Methodology for the Collection and Handling of Geological Data

Rodrigo Ávila Cipullo
UnB – University of Brasília
Institute of Geosciences
Brasilia – DF, Brazil
racshalom@gmail.com

Henrique Llacer Roig
UnB – University of Brasilia
Institute of Geosciences
Brasilia – DF, Brazil
roig@unb.br

Abstract— **Universities and other geological information producers have been suffering the negative effects of a lack of data organization and standardization. The purpose of this work is to create a new spatial geology database and to present a new methodology concept for internal procedures of data acquisition. This proposal follows the precepts of Spatial Data and vector Structure (EDGV), which were approved in 2008 by the National Commission of Cartography (CONCAR), created by the Brazilian government in order to standardize the structure of spatial data, facilitating data sharing, interoperability and rationalization of resources between producers and users of data and cartographic information. This is an important step for developing a geological model database that can be embedded in the context of the National Spatial Data Infrastructure (NSDI/ INDE).**

Keywords— geology data; managing data; web services; web mapping; geotool.

I. INTRODUCTION

Geoscience is a branch of science that requires the collection and processing of large databases, especially regarding spatial and geological phenomena. However, this development is overdue for improvement in safe storage,allowing a more secure use of databases in collaborative environments.

Many steps are being taken, in Brazil and internationally,to create an interoperability culture and standard of spatial data. In Brazil, the Decree No. 6666, from 2008, instituted the NSDI (National Spatial Data Infrastructure), which aims at establishing metadata and interoperability standards for basic cartography in Brazil.

These types of initiatives attempt to carry out the interoperability of spatial data. In this context, the OGC (Open Geospatial Consortium) must be highlighted. The OGC is defined on its own website [9] as a non-governmental, nonprofit organization, formed by volunteers from around the world, who intend to establish standards for spatial data and services.

Despite many efforts to set standards, the amount of geological information could not match the pace of the evolution of its standardization, and one of the only projects with global prominence is the GeoSciML [10]. This project has the goal of providing a data interoperability architecture that allows institutions with the most different types of databases to exchange information without changing their structures. The GeoSciML is an initiative of major geological surveys in the world, such as the IUGS [11] and the British Geological Survey [12].

A. Objective

The aim of this paper is to offer a new model of spatial geological databases, a new methodology for geological project management and the development of a tool that enables the implementation of the proposed model and methodologies.

Not every goal will be achieved at this first stage of the work, and the products presented in this article are:

- A conceptual database model,
- A proposed methodology for acquisition and management of geological data, and
- Easy use of web application for data visualization.

B. Structure of the Paper

This paper begins with an abstract and an introduction (I) that includes this topic. The other content topics are structured as follows: Proposed Architecture (II), Proposed Methodology (III), Data Acquisition (IV), Conclusion and Future Work (V) and References (VI).

II. PROPOSED ARCHITECTURE

The discussion about the ideal architecture for the provision of GIS has already been discussed by several authors. Harrower [5] defended that the use of the Internet would allow new services to GIS. According to Harrower, the Internet has revolutionized the way we work with cartography, featuring some important points that led to this new paradigm of GIS distribution:

- Ease of availability and distribution of cartographic products,
- Universal access to map data,
- Significant increase in demand for mapping services,
- Emergence of tools that allow the development of an "on-demand" application.

 ISBN: 978-1-61208-178-6

Another important concept presented by Harrower [5] is the "on-demand maps." According to the author, one of the main benefits of modern GIS is that it allows the user to manipulate and organize his or her own data, in which case the maps are not ready and static, but are constructed according to the needs of the user.

While we consider the globalization of access to GIS tools, we must also rethink the concepts of usability of these tools.

We have a growing public that is hungry for information, trained in a digital world, but does not necessarily have the adequate training to properly handle a GIS. The development of new GIS must consider this problem, during all the process from architectural definition to manufacturing the final product layout.

