GeoSemantica as a Technological Platform to Develop the Colombian Spatial Data Infrastructure

Julián ESCALLÓN, Colombia

Key words: Colombia, ICDE, Geosemantica, Ontology, Standards.

SUMMARY

At present, Colombia is participating in the GeoSemantica initiative funded by the Canadian government under the framework of the *Multinational Andean Project: Geoscience for Andean Communities (MAP:GAC)*. GeoSemantica is a set of computer tools and practices that allows efficient data integration through the use of information technologies that have been set up for integrating, translating and sharing information and knowledge in a distributed network environment.

Within the Colombian Spatial Data Infrastructure - ICDE (from its Spanish acronym) framework, it has been created a pilot project where a first integration of basic territory layers of several organizations was possible by using the GeoSemantica tools.

The layers have been placed and combined to form maps. They, in turn, have been published in the Web proving to be a very powerful, cost effective and efficient way of data integration and land information management. This application is currently being evaluated by different organizations at the government level in Colombia as a platform that suits present and future needs of data integration and public services of geographic data.

Thus, GeoSemantica is to be used as a leverage to access the data that many organizations in the country manage and maintain in order to share and integrate information from disparate institutions each contributing specific and diverse information and expertise, that when integrated improve the decision-making capabilities of governments and private sectors for territorial planning and development.

The generalized usage of GeoSemantica tools among data providers and users will definitely foster the achievements of the Colombian National Data Infrastructure in the near future. The main contributions will be on helping overcome the lack of knowledge about international standards for geographic data integration, and also on providing the country with a robust technological backbone for data discovery and visualization, allowing inter-operability with other nodes.

RESUMEN

En el presente, Colombia participa en la iniciativa de GeoSemántica, financiada por el Gobierno Canadiense bajo el marco del proyecto *Multinacional Nacional Andino: Geociencias para las Comunidades Andinas (PMA-GCA)*. GeoSemantica es un conjunto de herramientas computacionales y prácticas que permiten la eficiente integración de datos por medio del uso de tecnologías de información adecuadas para integrar, traducir y compartir información y acervos de conocimiento (geo-espacial y de noticias) en un ambiente de red distribuida. Para lograr estas metas, GeoSemantica usa los estándares del Open GIS Consortium, así como el protocolo Z39.50 para compartir y acceder a metadatos de manera global.

Al interior de la Infraestructura Colombiana de Datos Espaciales - ICDE, se ha creado un proyecto piloto donde, utilizando las herramientas de GeoSemántica, fue posible lograr una primera integración de capas básicas del territorio de varias organizaciones.

Estas capas han sido dispuestas y combinadas para formar mapas. Ellas a su vez, han sido publicadas en la Web, probando así que GeoSemantica representa una forma de integración de datos y manejo de información del territorio muy poderosa, costo-efectiva y eficiente. Esta aplicación está en el presente siendo evaluada por diferentes organizaciones en el nivel gubernamental en Colombia como una plataforma que satisfaga las necesidades presentes y futuras de integración de datos y servicios públicos de información geográfica.

Así, GeoSemántica puede ser usada como una palanca para acceder a los datos que muchas organizaciones en el País manejan y mantienen con el fin de compartir e integrar información de diversas instituciones, cada una de ellas contribuyendo con información específica y diversa y con experiencias que al ser integradas faciliten la capacidad de toma de decisiones del gobierno y del sector privado para la planificación territorial y el desarrollo.

El uso generalizado de las herramientas de GeoSemantica entre proveedores y usuarios definitivamente impulsará los logros de la Infraestructura Colombiana de Datos Espaciales en el futuro próximo. Esta contribución será indispensable para superar la falta de conocimientos en el País acerca de estándares internacionales que permiten lograr la integración de información geográfica, así como para suministrar una plataforma tecnológica para el descubrimiento y visualización de datos geográficos, permitiendo la inter-operabilidad con otros nodos.

