Global Geoportal for Remote Sensing Images Centers

Vanessa Cristina Oliveira de Souza, Gilberto Câmara

National Institute For Space Research
Av dos Astronautas, 1.758 - Jd. Granja - CEP: 12227-010
São José dos Campos – SP - Brazil

vanessa@dpi.inpe.br, gilberto.camara@inpe.br

Abstract. *The objective of this work is to propose a geoportal capable of integrating the catalogues of the different images centers without these centers having to migrate their current computational architectures to a specific architecture. In this work, the mediator architecture is considered a flexible and efficient way of building a global geoportal for remote sensing images. The three components of this architecture, portal, mediator and data sources, will be analyzed and specified. The web services technology, as a solution for implementing this architecture, will also be analyzed. As a demonstration prototype, we have integrated the CBERS data distributed by INPE in an existing geoportal.*

1. Introduction

Remote sensing images are used in various fields with various objectives. During the last decades of the 20th century, the repositories of remote sensing images, the images centers, organized their data in offline tape archives and each image was individually generated for the user. At the end of the 1990's the centers changed the way their images were distributed. With the popularization of the Internet, these repositories began to convert their data to online access.

Each center created its own images catalogue using its own interfaces, software and hardware platforms. It became difficult for the user to interact with different interfaces and to find the catalogues online since the current search engines were not designed to find geographic information. Therefore, the user must find the catalogue, learn how to use it and later combine the data manually.

In this context, a geoportal where the various images centers around the world could be accessed, making the particularities of each one clear to the user, is desirable.

In this work we will analyze the problem of organizing *online* distribution so that the centers can work in cooperation, increasing the amount of information available to the user. We will try to answer the question: How can we conceive and build a geoportal with search and recovery services for remote sensing images centers that function in cooperation?

We assume that a mediated architecture [1] is a flexible and efficient way of building the geoportal for the images centers. We will study the concept and building of mediators based on the *web services* technology [2-5] as a solution to the problem.

After studying some existing online images catalogues [6] and analyzing the necessary components of this mediated architecture, we propose a solution based on the current *web services* technologies. We also carried out an experiment that integrated the data from the CBERS[1] images catalogue, distributed by INPE[2], in an existing geoportal (eoPortal) to validate the use of the mediated architecture and to analyze the positive aspects and the deficiencies of this geoportal.

2. Theoretical References

A geo-spatial portal (geoportal) is the human interface for a collection of *online* geo-spatial information, including services and datasets. Technically speaking, geoportals are *sites* connected to *web* servers that provide metadata on their geographic data or services [7, 8]. For Tait [9], geoportals provide applications that can search, map, publish and administrate geographic information.

Data integration is the problem of combining data from different sources, offering the user a unified vision of them (global schema), and defining a set of queries that determine how each element of the global schema is obtained in function of the data stored in the local data sources [10, 11]. A schema for the mediated is an unified vision available from the mediator. Queries can be made against that schema and submitted to the mediator. So, they are decomposed at run time into queries on the local data sources. The results from these queries on the local data sources are translated, filtered and merged, and then the final answer is returned either to the user or to the application[12].

The data integration systems follow two main approaches: the materialized approach and the virtual approach. The virtual approach is the most applied to our problem, as the data remain in the sources and the information is extracted directly from them when a consult is requested. The main advantages of the virtual approach are the non-replication of the data and the fact that the information recovered is always updated. The disadvantages of this approach include a possible inaccessibility of the sources and the long response time [12-14].

The virtual approach is generally modeled using the mediated architecture [1]. In this architecture there is a *software* module, the mediator, that receives and treats the searches submitted to the integration system, decomposing these queries into sub-queries that will be submitted to the data sources [15]. The *wrapper* is a program used to make the translation between the sources' data model and the data model defined by the mediator (Figure 1).

The Web Services (WS) are currently the most promising way to integrate applications on the Internet [3]. A web service is a software application identified by a URI, whose interfaces and binding are capable of being defined, described and discovered by XML artifacts and supports direct interactions with other software applications using XML based messages via internet-based protocols [16]. The virtual approach modeled by mediated architecture is generally implemented by web services.

