(Draft) Proceedings of the 11th EC-GI&GIS Workshop: ESDI: Setting the Framework. 2005, p. 91-95.

MULTI-SOURCE FRAMEWORK FOR SEAMLESSLY EXPLOITING AND LEVERAGING DISPARATE SPATIAL DATA CATALOGUES

Oscar Cantán[1], F. Javier Zarazaga-Soria[1], Javier Nogueras-Iso[2]
[1]Computer Science and Systems Engineering Department, University of Zaragoza, Spain
{ocantan, javy}@unizar.es
[2] Institute for Environment and Sustainability, Joint Research Centre, European Commission
javier.nogueras@jrc.it

ABSTRACT

In the context of a Spatial Data Infrastructure (SDI), the loss of time and resources in searching for existing spatial data or in discovering if they may be used for a particular purpose is a key obstacle to the full exploitation of the data available. A new GIS model based on interoperable geospatial services is required to combine accurate, quality data from multiple sources and offering it to broader user communities. Unfortunately, the proliferation of specifications, standards and protocols, some proprietary, hampers the interoperability goal. In this work a multi-source framework for exploiting disparate spatial data catalogues is presented. Using this technology users are offered coherent geospatial services capable of finding, accessing and combining existing data from various sources adhering to different protocols.

KEYWORDS: Interoperability, Spatial Data Infrastructures, Information Retrieval, Web Services Model

INTRODUCTION

The situation of spatial information in Europe is one of fragmentation, gaps in availability of geographical information, duplication of information collection and problems of identifying, accessing or using data that is already available. This situation makes it really difficult for users to find the quality spatial resources needed for their activities.

Additionally, in the last few years the GIS field is moving into the information technology (IT) mainstream. This implies dealing with heterogeneous systems in a seamlessly manner and, consequently, brings about new problems in the interoperability arena. It is out of discussion that the geospatial community is changing into a very broad-based community that works in many different operational environments, as shown in the information discovery continuum of Figure 1 [OGC 04-021r2, 2004]. The current tendency is moving towards the left of the illustration although there will continue to be domain specific and highly specialized GIS systems –right extremum in Figure 1- with punctual accesses to less specialized systems.

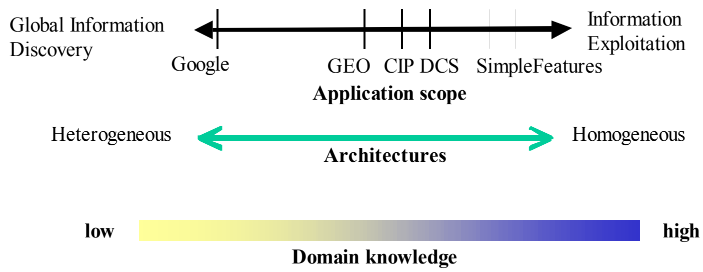

Figure 1: Information discovery continuum

Fortunately, awareness is growing at different levels about the need of delivering integrated spatial information services and infrastructures capable of exploiting information no matter its location or operational particularities. In this line of affairs, INSPIRE, an initiative prepared by the Commission to support the availability of spatial information for the formulation, implementation and evaluation of Union policies, has a prominent role to play. It intends to set the legal framework for the gradual creation of a European Spatial Data Infrastructure.

TOWARDS THE EUROPEAN SPATIAL DATA INFRASTRUCTURE (ESDI-INSPIRE)

According to the Directive of the European Parliament and of the Council establishing an infrastructure for spatial information in the Community (INSPIRE) [CEC, 2004], one of the main objectives of a Spatial Data Infrastructure (SDI) is to promote the broad dissemination and use of spatial data, not by means of collecting, as has been done until now, but by exploiting the data that is already available. Sadly enough, there is an important problem hindering this objective: incompatible information models, standards, protocols and interfaces which obstruct the interoperability demand.

