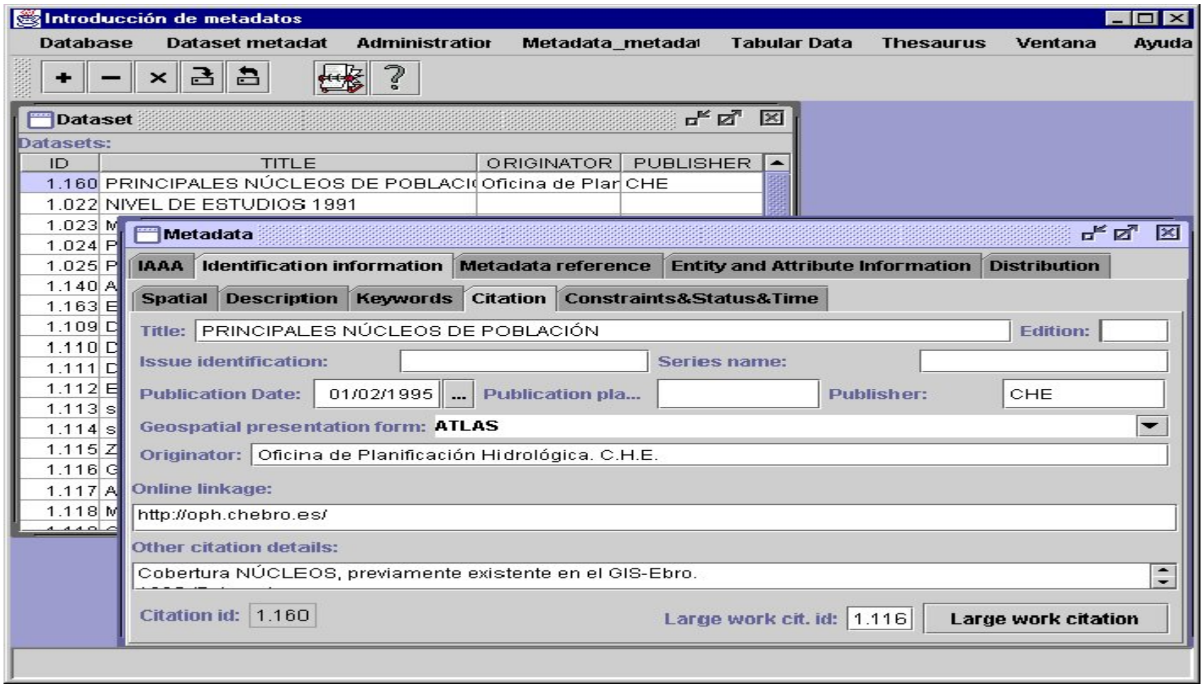

Figure 4: Metadata Edition

- *Metadata Edition* tool, as its own name indicates, enables edition and visualisation of metadata entries for local purposes in our catalogue. Figure 4 shows the tab windows aspect of this tool.

