Paper presented at the 3rd AGILE Conference on Geographic Information Science in Helsinki, May 25-27, 2000.

First Steps to Set Up Java Components for the OpenGIS Catalog Services and its Software Infrastructure

F.J.Zarazaga¹, R.López, J. Nogueras, O. Cantán, P. Álvarez, P.R.Muro-Medrano²

Department of Computer Science and System Engineering University of Zaragoza María de Luna 3, 50015 Zaragoza (Spain) ¹javy, ²prmuro@posta.unizar.es http://iaaa.cps.unizar.es

Abstract. The OpenGIS Consortium uses the term "Catalog" to describe the set of service interfaces supporting organization, discovery, and access of geospatial information. Catalog services help users or application software to find information that exists anywhere in a distributed computing environment. A Catalog can be thought of as a specialized database of information about geospatial resources available to a group or community of users. This paper presents an approach for the development of a GIS Catalog in compliance with "OpenGIS Catalog Services". The catalogue infrastructure has been expanded with several components to facilitate related tasks for metadata entry and automatic acquisition, keyword utilities, management and maintenance. The paper also illustrates our experience in the use of Java for software construction.

1. Introduction

The OpenGIS Consortium uses the term "Catalog" to describe the set of service interfaces which support organization, discovery, and access of geospatial information. Catalog services help users or application software to find information that exists anywhere in a distributed computing environment. A Catalog can be thought of as a specialized database of information about geospatial resources available to a group or community of users. The Catalog 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 OpenGIS Technical Reference Model (see [1]).

This work presents the development of a Catalog in compliance with "OpenGIS Catalog Services" as specified in [5] and using Java as programming language. It is part of the ISIGIS project at the IAAA R&D Group at the University of Zaragoza. The main goal of this project is to develop a multiple "layer" of tools and services above the Internet and the Web that increases the value of the information distributed across the Internet and catalyzes the creation of new information and knowledge. This information must have any kind of geospatial component such as geographical coordinates or reference location.

The decision to use Java does not appear now as essential (even in [4] the specification in Java is not presented). However, our experience working with Java and the benefits offered by this language to work in Internet made take our decision in this way. The Java programming language has gained quick and large acceptance as a prominent programming language in recent year's [2]. Its programming capabilities, multiplatform 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.

The GIS community is living an impressive increment of interest for the Java-web technologies. In this sense, the most relevant GIS and spatial data management commercial companies are devoting significant resources to develop software products with internet facilities, much of them written or with the ability to interoperate with Java (Esri, MapInfo, Oracle, Intergraph, ...), the reader can refer to [3] for product reviews and comparisons. Dozens of applets with some GIS functionality can be found in the Internet, many of them publicly available (see [2] for a review with interesting web links).

Efforts of standard organizations have been displaced to the Java and Internet worlds (the OpenGIS Consortium has prioritized Java Implementation and Web services in its latest standards).

The rest of the paper has been structured as follows. Next section presents the catalogue architecture and their main components. Section three shows some libraries and tools developed out of the scope of the OpenGIS Catalog. This paper ends with a conclusions and future work section.

2. The catalogue architecture

The implementation of the Catalog is structured in three levels: database, Catalog server and software for creating client applications (see Figure 1).

The low level consists of the database representation of the object metadata model we have developed. This model is an object representation of the FGDC Standard [6]. The use of this standard makes the interoperability labours easier. The recommendations on metadata proposed by the CEO [7] are being used in order to manage the controlled keywords. These lists of controlled keywords have been extended in order to support the Spanish political units organization using the codification of the Spanish Institute of Statistic (INE: http://www.ine.es). Oracle 8i is the relational database server used. This database engine offers interesting utilities for geospatial data management. It also provides a tool, named thesaurus, specially conceived for controlled keywords management. The database representation is completed with a set of utilities for maintaining the controlled keyword lists and adding new ones, and for labours of metadata maintaining out of the scope of the Catalog (data relocation, load balancing. etc.).