As stated in the INSPIRE proposal, the interoperable spatial data and services should be achieved by all stakeholders adopting and implementing common standards and specifications. There are currently three consensus building organisations dealing with GI and GIS interoperability that have the industry's attention: the International Organization for Standardisation (ISO), the European Committee for Standardisation (CEN) and the OpenGeospatial Consortium (OGC). Nonetheless, and even though great progress has been achieved due to these initiatives, the fact is that at present there are far too many catalogue services' implementation profiles and standards available, even just inside OGC –for example Z39.50, CORBA-IIOP, CSW or SRW [OGC 04-038r2, 2005]. And that without even taking into consideration the work of private vendors promoting their own proprietary interfaces. The consequence is that spatial information is scattered under different interfaces, some proprietary, like data islands which need to be bridged.

From and organizational point of view, Spatial Data Infrastructures should be built by levels, providing interoperable services among them [Rajabifard et al., 2000]. This, precisely, is one of the most urgent and important challenges: getting all the composing systems to interoperate and share information. But as it can be seen in Figure 2, there exist so many spatial nodes, each with its own entity, attributes and responsibilities, as to pretend that a complete harmonisation of standards and protocols in the short run will be widely adopted. As Hillman and Westbrookes advanced in their book [Hillman D. et al. 2004], waiting for emerging standards to settle down is a futile exercise because it will probably not happen in our lifetimes. Actually, a more realistic step-wise approach would involve providing for translational frameworks between nodes. Otherwise some of them, especially the lower levels ones, more unresponsive to directives, will continue operating as usual with their partners and customers.

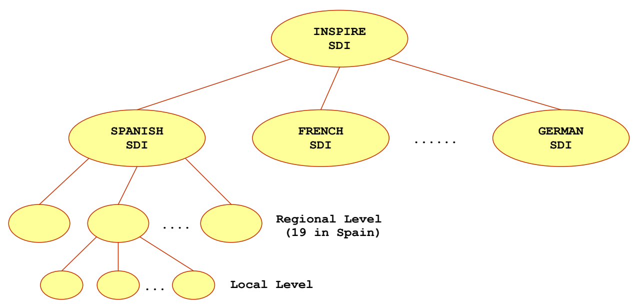

Figure 2: Hierarchical structure of SDI nodes

TOWARDS A MULTI-SOURCE, TRANSLATIONAL FRAMEWORK

Geographic Information Technology has developed as isolated, standalone, monolithic, proprietary systems as illustrated in Figure 3. Through private Application Programming Interfaces this kind of systems communicates with some sort of spatial data middleware in order to gain access to geographic data. The lower part of the SDI hierarchy of Figure 2, encompassing the local and corporate levels, rely heavily on this type of infrastructure in which interoperability is not a concern. How to walk towards the goal of interoperability from here is the subject of this work.

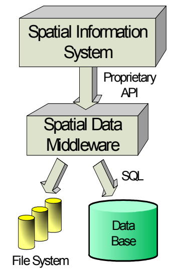

Figure 3: A typical primitive GIS

The work done in the Advanced Information System Laboratory of the University of Zaragoza constructing some Geospatial Information Systems for several organizations has shown that although these systems can be implemented using different SDI models the fact is that all of them share the same paradigm. This paradigm is described in the GSDI Cookbook [GSDI, 2004] and consists of three phases: Resource Discovery –finding resources-, Resource Evaluation –determining if what was found is what is needed- and finally Resource Exploitation –that is, making use of the underlying geographic data or services. One of our first projects in the GIS field presented a basic three-tier architecture for the aforementioned Resource Discovery, Evaluation and Exploitation phases –see Figure 4 and [Cantán et al., 2003]. In addition to the typical web client and server, the system relayed on our catalogue server, named CatServer [Tolosana et al. 2005], reachable through an RMI connector.

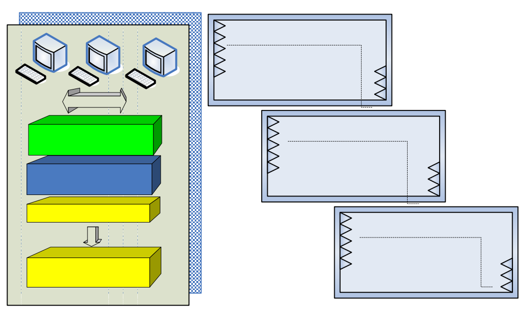

Figure 4: Basic Architecture

Nonetheless, and although a common paradigm is shared among all systems, the fact is that each application domain and organization has its own particularities, ranging from different information models to specific network services. Figure 5 shows three modifications to the basic architecture which lead to greater interoperability. On the left of the illustration the basic system is obtaining the geospatial resources from a CSW-compliant server instead of from the proprietary CatServer. On the right the resources of CatServer are accessed from SRW-compliant clients. Finally, in the middle of the illustration a source is being harvested using OAI. CSW –Catalogue Search for the Web [OGC 04-038r2, 2005], SRW –Search Retrieve Webservice [Sanderson, R., 2004] and OAI –Open Archives Initiative [Lagoce C. et al. 2004]- are well-known information retrieval protocols widely accepted in the GIS community.