Another way that technology has taken hold of late is with cell phones and tablets. Increasingly powerful and with characteristics very similar to computers, these devices enable the technology to be part of every moment of everyday life.

Considering that a geologist uses several tools in a field study, it is possible that mobile devices provide the missing link between the data acquisition and storage/ distribution of information.

The use of mobile data acquisition not only represents a huge reduction in operational costs, but mainly simplifies the publishing process and data interpretation.

According to this work, this type of technology contributes greatly to the successful implementation of the new GIS, which allows the technology to spread with greater speed and makes possible the creation of a technological culture that is conducive to the advancement of real security and interoperability of geological information.

All points shown indicate the following architectural features:

- Centralized storage and web presentation,
- Centralized data processing,
- Acquisition of data using mobile devices.

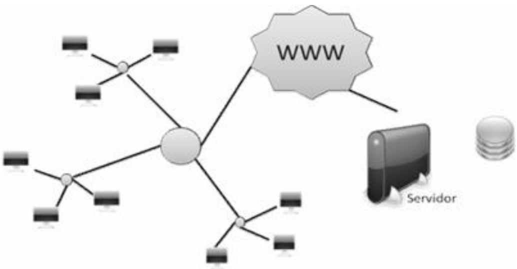

Figure 1. Architecture of centralized storage and processing type Legend: Servidor (Server)

A key to the success of the proposed methodology is to ensure that data is always available, and ensure that the database is always updated. For this, the tool uses web servers and the update will be done directly from mobile devices.

In order to provide this kind of service on the web, we need some essential tools, which are a web server and a map database with support for spatial data.

The web server application is responsible for providing a website or application hosted on one computer to another within a network, the global network known as the World Wide Web. The application server used is the Apache Web Server [14].

Map Server is the software that allows us to publish geographic data on the web. Through this application, we can provide a spatial database on the web through a series of specifications established by the OGC (Open Geospatial Consortium). Because they are published based in international open standards, the information can be reached by a wide variety of web and desktop software.

The application Geoserver [13] maintained by the Open Planning Project, was chosen in this project to serve the geological data, as shown in Figure 2.

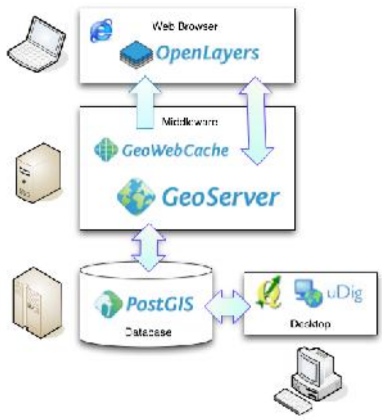

Figure 2. Basic publishing architecture of spatial data based on open source software [15].

III. Proposed Methodology

The proposed methodology is based on three steps:

- The standard remote data acquisition,
- Topologically-consistent database,
- Decentralized data management.

A. Data Acquisition

The data acquisition step represents the most important stage of the process, and the quality of data acquired following minimum standards is crucial to the success of the other steps. Thus, it is important that the data collection tool triggers a simple and standardized environment, ensuring flexibility in the collection that results in high quality data.

As shown in Figure 3, all acquired data are stored in a small SQLite database, exported by the system in CSV format and then imported by the Web tool.

This process will change in the next version of the mobile tool, when it will automatically synchronize the data with the server.