[1] China-Brazil Earth Resources Satellite - http://www.cbers.inpe.br/
[2] Brazilian National Institute For Space Research

In W3C WS, XML is used to encode data. The protocol of communication is SOAP (Simple Object Access Protocol). The services' description is standardized by the WSDL (Web Services Description Language) and the services' discovery by UDDI (Universal Description, Discovery and Integration).

Compared with other protocols, the SOAP has a low performance, considering that the messages are described in text (XML), while in RPC (Remote Procedure Call) systems, they are exchanged in binary format. Govindaraju et al. [17] say that the SOAP message is four to ten times larger than the binary representation the same.

The eoPortal[3] is the main entrance of an architecture made available by ESA[4] called Service Support Environment (SSE), in which the users and the service providers can interact automatically and dynamically. In this portal, the providers register and offer their services and data and the users access them [18].

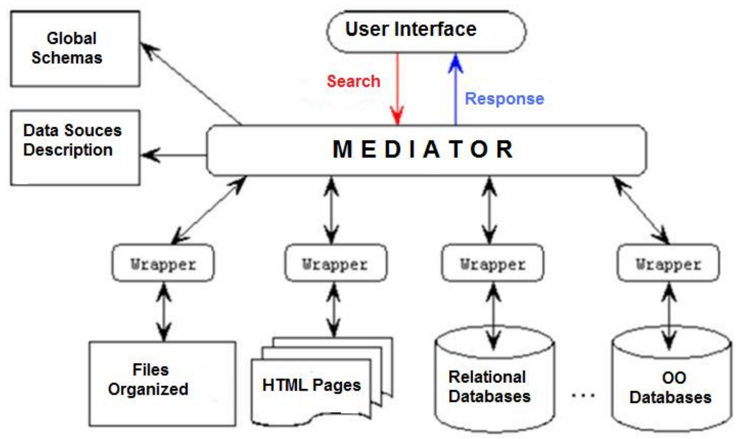

Figure 1 : Virtual approach modelled by mediators.

The SSE is a system open to remote integration of services with domains focused on observation of the Earth and on the geographic information systems. The architecture of the SSE can be seen in Figure 2. In the system, services can be integrated without having to re-compile or re-design it [19]. Its architecture is constituted by a set of functionalities, components and gateways, providing a distributed environment oriented towards services. This architecture is non-proprietorial, open, scaleable and robust for discovering and distributing services.

In the area aimed at service providers, presented in Figure 2, it is the Toolbox, a development environment provided by SSE to help the providers to create their services.

3. Global Geoportal for Remote Sensing Images Centers

We will begin this geoportal proposal with its architecture. We propose a mediated integration architecture based on the virtual approach and implemented by web services. Therefore, the mediator component in Figure 1 is a web service, as are the wrappers. The sources are described by WSDL and UDDI. The integrated vision is a schema

[3] http://eoportal.gov/
[4] European Spacial Agency

XML that defines the content of the SOAP messages exchanged between the mediator and the wrappers (Figure 3).