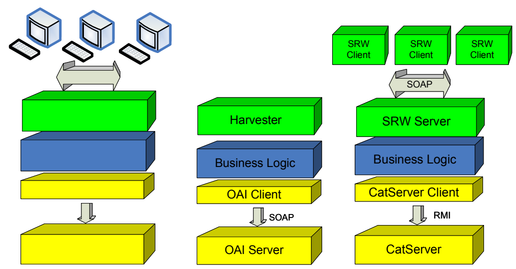

Figure 5: Some variations to the basic architecture

Problematic aspects of GIS interoperability

As new extensions to the basic architecture are demanded in order to cope with novel geographic information discovery and retrieval situations, there emerge a series of problematic elements which have to be taken into consideration in order to devise a multi-source framework. Among them the following are worth mentioning:

- Information models: Description of information resources that can be managed by GIS services. Supported query languages, core queryable and returnable properties, metadata schemas (e.g. ISO 19115:2003), data bindings and conceptual models belong to this category.
- External interfaces: Externally visible behaviour of the system, request and response message structures, operation encoding, named groups of properties.
- Network services: Protocol bindings (e.g. HTTP, Z39.50, etc), discovery services, transformation services, security, technical architectures and protocols.
- Semantics: Mapping of queryable and retrievable properties against other public metadata models, information derivation from raw data.
- Data and services sharing: Types of rights –ownership, rights of use, copyright-, types of access –retaining, sharing, trading-, types of use –discover, view, download.

MULTI-SOURCE FRAMEWORK

The problems of separately maintaining the GIS systems depicted in Figure 5 are manifold because they all share a great deal of commonalities:

- The abstract architecture is the same for them all.
- Some architectural layers are similar, so having this logic replicated is a source of engineering problems.

- There are layers differing only on the implementation details but sharing the same responsibilities (e.g. retrieving resources from a source).

More importantly, an aggregated framework encompassing different information models, protocols and interfaces would make it possible the interoperability claim as its users could access different sources using the same interfaces and, symmetrically, a source could be presented to the outer world under diverse protocols. The result of the aggregation of functionalities is the multi-source framework of Figure 6.

This multi-source framework consists of five main layers named SDI Server, Client Services, Business Model, MetadataAccessToolkit and Connectors. Its objective is to accelerate the construction of new clients and the empowerment of their capabilities as it offers interesting additional functionalities like multi-source querying, high-level services and protocol translation which are very demanded in any SDI-node. In the same vein, we pretend our clients to be written once and be able to access any data source. In the following subsections a brief description of each layer is given.

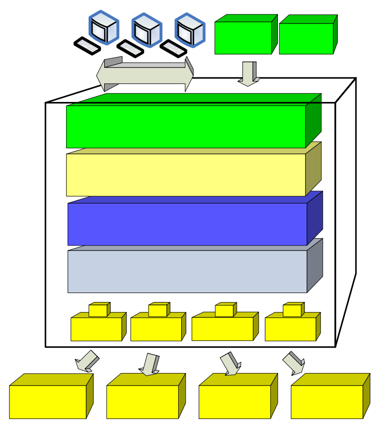

Figure 6: Multi-source framework

SDI Server Layer

The Spatial Data Infrastructure (SDI) Server Layer is composed of a series of network protocol listeners. This makes it possible for the Client Services Layer to be accessible from varied types of clients. At present the SDI Server Layer can understand requests received directly over HTTP, encoded in SOAP or issued using the Java programming language.