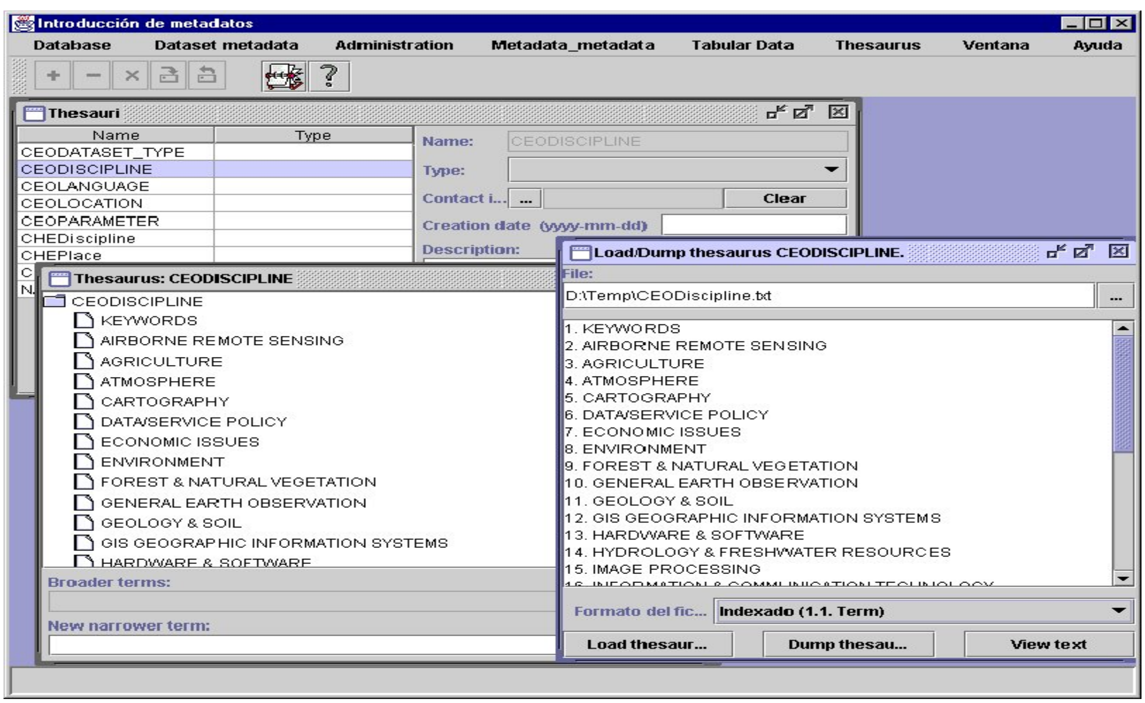

Figure 5: Controlled Keywords Management

- *Controlled Keywords* management tool permits the management of thesaurus supported by our Controlled Keywords database. The main functions of this tool are: Creation/Deletion/Modification of thesauri; Edition/Visualisation of terms in a hierarchical structure; and Import/Export from/to text files in different formats. In the future, metadata suppliers will be given the chance to introduce their own Controlled Keyword list over our Web Server. Figure 5 shows this tool.

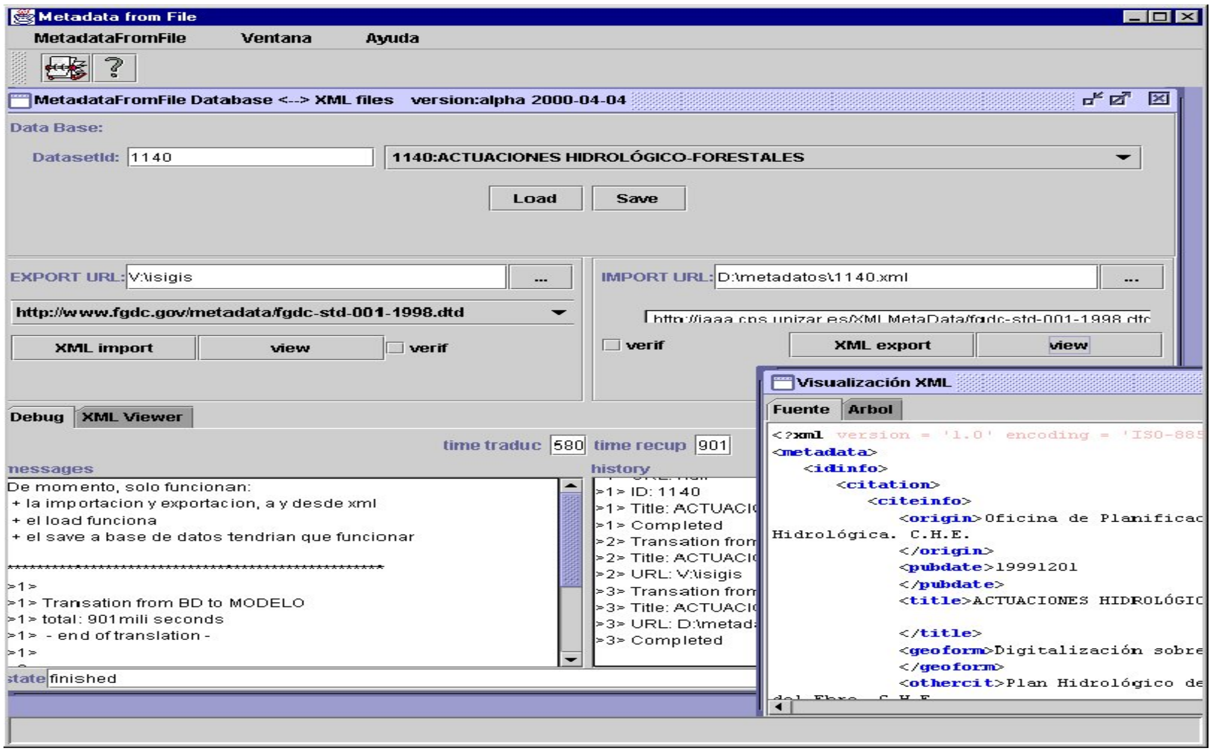

Figure 6: XML Import/Export

- *XML Import/Export* tool (see Figure 6) enables translation between XML files (plain text files) conforming to CSDGM (of FGDC) and our metadata object model. As we mentioned in section 3, the way to intercommunicate metadata was through XML file format conforming to standards. In the future, conversion to other possible standards (ISO TC211, CEN/TC 287) will also be available.