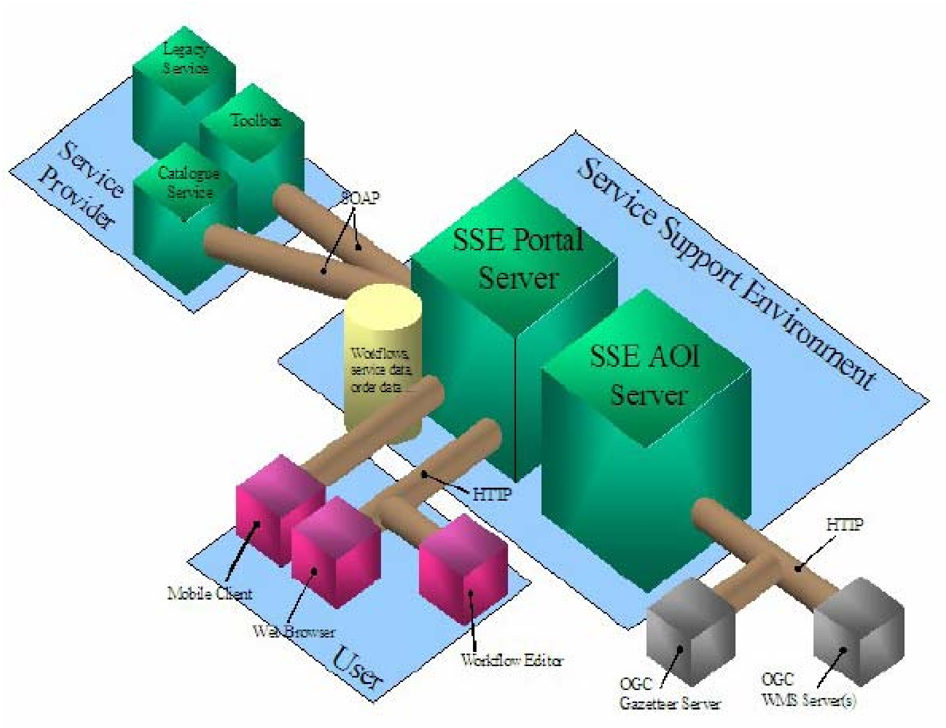

Figura 2 : SSE architecture.

The architecture is composed of: portal, mediator and data sources. The mediator component of this architecture is detailed in Figure 4. As the SOAP, the virtual approach has performance problems. Therefore, the purpose of the mediator filter is to optimize the transmission of the messages, sending them only to the sources that can actually respond to the consult made. This filtering is done using the WSDL and UDDI. Ideally, the sources would register the image sensors, the area of coverage of their scenes and the period of time in the UDDI. After filtering the sources, the messages are sent to them and the results are integrated into one result, which is sent back to the portal.

The sources have their own data models and to transform them into the integrated vision, and to standardize the SOAP messages exchanged between the mediator and the sources, a web service wrapper must be implemented. Basically, the operations of this service follow Figure 5. After receiving an xml message with the search parameters, the service must decode this xml to remove the parameters, consult the database and write a new xml message, which will return the result of the consult to the geoportal.

All the interaction with the user occurs in the portal component of Figure 3. It is a web interface that functions as a client of the mediator component. This portal contains:

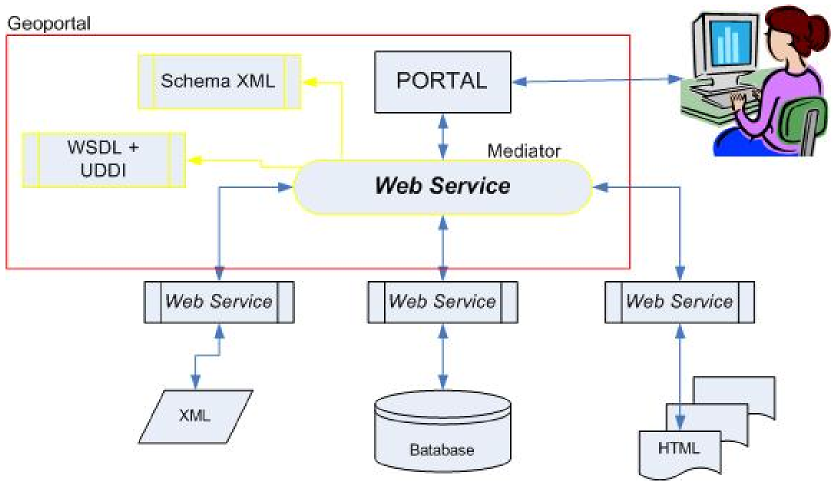

Figure 3: Architecture proposed for the global geoportal implemented by web services.