Client Services Layer

The Client Services Layer is responsible of providing high-level services to the SDI servers interacting with the clients. These services follow the General Catalogue Interface Model as specified in the OpenGIS® Catalogue Services Specification [OGC 04-021r2, 2004], and contain among others services for resource discovering, browsing and exploitation. Moreover this layer offers a series of high-level abstractions for representing client and search contexts and allows for concurrent access by means of an inversion of control pattern. As client requests reach the SDI server they are headed towards the corresponding service proxies–top of Figure 7-. These proxies can recover the Client Context structures held in the session so clients can be identified, their state remembered and thus be treated accordingly. Finally the actual work is delegated to the services' implementations in the Business Model Layer.

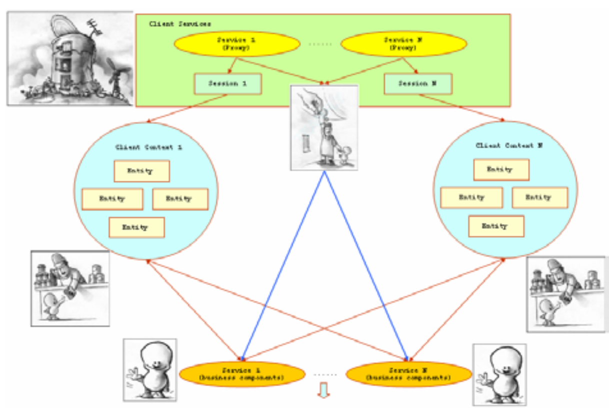

Figure 7: Operational flow of the Client Services Layer

Business Model Layer

The Business Model Layer is the real heart of the multi-source framework. Its responsibility is to receive the requests from the upper layer, apply the convenient processes, logic and information retrieval operations and finally return the results in the appropriate format.

The Business Model adopts the Geospatial Resource Access Paradigm shown before in Figure 4 and which in its simplified representation consists of three states: Query Construction and Refinement, Results Evaluation and Results Exploitation –Figure 8.

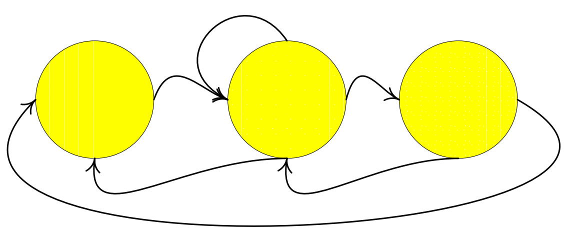

Figure 8: State diagram of the Business Model Layer

As the multi-source framework can deal with different standards and geospatial information representations, the Business Model also makes use of a crosswalk subsystem in charge of translating metadata between standards.

The Business Model also is capable of increasing the interfaces' level of abstraction by supporting high level concepts (e.g. the concept 'theme' maps to a group of lower level metadata properties) and it is responsible of results management, filtering and prioritization so users can get the best results first.

Finally the Business Model communicates with the MetadataAccessToolkit Layer in order to send the requests to the underlying metadata sources. When the responses of the underlying servers are gathered, it is this layer's duty to homogenize them for the Client Services layer's subsequent processing.

MetadataAccessToolkit and Connectors Layers

The last two layers of the multi-source framework are the MetadataAccessToolkit and the Connectors. They are presented together as the latter is closely plugged into the former.

The MetadataAccessToolkit Layer can be conceptualized as a virtual catalogue based on the OGC specifications. Thus it offers to the upper layer, that is the Business Model Layer, a series of access, management and, fundamentally, DiscoveryServices –Figure 9.

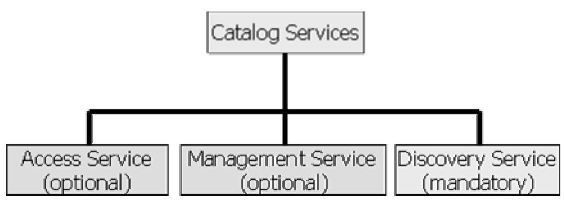

Figure 9: Abstract Catalogue Services of the MetadataAccessToolkit

These services are offered by means of an object oriented interface, something a bit similar to ODBC (Open DataBase Connectivity) or JDBC of Java. It also offers added value services like results caching, sorting, filtering, merging and session simulation. The most important consideration to stress is that the MetadataAccessToolkit Layer makes all the particular underlying sources look similar for the Business Model.