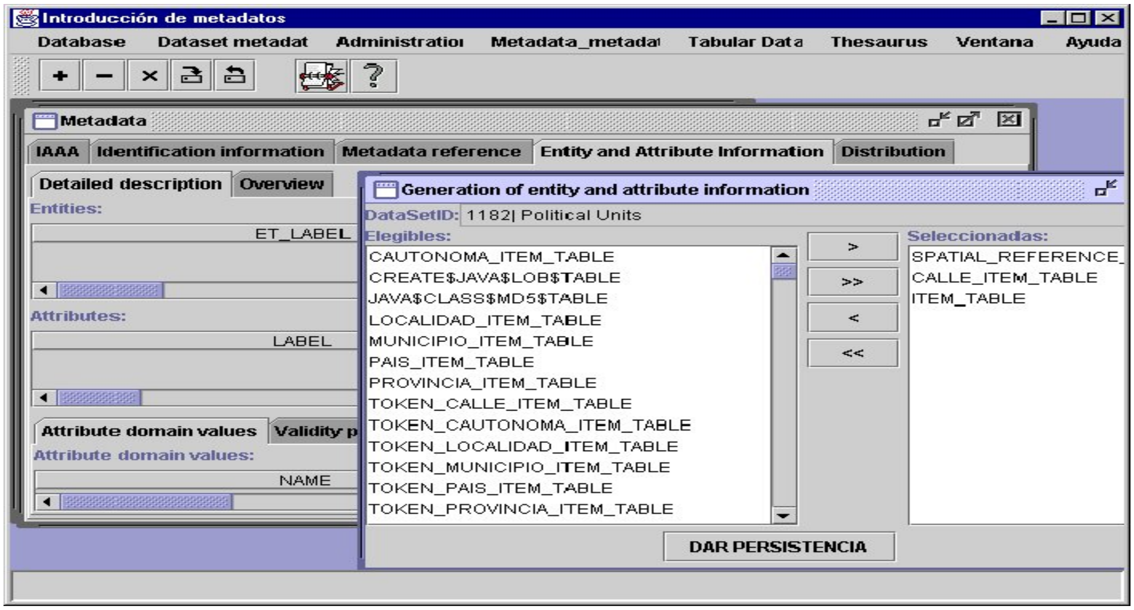

Figure 7: Metadata generation

- The *Metadata generation* tool (see Figure 7) enables the semi-automatic generation of metadata by means of analysis of a geospatial information data source. An example to understand the utility of this software is a component for obtaining descriptive data from an ESRI shape file usually used by ArcView. Other utilities of this tool are the extraction of metadata from Arc-Info files (although the arrival of Arc-Info 8 has not got much sense); or the analysis of the relational structure of a tabular source (Excel, Access, Oracle, ...) in order to obtain automatically the "Entity And Attribute section" of CSDGM.

Conclusions

This paper has presented our experience in a Java approach to build Catalog services guided by the OpenGIS Consortium Catalog Services specification. The use of Java to support catalogue services on the Internet is making our labour easier for different reasons. First of all, we have the classic benefits of using Java such as multi-platform execution and integration into web pages. In the same way, as same software can be used over different platforms, if the size of the catalogue information grows or the use of the web server increases, it is possible to use bigger machines, even with different operative systems, without any change in the software. This way, the use of a database server such as Oracle, with versions for different types of machines, completes this machine growth possibility. Furthermore, Oracle 8i offers utilities enabling the inclusion of Java programs into the database server. This Java integration in the database is a good chance for making the search services faster since it reduces the communication process between the catalogue server and the database.

One of the most relevant problems we have found is the un-standardised world or proliferation of geospatial metadata standards. It has been necessary to prepare our components to be adapted to new standards, and new versions of the current ones, easily.

The process to build the Catalog has begun with the development of the services, and goes to adjust to the OpenGIS Interface Specification. This permits to delay the implementation of the interface allowing the incorporation of latest modifications of the OpenGIS specification. Moreover, this developing policy provides the possibility of showing cataloguing functionality earlier in the building process.