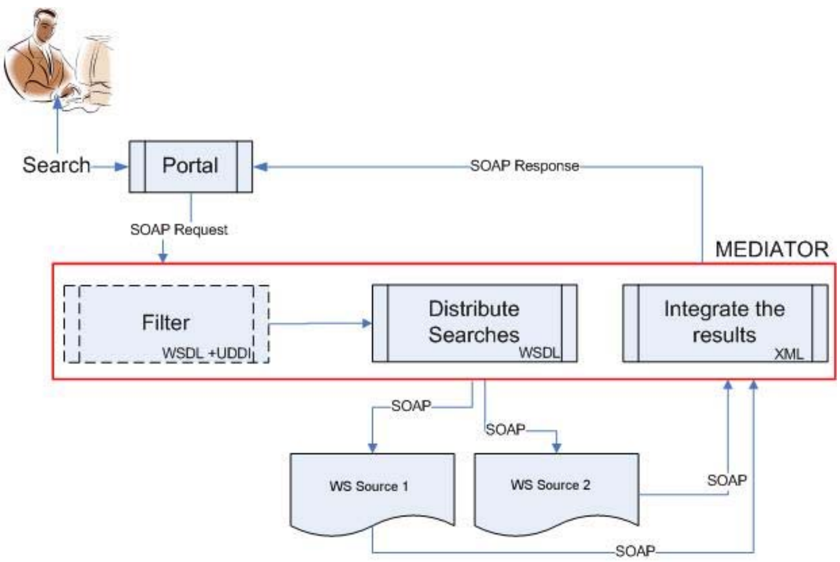

Figure 4 : Mediator component.

a) Images catalogue:

Catalogues publish metadata and provide mechanisms to consult and recover information from distributed repositories. This component is a web interface which composes queries and presents the information from the repositories to the user, via a graphic interface. We can, however, associate a data access interface to the catalogue, from where the user could obtain the actual scene.

The metadata presented in this work are the result of a study of major catalogues of images available today [6]. The ISO 19115 [20] was also used in this work. In the tables below, the metadata were divided into mandatory, optional and additional fields. This division was made to evaluate the catalogues. At the interfaces of this proposed catalogue, all metadata should be presented.

The search interface must contain the maximum number of methods for locating the area of interest, such as gazetter, geo-coding, editing map, file upload and coordinate entry. The main role of this interface is to help the user find the data he is searching for in the best possible way. We suggest that all the parameters in Table 1 should be included in this interface. These fields will be the parameters used by the data sources during a consult.

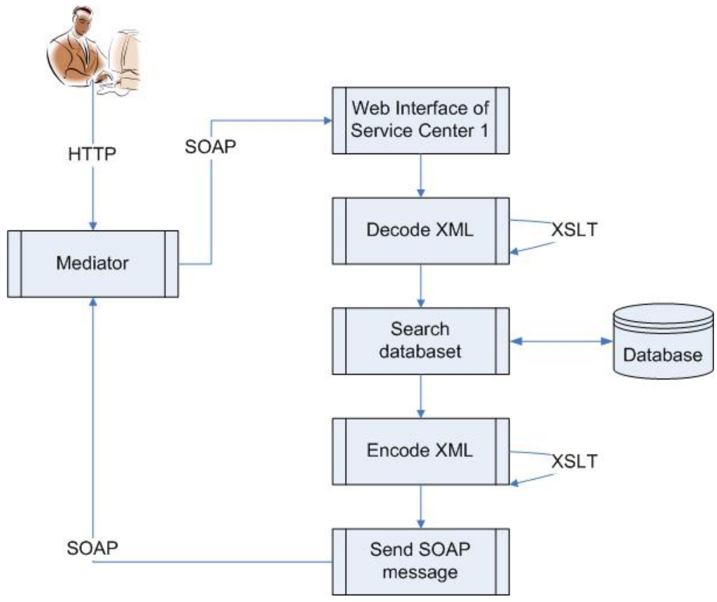

Figure 5 : Images centers' services.

Apart from these parameters, common in existing images catalogues, we also propose an interface that uses parameters that are less punctual and more semantic. Therefore, instead of searching for the sensors by their names, the user would search for them according to spatial, temporal, radiometric and spectral resolution, as well as according to a radar's polarimetry. This form of consult does not require knowledge of the sensor system, allowing the user to search for any sensor that best meets his needs, making the interface more flexible and intelligent.

The interface for publishing the metadata will present to the users the metadata returned by the consult. We classify these metadata in Table 2. The fields referenced with * were taken from ISO 19115 [20].

Table 1: Mandatory, optional and additional fields of a geoportal's search interface for RS images centers.