Finally the MetadataAccessToolkit delegates the work of accessing the spatial information sources to the concrete pluggable cartridges –Connectors Layer. This layer consists of a cluster of protocol specific pluggable cartridges implementing the MetadataAccessToolkit abstract interfaces. Worth mentioning are those for RMI, CSW, Z39.50 and SRW but of course as new needs emerge new connector can be added without disrupting the upper layers.

CONCLUSIONS

In the European Union massive amounts of geospatial information have been collected in recent years, particularly at regional and local levels. However this information is fragmented and hard to be accessed by broad user communities. INSPIRE recognizes these problems and proposes creating a Spatial Data Infrastructure (SDI) to deal with them. This SDI will allow users to easily find, access and combine existing data form various sources. In order to achieve this objective, incompatible geographic information systems have to be put working together. This, in its turn, implies interoperability between protocols, interfaces and information models. In this work we have presented an expandable multi-source framework developed at the University of Zaragoza which enables seamlessly access to different geospatial data catalogues, even under different protocols.

ACKNOWLEDGEMENTS

This work has been partially supported by the Spanish Ministry of Education and Science through the project TIC2003-09365-C02-01 from the National Plan for Scientific Research, Development and Technology Innovation.

BIBLIOGRAPHY

Cantán O., Nogueras-Iso J., Torres M.P., Zarazaga-Soria F.J., Lacasta J. (2003). On the Problem of Finding the Geographic Data We Are Looking For, 9th European Commission-GI&GIS Workshop ESDI: Serving the User A Coruña, Spain.

Commission Of The European Communities (CEC), (2004). Proposal for a Directive Of The European Parliament And Of The Council establishing an infrastructure for spatial information in the Community (INSPIRE) http://www.eurogeographics.org/eng/documents/COM_2004_0516_F_EN_ACTE.pdf

Global Spatial Data Infrastructure Association (GSDI), (2004). Developing Spatial Data Infrastructures: The SDI Cookbook, Editor: Douglas D. Nebert, http://www.gsdi.org/docs2004/Cookbook/cookbookV2.0.pdf

Hillman D., Westbrooks E. (2004). Metatada in Practice. American Library Association.

INSPIRE WP-PP-v4.5.3 (2004). Work Programme Preparatory Phase 2005-2006. Infrastructure for Spatial Information in Europe.

Lagoce C., Van de Sompel, H. (2004). The Open Archives Initiative Protocol for Metadata Harvesting. http://www.openarchives.org/OAI/openarchivesprotocol.html

OGC 04-021r2 (2004). OpenGIS® Catalogue Services Specification 2.0. Open Geospatial Consortium Inc. http://portal.opengeospatial.org/files/?artifact_id=5929&version=1

OGC 04-038r2 (2005). OpenGIS® Catalogue Services Specification 2.0 - ISO19115/ISO19119 Application Profile for CSW 2.0. Open Geospatial Consortium Inc. https://portal.opengeospatial.org/files/?artifact_id=6495

Rajabifard, A., Williamson, I.P., Holland, P., Johnstone, G., 2000. From Local to Global SDI initiatives: a pyramid of building blocks. In: Proceedings of the 4th GSDI Conference, Cape Town, South Africa.

Sanderson, R. (2004). SRW: Search/Retrieve Webservice, Version 1.1. SRW Editorial Board. http://srw.cheshire3.org/SRW-1.1.pdf.

Tolosana-Calasanz R., Portolés-Rodríguez D., Nogueras-Iso J., Muro-Medrano P.R., Zarazaga-Soria F.J., 2005: CatServer: A Server of GATOS. 8th Agile Conference on GI, Estoril, Portugal.

Multi-Source Framework for Seamlessly Exploiting and Leveraging Disparate Spatial Data Catalogues

Oscar Cantán [1], F. Javier Zarazaga-Soria[1], Javier Nogueras-Iso[2]

[1]*Computer Science and Systems Engineering Department*
University of Zaragoza, Zaragoza, Spain
[2]*Joint Research Centre - European Commission*

According to the Directive of the European Parliament and of the Council establishing an infrastructure for spatial information in the Community (INSPIRE) [CEC, 2004], one of the main objectives of a Spatial Data Infrastructure (SDI) is to promote the broad dissemination and use of spatial data, not by means of collecting, as has been done until now, but by exploiting the data that is already available. Sadly enough, there is an important problem hindering this objective: incompatible standards, protocols and interfaces.