GeoSemantica as a Technological Platform to Develop the Colombian Spatial Data Infrastructure

Julián ESCALLÓN, Colombia

1. INTRODUCTION

As it is the case in all countries of the world, there is an important need in Colombia to use large volumes of information with tools that maximize the security of the information yet optimize the processes of uploading, accessing and integrating information.

Colombia in the past has been quite active in developing its National Spatial Data Infrastructure, taking special care in addressing important implementation issues. Indeed, the production of fundamental data (framework) has been widely discussed, including standards, technical specifications and national production programs of geographic information. Additional effort has been put into developing mechanisms to improve data access and use by the public; they include metadata query systems for data discovery and recovery (Nebert, 2004).

Unfortunately, the concepts and benefits of a national spatial data infrastructure have not been easy to explain to government authorities and private companies, especially, because there were no real Colombian data being managed within the spatial data infrastructure philosophy. This drawback has been a key obstacle in developing the Colombian Spatial Data Infrastructure - ICDE, since the integrating infrastructure scenario had not been possible to be shown in a practical way to relevant government planning and decision making authorities.

Despite many organizational and policy issues are still to be fully addressed, it is very important to start working real data, including metadata and map services in order to show, in a straightforward way, the great advantages and savings a country could get just by managing its geographic information with a spatial data infrastructure.

At present, Colombia is participating, along the Andean countries of South America, in the *GeoSemantica* initiative funded by the Canadian government under the framework of the *Multinational Andean Project: Geoscience for Andean Communities (MAP:GAC)*. *GeoSemantica* is a set of computing tools and information management practices that allows efficient data integration through the use of information technologies in a distributed network environment. To achieve these goals, GeoSemantica uses the standards and recommendations of the Open GIS Consortium as well as the Z39.50 protocol for sharing metadata worldwide.

2. GEOSEMANTICA AND ITS FUNCTIONALITY TO HELP INTEGRATE THE COLOMBIAN SPATIAL DATA INFRASTRUCTURE

2.1 The concept of GeoSemantica

GeoSemantica is an ontology based web services architecture designed for integrating, translating and sharing information and knowledge assets (geospatial and news media) in a distributed network environment (http://www.pma-map.com/en/gac/index.html).

GeoSemantica is part of a much broader initiative within the Earth Science Sector of Natural Resources Canada to build a web-based architectural framework to situate and promote the use of integrated earth science information, knowledge and expertise within a broader societal context. Its objectives focus on the research and development of integrated knowledge systems and semantic web applications that will facilitate the understanding and uptake of Earth science information in support of interdisciplinary research, planning and decision-making on issues of public safety, resource management and sustainable development.

Combining semantic web browsing with Internet GIS and knowledge discovery tools, GeoSemantica is intended to help broaden and deepen both, an awareness and understanding of sustainability issues in the context of collaborative learning and decision support.

GeoSemantica will be fully OGC compliant with Web Mapping Services (WMS) and Web Feature Services (WFS) by using MapServer as Internet mapping tool. This feature ensures that it will be able to connect to other nodes worldwide and will allow us to integrate and share geographic information. Moreover, the usage of technology supported on the Z39.50 standard (ANSI/NISO, 2003) for metadata, guarantees inter-operability with other digital libraries and other clearinghouse nodes at the international level.

GeoSemantica uses XML as a core technology in a wide range of tools for creation and editing and for many data exchange operations, especially metadata and geographic information. This technology brings the advantages of inter-operation, keeping the semantics of the data, thus data may be integrated and understood by other systems, even if they do not have the same data model.

Taking into account that the whole package is going to be used by government agencies in South America, the maintenance and support costs became an important issue. Hence, in order to improve the sustainability, the GeoSemantica development team decided to use free software resources and in-house developments that are expected to be maintained and updated without any charge for licensing fees in the future.

2.2 Prototype of GeoSemantica and release of version 1.0

GeoSemantica has been developed mainly by the Geological Survey of Canada in partnership and with input from the seven Andean nations (Argentina, Bolivia, Chile, Colombia, Ecuador,

Peru, and Venezuela). Since 2003, a functional prototype is available in the Geological Survey of Canada regional office of Vancouver as a Web portal (http://node.geosemantica.net), which is based on a three-tier architecture and uses PHP, MapServer 4.1 and Macromedia Flash.