Mandatory Fields	Optional Fields	Additional Fields
Geographic Location	Sensors	Illumination and Azimuth angles
	Initial and final dates	Quality of the image
	Cloud coverage	Source of the data
	Online availability of the product	Path/row
		SceneId

Apart from these parameters, common in existing images catalogues, we also propose an interface that uses parameters that are less punctual and more semantic. Therefore, instead of searching for the sensors by their names, the user would search for them according to spatial, temporal, radiometric and spectral resolution, as well as according to a radar's polarimetry. This form of consult does not require knowledge of the sensor system, allowing the user to search for any sensor that best meets his needs, making the interface more flexible and intelligent.

The interface for publishing the metadata will present to the users the metadata returned by the consult. We classify these metadata as in Table 2. The fields referenced with * were taken from ISO 19115 [20].

Table 2 : Mandatory, optional and additional fields of a geoportal's interface for presenting results for RS images centers.

Mandatory Metadata	Optional Metadata	Additional Metadata
Scene Identifier	Format of the image	Condition of the image*
Geographic coordinates	Cost	Quality of the image*
Path/Row	Source of the data	Level of processing*
Day and time of the acquisition	Órbit	Radiometric calibration data *
Sensor system	Nadir	Camera calibration data *
Cartographic Projection	Direction of the orbit	Distortion data*
Form of access to the data	Azimute *	Triangulation*
Cloud coverage	Elevation*	

The interface proposed presents all the metadata in Table 2, divided in categories. All the metadata have a key that explicits their function and possible values, assisting less experienced users. All the scenes must be drawn over a reference map. The mandatory metadata must be presented for all the scenes resulting from the consult. The others must be presented only if requested by the user. The order in which the scenes appear must be decided by the user. By default, the more recent scenes appear first.

Coupled with the interface for publishing the metadata there must be an interface that allows the user to request the desired scene. This interface should send the requests to the provider centers and deliver the products to the user of the geoportal. The scenes may be downloaded or delivered to the user's e-mail.

b) Administrative interface:

The administrative interface will manage the registration of the users and of the images centers. The registration of the users will enable them to make full use of the geoportal's functions, as well as composing the portal's statistics. The registration of the centers is also important since they then gain the status of service providers and can publicize their data.

c) Extra functions:

Many functions can be added to the geoportal to attract and help users. We suggest the list below. Only the last can be used by users not registered in the geoportal:

- **Storing search parameters and consult results:** this function is present in the Earth Explorer5 and its objective is to save time. Saving the search criteria is interesting for users who always make the same consult, changing only the date, for example.
- **Forum:** as it is a niche environment, it is common for the portals to have forums so that theirs users can communicate with each other, exchange information and clarify doubts.
- **News:** the objective of this function is to keep the users updated on the news of the geoportal itself or other news in the field of RS images. In this way, the user can be informed, via e-mail, of a new content or a new data source in the portal.
- **Compare sensors:** this function helps the user to choose one or another sensor based on the comparison of some characteristics, such as, for example, the sensors' spatial resolution and the cost of the product.

4. Integration of the CBERS catalogue in the eoPortal

4.1. Current architecture of the CBERS catalogue

The cooperation between the Brazilian and Chinese governments established in 1988 led to the development of imaging satellites, which help monitor the natural resources of both countries. In Brazil, the CBERS program includes a policy of free distribution of the images, first inside the national territory and, more recently, to other South-American countries. This distribution is done through a catalogue available on the Internet.

The current catalogue is built on a structure based on PHP scripts and mysql relational database. The catalogue's search interface does not have an editing map, a geo-coding system and file upload. Although the catalogue provides data from many sensor systems, the searches are not multi-sensor. The interface for publishing the metadata lacks a reference map and the obligatory metadata available is minimal.

In the data access interface all communication with the user is done through e-mails. When the scene requested is ready, it is copied to an area on the Internet. The area's address is sent to the user, who has five days to complete the download.

4.2. Data model for the unified vision

The Interface Control Document (ICD) defines the external interfaces of the SSE, particularly the interfaces between the SSE Portal and the remote services of the providers. For this work, the most relevant interfaces are ESRIN EOLI [21] and MASS. EOLI defines the messages that are exchanged between the SSE and the catalogue image services that implement the "Search" and "Present" operations. MASS defines many operations, among them the operation for carrying out orders through the portal, "Order".