Since 1994 the Open Geospatial Consortium (OGC) has been working in the promotion and adoption of open standards and interfaces in the GIS field. Though great progress has been achieved due to these initiatives, the fact is that at present there are far too many catalogue services' implementation profiles and standards available, even inside OGC –for example Z39.50, CORBA-IIOP, CSW or SRW [OGC, 2004]. And that without even taking into consideration the work of private vendors promoting their own proprietary interfaces. Even the INSPIRE initiative, aimed at setting the legal framework for the gradual creation of a spatial information infrastructure, recognises the fact that most of the quality spatial information is available at local and regional level and that this information is difficult to exploit in a broader context for a variety of reasons. Interoperability is one of the most relevant among these causes.

According with the model proposed in [Rajabifard et al., 2000], SDIs should be built by levels, providing interoperability services among them. As it can be seen in Figure 1, there are so many actors and existing metadata systems as to pretend that a complete harmonisation of standards and protocols in the short run will be widely adopted. Actually, a more realistic step-wise approach involves providing for translational frameworks in order to make metadata accessible as soon as possible. Otherwise the local level may find resistance to adopt new models and continue instead operating as usual with their partners and customers.

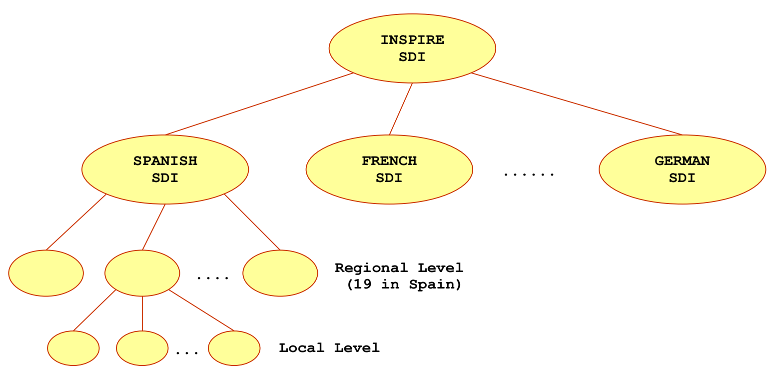

Figure 1: Hierarchical structure of SDI nodes

One of our first projects in the SDI field presented a basic three-tier architecture for the resource discovery and evaluation subsystem, as can be seen in Figure 2 [Cantán et al., 2003]. In addition to the typical web client and server, the system relayed on our catalogue server, named CatServer [Tolosana et al. 2005], reachable through a RMI connector.

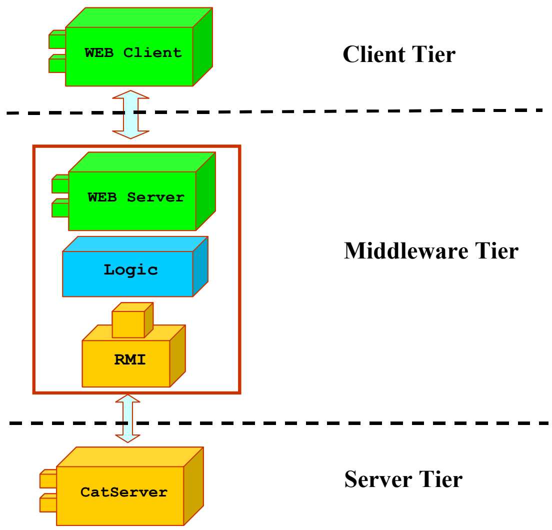

Figure 2: Basic architecture

Nonetheless for each application domain and organization the basic architecture of Figure 2 had to be replicated. New clients implied new ways of interaction, logic, connectors and servers, Figure 3.