In this application, it has been possible to prepare more than 500 layers of geographic information including vector and raster data in many different geographic projections for the seven countries of the Andean region. Among these layers, a full coverage of processed optical LandSat images (http://www.tectonique.net/bf/sat/mosaic.html) for South America are now available, as well as some processed images with the results of the Shuttle Radar Topography Mission (SRTM) at a resolution of 90 meters, accomplished by NASA in 2000 (http://srtm.usgs.gov/).

Although the prototype does not provide all the flexibility and ease of use designed for the final application (which is currently being written for a .NET platform), it is completely functional and allows testing of the advantages and power of geographic information integration. The metadata server is not implemented in the GeoSemantica prototype yet, but it will be released with GeoSemantica version 1.0 by mid 2005.

The prototype has been used to handle geographic information in a workshop series that have been carried out with several institutions and in six main cities of Colombia.

In the prototype, there are six important modules as follows:

- News and information: This module can hold connections to selected thematic scientific news sources that are updated in an automatic basis.
- Knowledge browser: By using ontology tools, an easy search and query function is being implemented to look for specific information in many metadata sources.
- Digital Library: This module allows a user to store and read digital contents, including maps, multimedia and reports. It has the capability of connecting to other OGC compliant services in the network in a straightforward fashion. Development of a metadata discovery portal will be included in version 1.0.
- Projects: This set of functions is intended to allow working groups enjoy a safe and robust environment for communication among members, programming activities, assigning tasks, and producing and sharing geographic information in a private instance.
- Field Notes: This module provides the user with functions to capture data on the field that are geo-referenced to a custom map of a particular area. Different types of data are supported, including text documents, alphanumeric contents, maps, images and multimedia. Some functionality is currently being developed to support PDA's and mobile computers to allow operations outside of a network environment that will automatically synchronize their data once they are brought back to the GeoSemantica network.
- Map Engine and live maps: By using all the searching tools, maps may be assembled, saved, stored and authored in the Web by the user with layers available at the local node and at other nodes by using OGC's WMS and WFS. A mechanism to provide Near Real

Time Maps (NRTM) in GeoSemantica has also been included in recent developments. By using technologies such as XML and MapServer`s OGR tools, the user can integrate live data as just another coverage or layer in the digital library.

3. HOW GEOSEMANTICA COULD BE USED BY GOVERNMENT AGENCIES

3.1 The availability of GeoSemantica to government agencies

After evaluation of the prototype capabilities, it was quite clear that the potential of the GeoSemantica tools expands far beyond the scope of geological survey organizations. Through very simple operations, it was shown that it could be used as an integrating platform for all the territorial layers in a country.

The participation of many government agencies of Colombia that contribute information within the country has been welcomed by the GeoSemantica coordinators and has given an additional boost to the initiatives of spatial data infrastructure, giving a higher profile to the project. Indeed, GeoSemantica has been offered to government agencies in Colombia that wish to become information nodes with all the capabilities to integrate and exchange data, both for metadata and geographic layers and products. This platform will be available to be used locally after Version 1.0 is released by mid 2005.

3.2 How nodes will be connected when GeoSemantica is used by government data providers and other organisms

The first step will be installing all the applications of GeoSemantica in a proper server (Intel Architecture on Windows 2003), with an appropriate internet link that should be setup for providing Web services.

As a second step, the sources of local data should be located, either as a spatial database engine or as simple file storage in directories. GeoSemantica has been thought to support either scenario, making it possible to start using the application right after its initial installation, even if the source data is not fully structured.

The metadata of the organization should be available complying with the Colombian metadata standard (ICONTEC, 1999). These metadata are then published in a Z39.50 metadata server using a profile developed for the project, very close to the FGDC GEO profile (Blue Angel Technologies, 2003).