Communication between the portal and the catalogue service is established using SOAP on HTTP. The SSE provides pre-defined workflows that make the

[5] http://edcsns17.cr.usgs.gov/EarthExplorer/

conversion between the XML messages of the SSE Portal and the EOLI interface messages [19, 21].

4.3. Integration

The CBERS catalogue was integrated in the eoPortal through the implementation of a web service, whose operations are: Search, Present and Order. This service will perform the function of wrapper between the CBERS relational data model and the interfaces EOLI and MASS (Figure 6).

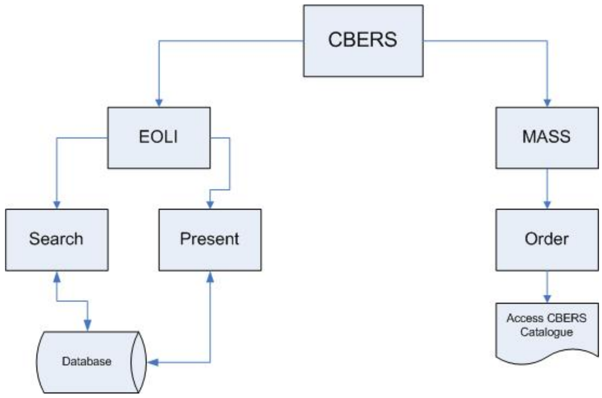

Figure 6: Implemented CBERS service.

We used the Toolbox to implement the service, which runs on an Apache Tomcat server. The operations implemented are all synchronous and are detailed below.

The scenes' metadata recovered from the database are translated to XML, following EOLI, by a second XSL script. This script is the actual wrapper, as it literally translates the CBERS catalogue schema into the EOLI data model. The XML generated is encapsulated in a SOAP message by Tomcat and sent back to SSE, which applies a style to it, codifies the scenes coordinates in GML[6] and presents both the textual and the graphic data in the eoPortal.

The Present operation functions in the same way as the Search. However, in the Present operation, the search parameter will always be the identifier of one single scene and the response requested to the database contains a larger number of metadata.

The operation Order follows the MASS interface data model. It receives as entrance parameter the identifiers of the desired scenes, the user's name and e-mail, and returns the scene, through a link, to the file stored in disk. This operation has been implemented, but not yet operationally. The service (Figure 7) can be accessed in: http://services-test.eoportal.org/portal/service/ShowServiceInfo.do?serviceId=DF80CA80.

At the moment, the service is in the test phase. The liberation for all community will be done in December.

[6] Geography Markup Language

5. Conclusions

This work discussed the potential benefits of integrating the different remote sensing images centers. We indicated how this integration can be achieved using a mediated architecture and we made a demonstration prototype.

Analyzing the current major catalogues, we observed how the integration was necessary. As the catalogues already exist in their software and hardware platforms, it is important to have a mediated architecture that unites them in one simplified interface. This work has proved that it is possible to build a mediated architecture that functions well, since the aspects that are common to all the catalogues are stronger than the differences between them. The integration between the eoPortal and the CBERS catalogue is an example of this.

The SSE is an open, free system based on free technologies that provides a high quality support. Providing services using the system does not imply any costs. In the case of images catalogues, the great advantage is the EOLI interface.

The greatest difficulty to integration with eoPortal lies in the quantity of languages that must be mastered (TSCRIPT, XML, XSL, XSLT, Xpath, WSDL, UDDI) and in translating local metadata into EOLI. There are no notation or semantic difficulties, but there is a difficulty in storing the information in the same standard as EOLI.

For the users, the implementation of this geoportal would facilitate the search for and recovery of the images and would increase the volume of data available. For the centers, this geoportal would decrease the effort of implementing and maintaining geographic web interfaces, as the providers would need only to keep the database updated and implement the wrapper service. It is essential to use the XML, and all the technology associated to it, for this architecture to function with total interaction between the parts.

Figure 7 : INPE Catalogue Service in eoPortal.