Figure 1. Three levels architecture

The Catalog Services server is built over the metadata model representation. This server has an interface compatible with the Coarse-Grain Structural Model proposed in [4]. 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 (named Discovery Service). The Catalog Manager enables management and update of the Catalog. The Discovery Service permits local information searching and distributed query. It has a component to break the queries in portions destined to a specific OpenGIS compatible catalogue, either local or remote. Once partial results

are obtained, this component is able to join them in order to compose the answer to the query of the client. To make the distributed search it is necessary to have information relative to each remote catalogue as well as its syntax of interrogation. This information is stored in a local database. The client query is written in a neutral interrogation language: the OGC Common Query Language which is a kind of SQL subset (see [4]). A component that makes a semantic translation between the original query language and the remote catalogue query language is being developed. This component needs to store the possible semantics equivalence between both languages in the "Catalog Directory" database. The design of the search has been done using the factory pattern (see [9]) where the specialization is referred to local (our catalogue) or remote search (remote catalogues). In order to complete the catalogue capacity for searching information, some utilities able to derive new knowledge from metadata stored in the database are being built. These utilities are based on artificial intelligence techniques and allow us to work with semantic relations among geographical coordinates and controlled keywords, and controlled keywords among themselves. This kind of techniques facilitates not only horizontal integration (same type of data for different areas) but also vertical integration (different types of data for the same area), which is a strong need in a distributed geolibrary as emphasized in [8]. Finally, the Access Service provides the user with means to access the items located through the Discovery Service. This service has the ability to access the local repository, or to execute a remote access service. All functionality is encapsulated into one object using the facade design pattern (see [9]) that permits to offer one single interface. This object takes responsibility for the session management of each connection to the server.

3. Other services and tools out of the OpenGIS Catalog specification

Out of the OpenGIS Consortium Catalog specification, a set of libraries and tools are under development. They have two main purposes. The first one is to facilitate the development of information systems which can operate with the catalogue. The other one is to make the Catalog management and maintenance labours easier.

Some examples of these libraries allow the creation of software for semi-automatic metadata extraction from data with geospatial information and components for transforming plain text files with metadata according to the FGDC Standard into the object metadata model and vice versa. These libraries haven used in the development of tools for catalogue management tasks. Maybe the most interesting ones of these tools are:

• *Metadata Edition* tool, as its own name indicates, enables edition and visualisation of metadata entries for local purposes in our catalogue. Figure 2 (a) shows the tab windows aspect of this tool.



Figure 2. Metadata Edition (a) and Controlled Keywords Management (b)

- *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 2 (b) shows this tool.
- *XML Import/Export* tool 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 3. Metadata generation (a) and XML Import/Export (b)

• The *Metadata generation* tool 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; 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.

There is also a library under construction to provide some classes for the development of query applications. This kind of query applications offers two ways for querying the catalogue. The first one is the 'guided' manner where the system gives to the user some different alternatives for consulting the catalogue. There is a set of predefined queries with few alternatives depending on parameter values. These queries are aimed to search information related to particular themes. For instance a component for searching information related with the Spanish political units organization is in progress. The other querying method is the 'free' way. The system supports a query language that can be used by the user to generate specific 'ad-hoc' questions. In this case, there is a library to use the OGC Common Query Language for making generic searches in the database catalogue.

4. Conclusions and future work

This work has presented the first steps to build a set of Java components for the development of an OpenGIS Catalog Services. On o the most relevant problems we have found is the un-standardised world 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 began 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 was partially supported by the Comisión Interministerial de Ciencia y Tecnología (CICYT) of Spain through the project TIC98-0587.

References

1. OpenGIS Project Document 99-112, OpenGIS Consortium 1999. The OpenGIS Specification Model. Topic 12: The OpenGIS Service Architecture (version 32).

2. A. Sorokine, I. Merzliakova. 1998. Interactive map applet for illustrative purposes". Proceedings of the 6th International Symposium on Advances in Geographic Information Systems. pp. 46—51.

3. Cambray, C. Leclerc, J.R. Houllier. 1999. Software Architectures Based on Cartographical Products. TeleGeo'99, First International Workshop on Telegeoprocessing. pp. 31--39 Lyon, France. May, 1999.

4. OpenGIS Project Document 99-051s, OpenGIS Consortium 1999. OpenGIS – Catalog Interface Implementation Specification (version 1.0).

5. OpenGIS Project Document 99-113, OpenGIS Consortium 1999. The OpenGIS Abstract Specification. Topic13: Catalog Services (version 4).

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

7. 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).

8. Panel on Distributed Geolibraries, Mapping Science Committee, Board on Earth Sciences and Resources, Commission on Geosciences, Environment, and Resources National Research Council, National Academy Press, Washington D.C. 1999. Distributed Geolibraries: Spatial Information Resources. Summary of a Workshop. http://www.nap.edu/html/geolibraries

9. E.Gamma, R.Helm, R.Johnson, J.Vlissides, 1996. Design Patterns. Elements of Reusable Object-Oriented Software. Addison-Wesley.