Once the servers are setup and the data are connected, they may be used to share, classify, and publish the data. Also, it is possible to integrate other sources of information along with their own layers, as long as the other sources are OGC compliant.

The primary relationships with other nodes of GeoSemantica or other clearinghouses will be through the metadata servers which in turn will be able to actually connect to the data sources.

3.3 Weaving the Network

A node becomes useful when it is capable of both, serving data and geographic contents to others, and communicating with other nodes that could provide additional information that is complementary to the one served locally. This achievement can be done if a network of nodes is properly weaved.

As more nodes are connected, the organizational complexities increase. To work out these future complications, a national policy, or at least some basic agreements on what should be available on each node to others have to be set at the beginning in the framework of the Colombian Spatial Data Infrastructure -ICDE.

Following the policy discussed above, an online catalogue of what is available in each node should be ready for other nodes.

For each major data supplier, one GeoSemantica node would be deployed. The topology of such network would be as shown in Figure 1.

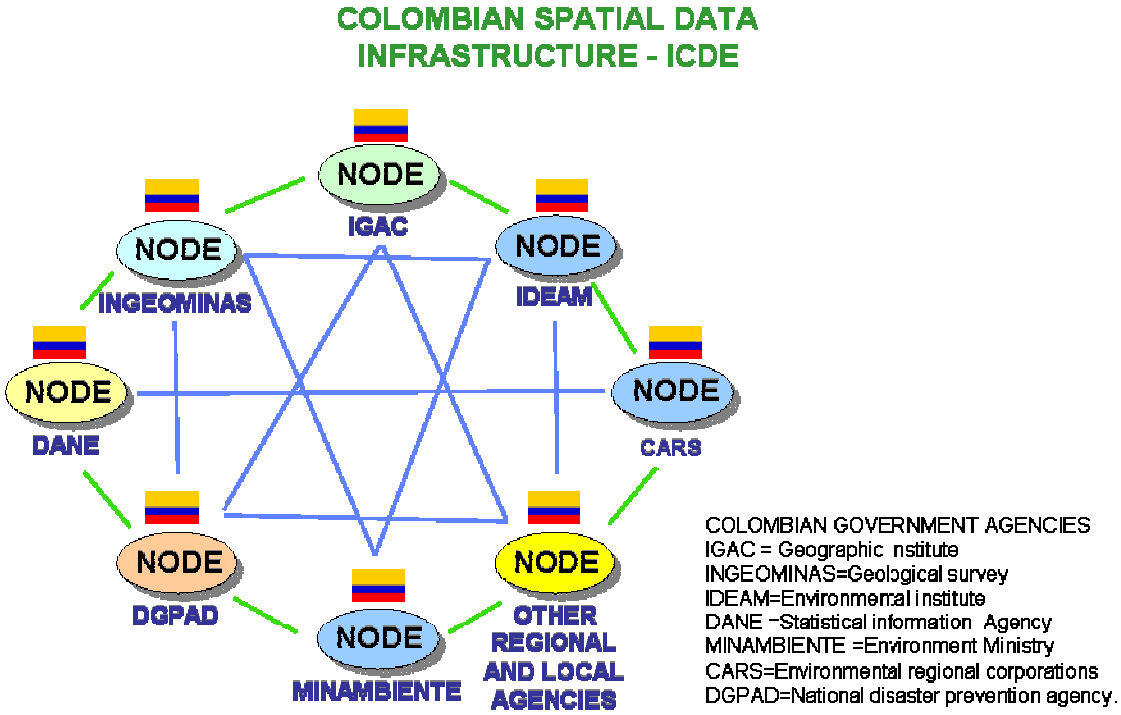

Figure 1. Government geographic data providers and users in Colombia linked as nodes. Some or all of them could rely on GeoSemantica tools.

Using WMS and WFS services, the relevant layers (or fundamental data) should be placed for being accessed by others. The individual layers may be available to anyone who wishes to connect to this service, in the specific address that identifies the node.