Currently, components are evolving to be in compliance with OpenGIS Coarse-Grain Structural Model interface. Next step should be to offer Discovery Services as a CORBA server which could be accessed by other OpenGIS Catalogs or other web applications.

Finally, we will extend metadata management capacity to support new standards, and new versions of the current ones as soon as they appear or are required.

Acknowledgements

This work has been partially supported by the Commission for Science and Technology of Spain through the project TIC98-0587.

References

1. Buehler, K., McKee, L. (eds.). *The OpenGIS Guide. Introduction to Interoperable Geoprocessing. Part I of the Open Geodata Interoperability Specification (OGIS).* OGIS Project6 Technical Commitee of the Open GIS Consorcium Inc. OGIS TC Document 96-001. 1996.
2. OpenGIS Project Document 99-113. OpenGIS Consortium 1999. *The OpenGIS Abstract Specification. Topic13: Catalog Services (version 4).*
3. European Commission, DG XIII-E. *GI & GIS Market Size Data* (http://158.169.50.95:10080/gi/en/markets/gimrktfigs.html)
4. Coleman, D.J., Nebert, D.D. "Building a North American Spatial Data Infrastructure" *Cartography and Geographic Information Systems.* Vol. 25, No. 3, pp. 151-160. Jul.1998.
5. OpenGIS Consortium Home page. http://www.opengis.org
6. OpenGIS Project Document 99-112, OpenGIS Consortium 1999. *The OpenGIS Specification Model. Topic 12: The OpenGIS Service Architecture (version 32).*
7. Sorokine, A., Merzliakova, I. “Interactive map applet for illustrative purposes”. *Proceedings of the 6th International Symposium on Advances in Geographic Information Systems*. pp. 46—51, 1998.
8. OpenGIS Project Document 99-051s, OpenGIS Consortium 1999. *OpenGIS – Catalog Interface Implementation Specification (version 1.0).*
9. Metadata Ad Hoc Working Group. *Document FGDC-STD-001-1998, Federal Geographic Data Committee (USA) 1998.* Content Standard for Digital Geospatial Metadata.
10. ISO/TC 211 committee draft for Geographic Information Metadata. ISO CD 15046 part 15, dated 1998. http://www.statkart.no/isotc211/
11. A User Guide provided by the Center for Earth Observation Programme (CEO programme) Of the European Commission. February 1999. *Recommendations on Metadata. Describing the data, services and information you have available! (version 2.0).*
12. European Committee for Standardization (CEN), The CEN/TC 287 Geographic Information European Prestandards, ENV 12657. Euro-norme Voluntaire for Geographic Information-Data description - Metadata. http://www.cenorm.be/
13. NASA Master Directory. Draft geospatial thematic keywords from the NASA Master Directory in short and long format for CSDGM of FGDC http://www.fgdc.gov/clearinghouse/reference/refmat.html
14. Alexandria Digital Library Project. http://www.alexandria.ucsb.edu

Proceedings of the 6th EC-GI&GIS Workshop: The Spatial Information Society - Shaping the Future. 2000

DAY 1:

PLENARY SESSION

0900	Opening of the Workshop *(Alessandro Annoni)*		0900
0915	DG Information Society: GI opportunities in eEurope *(Martin Littlejohn)*		0915
0940	GI and GIS within the Commission - Eurostat activities *(Gilles Decand)*		0940
1005	EUROGI: Past, Present and Future *(Anton Wolfkamp)*		1005