First Steps to Set Up Java Components for the OpenGIS Catalog Services and its Software Infrastructure

F.J.Zarazaga¹, R.López, J. Nogueras, O. Cantán, P. Álvarez, P.R.Muro-Medrano²

Department of Computer Science and System Engineering Centro Politécnico Superior, University of Zaragoza María de Luna 3, 50015 Zaragoza (Spain) ¹javy, ²prmuro@posta.unizar.es http://iaaa.cps.unizar.es

Abstract

The OpenGIS Consortium uses the term "Catalog" to describe the set of service interfaces which support organization, discovery, and access of geospatial information. Catalog services help users or application software to find information that exists anywhere in a distributed computing environment. A Catalog can be thought of as a specialized database of information about geospatial resources available to a group or community of users.

The Catalog 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 OpenGIS Technical Reference Model (see [1]).

This work presents the development of a Catalog in compliance with "OpenGIS Catalog Services" as specified in [2] and using Java as programming language. It is part of the ISIGIS project at the IAAA R&D Group at the University of Zaragoza. The main goal of this project is to develop a multiple "layer" of tools and services above the Internet and the Web that increases the value of the information distributed across the Internet and catalyzes the creation of new information and knowledge. This information must have any kind of geospatial component such as geographical coordinates or reference location.

The implementation of the Catalog is structured in three levels: database, Catalog server and software for creating client applications.

The low level consists of the database representation of the object metadata model we have developed. This model is an object representation of the FGDC Standard [3]. The use of this standard makes the interoperability labours easier. The recommendations on metadata proposed by the CEO [4] are being used in order to manage the controlled keywords. These lists of controlled keywords have been extended in order to support the Spanish political units organization using the codification of the Spanish Institute of Statistic (INE: <u>http://www.ine.es</u>). Oracle 8i is the relational database server used. This database engine offers interesting utilities for geospatial data management. It also provides a tool, named thesaurus, specially conceived for controlled keywords management. The database representation is completed with a set of utilities for maintaining the controlled keyword lists and adding new ones, and for labours of metadata maintaining out of the scope of the Catalog (data relocation, load balancing. etc.).

The Catalog Services server is built over the metadata model representation. This server has an interface compatible with the Coarse-Grain Structural Model proposed in [5]. 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 catalog (named *Discovery Service*). The Catalog Manager enables management and update of the Catalog. The Discovery Service permits local information searching and distributed query. It has a component to break the queries in portions destined to a specific OpenGIS compatible catalog, either local or remote. Once partial results are obtained, this component is able to join them in order to compose the answer to the query of the client. To make the distributed search it is necessary to have

information relative to each remote catalog as well as its syntax of interrogation. This information is stored in a local database. The client query is written in a neutral interrogation language: the OGC Common Query Language which is a kind of SQL subset (see [5]). A component that makes a semantic translation between the original query language and the remote catalog query language is being developed. This component needs to store the possible semantics equivalence between both languages in the "Catalog Directory" database. The design of the search has been done using the *factory* pattern (see [6]) where the specialization is referred to local (our catalog) or remote search (remote catalogs). In order to complete the catalog capacity for searching information, some utilities able to derive new knowledge from metadata stored in the database are being built. These utilities are based on artificial intelligence techniques and allow us to work with semantic relations among geographical coordinates and controlled keywords, and controlled keywords among themselves. This kind of techniques facilitates not only horizontal integration (same type of data for different areas) but also vertical integration (different types of data for the same area), which is a strong need in a distributed geolibrary as emphasized in [7]. Finally, the Access Service provides the user with means to access the items located through the Discovery Service. This service has the ability to access the local repository, or to execute a remote access service. All functionality is encapsulated into one object using the *facade* design pattern (see [6]) that permits to offer one single interface. This object takes responsibility for the session management of each connection to the server.

Out of the OpenGIS Consortium Catalog specification, a set of libraries enabling easier development of information systems and working with a compatible catalog is being developed. Some examples of these libraries could be: the creation of software for semi-automatic metadata extraction from data with geospatial information (an easier to understand example of this kind of software could be a component for obtaining descriptive data from an ESRI shape file usually used by ArcView); components for transforming plain text files (or even XML files) with metadata according to the FGDC Standard into the object metadata model and viceversa; or even some classes for the development of query applications. This kind of query applications provide two ways for querying the catalog. The first one is the 'guided' manner where the system offers to the user some different alternatives for consulting the catalog. There are a set of predefined queries with few alternatives depending on parameter values. These queries are aimed to search information related to particular themes. For instance a component for searching information related with the Spanish political units organization is in progress. The other querying method is the 'free' way. The system supports a query language that can be used by the user to generate specific 'ad-hoc' questions. In this case, there is a library to use the OGC Common Query Language for making generic searches in the database catalog.