In the same way, by using the Z39.50 protocol, the metadata could also be discovered and then accessed. Like in any spatial data infrastructure, it is desirable to have at least a server that acts as a common gateway, so that all the servers can be queried in selected or simultaneous searches.

4. DEVELOPMENT OF A PILOT PROJECT FOR TESTING GEOSEMANTICA ADVANTAGES AND FURTHER STEPS

4.1 Basic Pilot Project already in place

Within the Colombian Spatial Data Infrastructure framework, it has been created a pilot project where a first integration of basic territory layers of several organizations was possible by using the GeoSemantica tools.

It has been possible to integrate the territory layers of several organizations. In 2003, in the technical committee of ICDE, it was agreed to use GeoSemantica as a test bed for setting up national geographic layers of several thematic interest. Hence, information from the Geographic Institute (IGAC), the Geological Survey (INGEOMINAS), the Environment Ministry, the Mining-energy Planning Unit (UPME) and the National Parks Unit, amongst others. These layers were positioned as a pilot project for the Colombian Spatial Data Infrastructure and they are the first integration example of this nature in the Country.

The national Maps available at GeoSemantica are as follows:

Thematic Map	**Number of Basic layers available at Geosemántica**
Administrative and political boundaries (IGAC)	2
Indian and Black Communities (IGAC)	2
Nacional Parks (Environment Ministry)	1
Total Coal Potencial Map (UPME)	1
Basic Geology (INGEOMINAS)	3
Geochemistry Anomaly (INGEOMINAS)	1
Gravimetric Anomaly (INGEOMINAS)	1
Mining Cadastre (INGEOMINAS)	1
Mining Inventory (INGEOMINAS)	1
Metalogenic Map (INGEOMINAS)	3
Potencial areas for precious and Basic metals (INGEOMINAS)	2
Ground Water Points (INGEOMINAS)	1
Seismic Hazard (INGEOMINAS)	1
Instrumental Seismicity (INGEOMINAS)	1
Historic Seismicity (INGEOMINAS)	2
Galeras Volcanic Map (INGEOMINAS)	8
Machín Volcanic Hazard (INGEOMINAS)	12
Mass movement Fenomena (INGEOMINAS)	1
TOTAL	**43**

By using GeoSemantica tools, the layers have been placed and combined to form maps. They, in turn, have been published in the Web proving to be a very powerful, cost effective and efficient way of integrating and managing land information and data. This application is currently being evaluated by different organizations at the government level as a platform that suits present and future needs of data integration and public services of geographic data.

Figures 2, 3, and 4 illustrate some of the maps that have been assembled with layers available at GeoSemantica, and that use data from different agencies. Figure 5 shows Galeras Volcano hazard map, using information of radar topography to illustrate the features of the surface of the area.

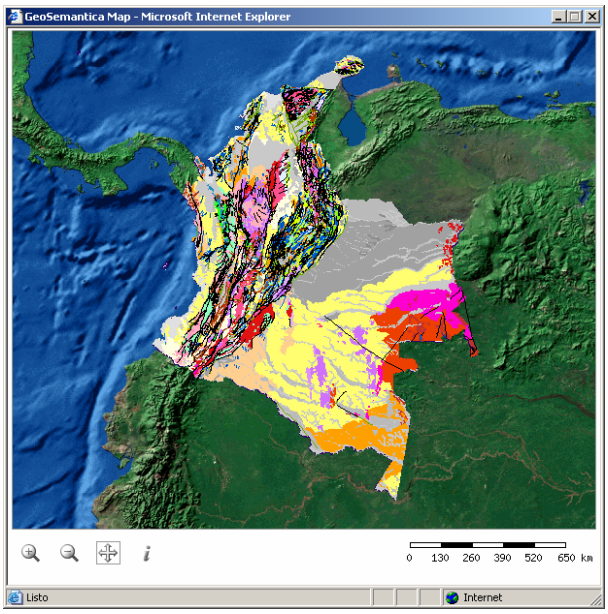

Figure 2. Geological Map of Colombia. (1:500.00)

If GeoSemantica is finally adopted by government agencies, it could be used as a leverage to access the data that many organizations in the country manage and maintain in order to share and integrate information from disparate institutions each contributing specific and diverse information and expertise that when integrated facilitate the decision-making ability of governments and private sectors for territorial planning and development.