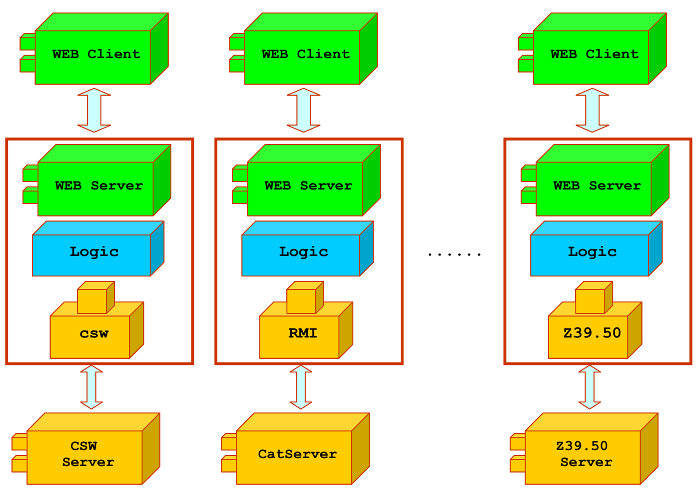

Figure 3: Basic architecture replication as new clients are added

The problems with the architecture depicted in Figure 3 are manifold: it is difficult to maintain, it is effort and time demanding to support new clients and it isolates rather than integrates the functionalities. Thankfully enough, in spite of each client's particularities there is an important deal of functionality that can be factored out. This gives rise to the architecture of Figure 4.

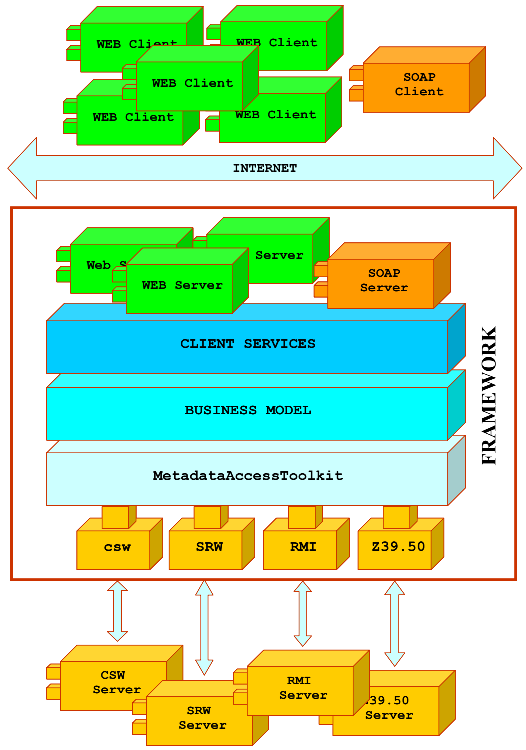

Figure 4: Search Framework

If by middleware in Figure 2 we meant any programming glue that serves to mediate between two separate entities, in Figure 4 the term framework is used to refer to a defined support structure in which another project can be organized and developed. This framework consists of four main layers named Client Services, Business Model, MetadataAccessToolkit and Connectors. Its objective is to accelerate the construction of new clients and the empowerment of their capabilities as it offers interesting additional functionalities like multi-source querying, high-level services and protocol translation which are very demanded in any SDI-node. In the same vein, we pretend our clients to be written once and be able to access any data source. In the following paragraphs a brief description of each layer is given.

The Client Services Layer is responsible of providing high-level services to the application servers interacting with the clients. These services are derived from the Geospatial Resource Access Paradigm, as described in the GDSI Cookbook [GSDI, 2004], and contain among others those for discovering resources, browsing results and recovering specific information. Moreover this layer offers a series of high-level abstractions for representing client and search contexts and allows for concurrent access by means of an inversion of control pattern.

The Business Model Layer recreates the GSDI Cookbook's paradigm, based on the resource discovery, evaluation and access phases. This layer translates the requests from above into invocations to the lower layer. It also provides for translation among different data representations and query languages. When the responses of the underlying servers are gathered, it is this layer's duty to homogenize them for the Client Services layer's subsequent processing.

The MetadataAccessToolkit layer (MAT for short) acts as a virtual catalogue. It makes accessible an object oriented interface, much alike JDBC in Java, to let the Business Model access uniformly any supported metadata source. Thus, no matter the kind of client sending the requests, the intermediate language is always the same in this layer. In addition to this, the MAT layer provides some capabilities like results caching, sorting, content-specific restrictions and session simulation when appropriate.