Currently, the object metadata model is completed and there is an initial version of the Catalog server operative. Besides, we have a version of some client applications which allow us to catalog and recover information with geospatial component.

Acknowledgements

This work was partially supported by the *Comisión Interministerial de Ciencia y Tecnología* (CICYT) of Spain through the project TIC98-0587.

References

- "The OpenGIS Specification Model. Topic 12: The OpenGIS Service Architecture (version 32). OpenGIS Project Document 99-112. OpenGIS Consortium 1999.
- [2] "The OpenGIS Abstract Specification. Topic13: Catalog Services (version 4)". OpenGIS Project Document 99-113. OpenGIS Consortium 1999.
- [3] "Content Standard for Digital Geospatial Metadata". Metadata Ad Hoc Working Group. Document FGDC-STD-001-1998. Federal Geographic Data Committee (USA) 1998
- [4] "Recommendations on Metadata. Describing the data, services and information you have available! (version 2.0)". A User Guide provided by the Center for Earth Observation Programme (CEO programme) Of the European Commission. February 1999.

- [5] "OpenGIS Catalog Interface Implementation Specification (version 1.0)". OpenGIS Project Document 99-051s. OpenGIS Consortium 1999
- [6] "Design Patterns. Elements of Reusable Object-Oriented Software". E.Gamma, R.Helm, R.Johnson, J.Vlissides. Addison-Wesley 1996.
- [7] "Distributed Geolibraries: Spatial Information Resources. Summary of a Workshop". Panel on Distributed Geolibraries, Mapping Science Committee, Board on Earth Sciences and Resources, Commission on Geosciences, Environment, and Resources National Research Council". National Academy Press, Washington D.C., 1999. http://www.nap.edu/html/geolibraries.

Proceedings of the 3rd AGILE Conference on Geographic Information Science. 2000, p. 168-170.

AGILE Association of Geographic Information Laboratories in Europe

Hanasaari Conference Centre

Helsinki/ Espoo May 25-27 2000



3rd AGILE CONFERENCE **ON GEOGRAPHIC** INFORMATION AGILE SCIENCE

- Proceedings of the 3rd AGILE Conference on Geographic Information Science. 2000, p. 168-170.
 Toivonen T, Luoto M, Finland. Using satellite imagery and GIS methods for evaluating habitat diversity in agricultural areas ➤ 158
 Kotzinos D, Greece. Using internet-GIS to develop traveler information systems ➤ 159
 - Luoto M, Finland.
 Dimensions of the landscape structure ➤ 160

15.30-17.00 Parallel session 7, Technical interoperability Chair: Mike Gould, University "Jaume I", Castellón, Spain.

- Lehto L, Finland. XML in web-based geospatial applications ➤ 162
- Zarazaga F.J, López R, Nogueras J, Cantán O, Álvarez P, Muro-Medrano P.R, Spain.
 First steps to set up Java components for the OpenGIS catalog services and its software infrastructure ➤ 168
- Bianchin A, Pastres R, Rinaldi E, Italy. Interoperability approach in designing an integrated GIS environment for the lagoon of Venice ➤ 171
- van Oosterom P, Verbree E, Kap A, The Netherlands. Storing and manipulating simple and complex features in database management systems ➤ 178

15.30-17.00 Parallel session 8, GI databases

Chair: William Huxhold, University of Wisconsin-Milwaukee, USA.

- Martin D, United Kingdom. Towards an integrated national socioeconomic GIS: the geography of the 2001 census in England and Wales ➤ 183
- Arleth M, Denmark. The use of GI in planning documents published on the internet ➤ 186
- Bokaty D, Markov N, Ostrast P, Russia. Decision making support GIS for managing socio-economical sphere of regions ➤ 187
- Annoni A, Peedell S, Christensen S, Italy. GIS for NATURA 2000 – managing Europe's nature conservation sites ➤ 191
- Stridsman F, Segerlund M, Sweden. A construction of a digital georeferenced and searchable public database > 192

ntenno en 3P5 2611 In ple features