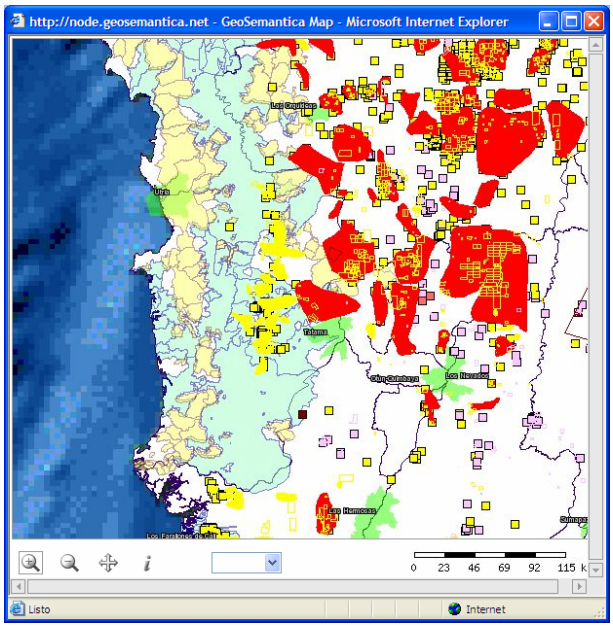

Figure 3. The Map of Mining themes (territory offer) and environmental and cultural restrictions (National Parks, Black and Indian Communities) allows users to access key information for planning future development.

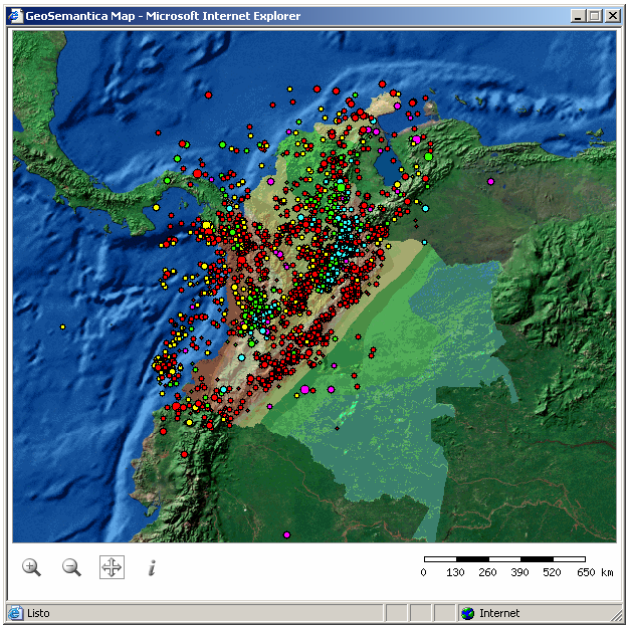

Figure 4. Seismic Hazard Map of Colombia with instrumenal seismicity on top.