6. References

1. Wiederhold, G., *Mediators in the Architecture of Future Information Systems* IEEE Computer, 1992. **25** (3): p. 38-49.

2. Breitman, K., M. Casanova, and W. Truszkowski, *Semantic Web – Concepts, Techonologies and Applications*. 2007, London: Springer. 327.

3. Curbera , F., et al. (2002) *Unraveling the Web Services Web - An Introduction to SOAP, WSDL, and UDDI*. IEEE Internet Computing **Volume**, 8

4. Newcomer, E., *Understanding Web Services - XML, WSDL, SOAP and UDDI*. 2002, Boston: Addison-Wesley. 332.

5. W3C, W.W.W.C. *Web Services Architecture*. 2004 [cited 2007 Janeiro/2007]; Available from: http://www.w3.org/TR/ws-arch/

6. Souza, V.C.O., *Geoportal Global para Centros de Imagens de Sensoriamento Remoto*, in *Sensoriamento Remoto*. 2008, INPE: São José dos Campos/SP. p. 99.

7. Maguire, D.J. and P.A. Longley, *The emergence of geoportals and their role in spatial data infrastructures*. Computers, Environment and Urban Systems, 2005. **29**(1): p. 3-14.

8. OGC, O.G.C. *Geospatial Portal Reference Architecture*. 2004 [cited 2007 Janeiro/2007]; Available from: http://portal.opengeospatial.org/files/?artifact_id=6669.

9. Tait, M.G., *Implementing geoportals: applications of distributed GIS.* Computers, Environment and Urban Systems, 2005. **29**(1): p. 33-47.

10. Lenzerini, M., *Data integration: a theoretical perspective*, in *Proceedings of the twenty-first ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*. 2002, ACM: Madison, Wisconsin.

11. Lóscio, B.F., A.C. Salgado, and V.M.P. Vidal. *Using Agents for Generation and Maintenance of Mediators in a Data Integration System on the Web*. in *XVI Simpósio Brasileiro de Banco de Dados*. 2001. Rio de Janeiro - Brazil.

12. Lóscio, B.F., *Managing the Evolution of XML-based Mediation Queries*, in *Doutorado em Ciências da Computação*. 2003, Universidade Federal de Pernambuco: Recife. p. 159.

13. Barbosa, A.C.P., *Middleware para Integração de Dados Heterogêneos Baseado em Composição de Frameworks*, in *Doutorado em Informática*. 2001, PUC-RIO: Rio de Janeiro. p. 167.

14. Batista, M.d.C.M., *Otimização de Acesso em um Sistema de Integração de Dados através do Uso de Caching e Materialização de Dados*, in *Mestrado em Ciências da Computação*. 2003, Universidade Federal de Pernambuco: Recife/PE. p. 128.

15. Moura, S.L.d., *Uma arquitetura para Integração de Repositórios de Objetos de Aprendizagem baseada em Mediadores e Serviços Web*, in *Mestrado em Informática*. 2005, PUC-Rio: Rio de Janeiro. p. 158.

16. W3C, W.W.W.C. *Web Services Architecture Requirements*. Web Services Architecture Working Group 2002 [cited 2007 12/03/2007]; Available from: http://www.w3.org/2002/ws/arch/2/wd-wsawg-reqs-03262002.xml.

17. Govindaraju, M., et al. *Toward Characterizing the Performance of SOAP Toolkits*. in *5th IEEE/ACM International Workshop on Grid Computing (GRID'04)*. 2004. Pittsburgh, USA: IEEE Computer Society.

18. SSE Team. *SSE Infocenter*. 2007 [cited 2007; Available from: https://services.eoportal.org/infocenter/index.jsp.

19. SSE Team. *SSE Interface Control Document*. 2006 [cited 2007; 156]. Available from: http://services.eoportal.org/massRef/documentation/icd.pdf.

20. ISO, *Geographic Information - Metadata.* 2003, International Organization for Standardization. p. 140.

21. Nencioni, P. *Earthnet Online XML Front-End*. 2006 [cited 2007; Available from: http://earth.esa.int/rtd/Documents/EOLI-XML-ICD.pdf.