PARALLEL SESSIONS

	Geodata for all: Metadata *(Chair: Daniele Rizzi)*	Geodata for all: Data *(Chair: Yves Reginster)*	
1050	Meta What!?! – An Overview of Happenings in the World of Metadata *(Tim Hancock)*	Geographic Data Infrastructure for North-Rhine Westphalia (GDI NRW) *(Heinz Bruggemann)*	1050
1110	Metadata: European Needs and User Perspectives *(Massimo Craglia)*	Spatial Reference System *(Claude Luzet)*	1110
1130	GEOCID: Implementation of a National Geographic Information Infrastructure for the Citizen *(Joana Hipolito)*	ABDS for the CEEC – An Answer to the Needs of Emerging Markets of the Region *(Katalin Toth)*	1130
1150	Metainformation System of CAGI *(Jan Ruzicka)*	Addresses and Address Data – A Key Role in Spatial Infrastructure *(Morten Lind)*	1150
	CEC: European-wide Initiatives *(Chair: Robert Peckham)*	GI Enabling Government *(Chair: Jean-François Dallemand)*	
1400	The OEEPE as a Mechanism for Problem Solving by means of Multinational Research on Common Geoinformation Problems in Europe *(Otto Kolbl)*	Integrated Regional Development Support System (IRDSS) *(Colm McColgan)*	1400
1420	PreANVIL: An Accompanying Measure for the Interoperability of Geographic Information Systems in Europe – A Technology Perspective *(Michael Gould)*	Computerisation of County Land Offices in Hungary *(Piroska Zalaba)*	1420
		GIS as Enabling Technology for an Information Environment in Public Administration *(Monika Heidemann)*	1440
1450	ETeMII: Integrating GI into the Information Society. An Accompanying Measure to Support the Setup of a European Territorial Management Information Infrastructure *(Yves Reginster)*		
	GI Education and Awareness *(Chair: Darius Bartlett)*	GI Enabling Government *(Chair: Angela Ionita)*	
1600	MUTATE: An Educational System for Multimedia and Distance Learning of GIS. *(Mauro Salvemini)*	GIS Based Dissemination of Air Quality Information in APNEE *(Gertraud Peinel)*	1600
1620	The First Steps, GIS in Education *(Henk van Dijk)*	EUROGISE *(Gavin Keith)*	1620

Proceedings of the 6th EC-GI&GIS Workshop: The Spatial Information Society - Shaping the Future. 2000

1640	Sustaining Health and Human Services: Encouraging the Creative Application of GIS *(Gregory Elmes)*	EO Data as Integrator Tool in a Town Planning Agency *(Helene de Boissezon)*	1640
1700	Leonetwork GISIG: A Training Network in GIS *(Giorgio Saio)*	Interoperating with GIS and Statistical Environment for Interactive Spatial Data Mining *(Didier Josselin)*	1700

DAY 2:

PARALLEL SESSIONS

	GI Research *(Chair: Nataliya Andrienko)*	GI Everywhere *(Chair: Françoise de Blomac)*	
0900	AGILE: A Cornerstone for the Development of Geographic and Territorial Information and Related Systems in the EU *(Mauro Salvemini)*	Possibilities of GIS Utilisation in Labour Market Analysis *(Jiri Horak)*	0900
0920	Automating the Administration Boundary Design Process using Hierarchical Spatial Reasoning (HSR) Theory and Geographic Information Systems (GIS) *(Francisco Escobar)*	The SITAI Project for the Industrial Areas of Sardinia: from a GIS Database to a Web-based Site Selection Tool *(Andrea Giacomelli)*	0920
0940	Object-Oriented Model for GIS Compressed Images *(Boris Rachev)*	GIS for Natura 2000 – Managing Europe's Nature Conservation Sites *(Stephen Peedell)*	0940
1000	Preliminary Results on Realistic Visualisation of Landscape Changes in a Small Watershed in Urdaibai Biosphere Reserve (Basque Country, Spain) *(Gonzalo Zavala)*	The Construction and Implementation of a Marine Environment Geographic Information System (GIS) *(Tesfazghi Ghebreegziabeher)*	1000
	GI Economy *(Chair: Alison Munro)*	GI Protecting the Citizen *(Chair: Mauro Salvamini)*	
1100	Efficient Pricing of Geo-Marketing Internet Services: European vs. American Approach *(Alenka Krek)*	Geographic Information Systems and Satellite Remote Sensing in Nuclear Safeguards *(Sergio Contini)*	1100
1120	Transition to Market Economy – Is it really possible in the case of GI? *(Katalin Toth)*	Agriculture – Environment – Health – Integration *(Walter Mayer)*	1120
1140	The Changing Face of the GI Economy *(Ed Parsons)*	A Web Based GIS Archive for Local Area Hazard Prevention and Mitigation *(E. Costamagna)*	1140
1200	Geo-Marketing at Your Fingertips with GISMO Internet Services *(Chris Verwoert)*		1200
	Technology: Mobile *(Chair: Yves Reginster)*	Technology: Interoperability *(Chair: Christophe de Dreuille)*	
1400	ASTRON (Applications on the Synergy of Satellite Telecommunications, Earth Observation and Navigation *(Francesco Pignatelli)*	A Networked Virtual Interoperability Laboratory for Europe - ANVIL *(John Rowley)*	1400
1420	HYPERGEO: Easy and Friendly Access to Geographic Information for Mobile Users *(David Hello)*	GIPSIE: Lessons Learned from Promoting Interoperability in Europe *(Maurits van der Vlugt)*	1420