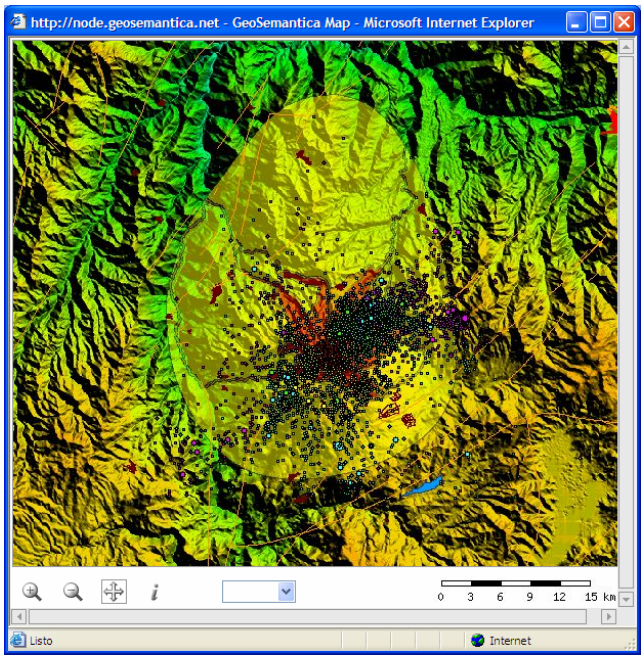

Figure 5. Geological Hazard Map of Galeras Volcano, Sourthern Colombia. It includes main municipalities and relief features from the topography information available from NASA.

4.2 Steps to follow

For the next few months, the Colombian Geographic Institute (IGAC) will complete the pilot project with some additional layers, especially those that are to be used in the census taking to be held in Colombia during 2005. Besides, IGAC is the main geographic data supplier and it is called to be the technological backbone for further integration and data exchange.

Other important organizations that are likely to continue contributing to the pilot project are as follows:

- The Colombian Geological Survey (INGEOMINAS), which is the very first GeoSemantica node set up for Colombia and is the agency with the highest number of layers available online.
- The Ministry of Environment with all the national, regional and local agencies committed to preserve and protect the environment.
- Bogotá, Colombia's capital city, which is already engaged in a local spatial data infrastructure initiative and is willing to follow further testing activities within GeoSemantica.
- Emergency attention agencies (such as the Bogotá's Direction for Attention and Prevention of Emergencies) which have to deal with real time information of several sources that are not easily accessed online.

5. CONCLUSIONS

The implementation of GeoSemantica with all the technologies involved in its development will allow many government agencies to implement, in a straightforward way, clearinghouse systems, with the same services shown by others already in place in several countries. The success of the implementation in the long term is closely related to the use of known standards such as those used by the GeoSemantica specification.

The intensive usage of GeoSemantica principles and integration tools among data providers and users could definitely foster the achievements of the Colombian Spatial Data Infrastructure in the near future at a relatively low implementation cost. It could help overcome the lack of knowledge about international standards for integration, and also to provide the country with a robust technological backbone for data management and inter-operability. Thanks to the pilot project deployment, the advantages and achievements of integration can now easily be shown to decision makers and future users, making it possible to involve more and more people at all levels.

The sustainability of GeoSemantica depends on how all the integration technologies and standards are going to be assimilated by personnel of the participating government agencies. It is also important to have a wide number of agencies that can invest resources for maintenance and development of new features in the future that can be shared with other partners within the Colombian Spatial Data Infrastructure Framework.

REFERENCES

ANSI/NISO Z39.50-2003. 2003. Information Retrieval (Z39.50): Application Service Definition and Protocol Specification. ISSN: 1041-5653. Revision of Z39.50-1995.

Blue Angel Technologies. Z39.50 Application Profile for Geospatial Metadata or "GEO", 2003.

ICONTEC. 1999. Norma Técnica Colombiana 4611 - Metadatos Geográficos.

Nebert, D. N., (Editor) 2004. Developing Spatial Data Infrastructures: The SDI Cookbook. GSDI Version 2.0. Technical Working Group Chair, GSDI.

BIOGRAPHICAL NOTES

Mr. Julian Escallon is a Civil Engineer from Los Andes University at Bogota (Colombia), and a MSc. in Geophysics from the University of Western Ontario (Canada).
He has worked for 16 years for the Colombian Geological Survey (INGEOMINAS) in Geophysics and Geoscience Information management.
He has been coordinator of the Technical Committee of the Colombian Spatial Data Infrastructure - ICDE for four years.
Currently he is a geographic information consultant involved with the GeoSemantica project and with the Colombian Spatial Data Infrastructure developments.

CONTACTS

Mr. Julián Escallón
Diag. 106 D #43-66 Apt. 102
Bogota, DC.
COLOMBIA
Tel. +571 6130104
Email: julianescallon@yahoo.com