Finally, the Connectors layer makes the connection with diverse resource servers possible. As can be seen in Figure 4, this layer consists of a cluster of protocol specific pluggable cartridges. Each one of them implements the MAT's source and protocol specific interfaces. The sources to connect to have to be indicated to the framework in an XML file in order for it to select the adequate connectors at launch time. Adding support for new sources calls for convenient cartridges, the rest of the framework remains the same so existing clients can at once access the new sources using the already provided functionality.

Summing it up, the framework presented in this article not only accelerates the construction of SDIs by factoring out common logic, but it also adds new and interesting capabilities both for clients and developers. One worth mentioning is that which enables the exploitation of underlying sources using any high level protocol. See Figure 5.

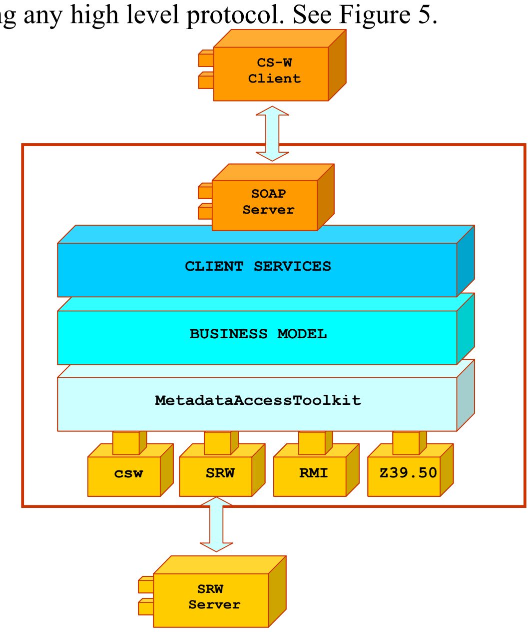

Figure 5: Connection of a CS-W client to a SRW server

ACKNOWLEDGMENTS

This work has been partially supported by the Spanish Ministry of Education and Science through the project TIC2003-09365-C02-01 from the National Plan for Scientific Research, Development and Technology Innovation.

BIBLIOGRAPHY

Cantán O., Nogueras-Iso J., Torres M.P., Zarazaga-Soria F.J., Lacasta J.: On the Problem of Finding the Geographic Data We Are Looking For, In 9th European Commission-GI&GIS Workshop ESDI: Serving the User A Coruña, Spain

Commission Of The European Communities (CEC), 2004: Proposal for a Directive Of The European Parliament And Of The Council establishing an infrastructure for spatial information in the Community (INSPIRE) http://www.eurogeographics.org/eng/documents/COM_2004_0516_F_EN_ACTE.pdf

Global Spatial Data Infrastructure Association (GSDI), 2004: Developing Spatial Data Infrastructures: The SDI Cookbook, Editor: Douglas D. Nebert, http://www.gsdi.org/docs2004/Cookbook/cookbookV2.0.pdf

Open Geospatial Consortium Inc., 2004: OpenGIS® Catalogue Services Specification 2.0 - ISO19115/ISO19119 Application Profile for CSW 2.0,

https://portal.opengeospatial.org/files/?artifact_id=6495

Rajabifard, A., Williamson, I.P., Holland, P., Johnstone, G., 2000: From Local to Global SDI initiatives: a pyramid of building blocks. In: Proceedings of the 4th GSDI Conference, Cape Town, South Africa.

Tolosana-Calasanz R., Portolés-Rodríguez D., Nogueras-Iso J., Muro-Medrano P.R., Zarazaga-Soria F.J., 2005: CatServer: A Server of GATOS. In 8th Agile Conference on GI, Estoril, Portugal.

(Draft) Proceedings of the 11th EC-GI&GIS Workshop: ESDI: Setting the Framework. 2005, p. 91-95.

ies
Institute for Environment and Sustainability

EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

Abstracts

ESDI: Setting the Framework

11th EC-GI&GIS Workshop

Alghero, Sardinia
29 June-1 July 2005

www.ec-gis.org/Workshops/11ecgis

INSPIRE · Infrastructure for Spatial Information in Europe

REGIONE AUTONOMA DELLA SARDEGNA

Table of Contents

(Draft) Proceedings of the 11th EC-GI&GIS Workshop: ESDI: Setting the Framework. 2005, p. 91-95.