Proceedings of the 6th EC-GI&GIS Workshop: The Spatial Information Society - Shaping the Future. 2000

1440	Value Added Services for Mobile Communications *(Henri Tobiet)*	OpenGIS and New Wave Info Dissemination *(Louis Hecht)*	1440
1500		Cataloguing and Recovering Distributed Geospatial Data: A Java Approach to Building the OpenGIS Catalog Services *(F. Javier Zarazaga-Soria)*	1500
	Technology: Data Access *(Chair: Maurits van der Vlugt)*	Technology: WWW *(Chair: Alfred de Jager)*	
1600	Spatial Knowledge Discovery: The SPIN!-System *(Michael May)*	XML-based Spatial Data Mediation infrastructure for Global Interoperability *(Ilya Zaslavsky)*	1600
1620	CommonGIS – Common Access to Geographically Referenced Data *(Natalya Andrienko)*	Web Mapping Interoperability in Practice *(Pedro Fernandez)*	1620
1640	A Prototype Spatial Object Transfer Format (SOTF) *(Peter Woodsford)*	A Spatio-Temporal Query Language for a Data Model based on XML *(Jose Eduardo Corcoles)*	1640
1700	Supporting Multiple Representations in Spatio-Temporal Databases *(Stefano Spaccapietra)*	Integration of 3D Geoscientific Visualisation Tools *(Serge Shumilov)*	1700

DAY 3:

PLENARY SESSION - THE SPATIAL INFORMATION SOCIETY

Managing the Mosaic *(Chair: Jean Claude Lummaux)*

0900	Global Mapping – The Shuttle Radar Terrain Mapper *(Walter Senus)*	0900
0925	National Geographic Information Policies in Europe *(Alessandro Annoni)*	0925
0945	Global Spatial Data Infrastructure *(Anton Wolfkamp)*	0945
	Shaping the Future *(Chair: Jean Meyer-Roux)*	
1030	The Information Society and its Impact on GI: Best Practice in Central and Eastern Europe *(Ulrich Boes)*	1030
1115	Round Table Discussion: The Spatial Information Society – Shaping the Future Participants: *Greg Elmes* *Louis Hecht* *Robert Laurini* *Jean Claude Lummaux* *Raina Pavlova* *Daniele Rizzi*	1115
1220	Closing Remarks *(Jean Meyer-Roux)*	1220

POSTERS

CREDO Flood Plain – an Example of a GI Cross-border Cooperation *(Petr Kubicek)*	Monitoring Crop Growth at National or Regional Scale with an Agro-Meteorological Model and Remote Sensing Data through a GIS Web-based Mapping Application *(Dominique Buffet)*

The SITAI Project for the Industrial Areas of Sardinia: from a GIS Database to a Web-based Site Selection Tool *(Andrea Giacomelli)*

Evaluating Tourism Potential in Mountain Areas *(S. Calvo Iglesias)*

GIS for Glaciology *(Tatiana Khromova)*

EuroFigures: A Digital Portrait of the EU's General Statistics *(Gennady Andrienko)*

Integration of GIS and Remote Sensing in Strategic Environmental Asssessment of Trans-European Networks *(Dirk Vanderstighelen)*

Metadata in Slovenia *(Tomaz Petek)*

Territorial Information System for Making Decisions in Spanish Municipalities *(Javier Albisu)*

The Construction and Implentation of a Marine Environment Geographic Information System (GIS) *(Tesfazghi Ghebreegziabeher)*

BioDiversity Information Systems: Designing WWW-based Spatial Information System *(David Pereira Jerez)*

Practising the ISO/TC 211 Approach to GI Interoperability *(Arne Berre)*

PANEL-GI: Pan European Link for Geographical Information *(Panel-GI Consortium Partners)*

Investigation on Spatio-Temporal Models and Languages *(Chiara Renso)*

Presentation of Grenoble GIS Research Activities *(Pierre Dumolard)*

Presentation of the LYON GIS Research Activities *(Robert Laurini)*

Presentation of the Activites of the Rhone-Alpes GIS research Group (SIGURA) *(Mireille Batton-Hubert)*

SVG Format Usage for GIS WWW Interface *(Lucie Friedmannova)*