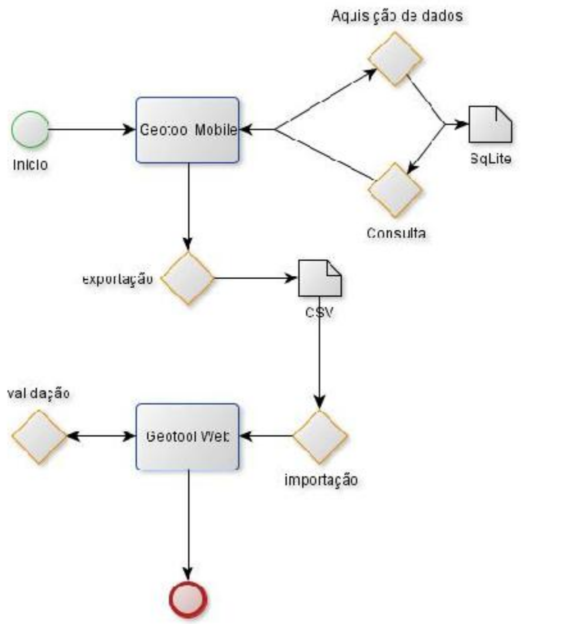

Figure 3. Flow data acquisition methodology. Legend: Início (Start); Aquisição da dados (Data Acquisition); Consulta (Query); Exportação (Export); Importação (Import); Validação (Validation).

B. Database

The second step of the methodology, the spatial database, was modeled using a methodology known as OMT-G.

The methodology OMT-G (Object Modeling Technique for Geographic Applications) was designed by Borges [4] based on one of the most popular models for modeling conventional databases, the OMT (Object Modeling Technique), which has the characteristic of representing the semantic aspects of data using an object-oriented approach. Thus, the OMT-G model, revised and extended by Borges [4], presents an object-based model that is also capable of representing objects and spatial relationships.

In addition to the OMT-G model, other proposals for modeling spatial data were created and must be remembered, such as the GeoOOA [1], MADS [2] and UML-GeoFrame [3].

The OMT-G brings together lots of geographic primitives proposed by various authors, as well as introduces new primitives that supply some deficiencies, such as the representation of multiple views from geographic entities. The OMT-G model is based on three main concepts: classes, relationships and spatial relationships [4]. This model works on the conceptual level as spatial classes featuring both conventional classes (Figure 8).

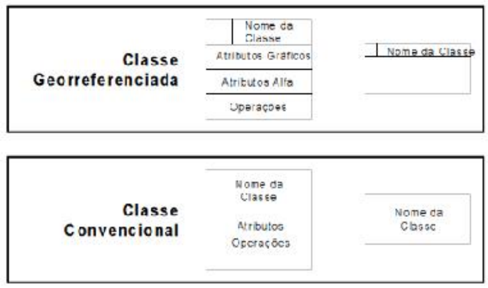

Figure 4. Types of classes and their representations in the OMT-G model. Borges et al. (2001). Legend: Classe georreferenciada (Spatial Class); Classe convencional (Conventional Class); Nome da Classe (Class Name); Atributos (Attributes); Operações (Operations).

Seeking an adequate representation of the types of spatial objects, spatial classes receive information from their geometry type. This information is known in the model as subclasses. The description of the main subclasses can be found in Figure 5.

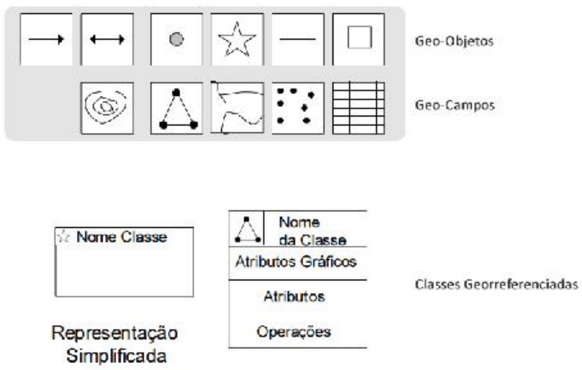

Figure 5. Georreferenced Classes model and its subclasses. Legend: Classe georreferenciada (Spatial Class); Nome da Classe (Class Name); Atributos (Attributes); Operações (Operations).

To generate the geographic database conceptual model, it was necessary to define some steps in the process of geological mapping. The following steps were also highlighted:

- Field survey (outcrops),
- Lamination,
- Geochronology,
- Geochemistry.

These steps were defined as superclasses of the model. These super classes do not have a real implementation, serves only as aggregators towards a better understanding of classes.

Each of the major classes that would be needed to correctly represent the stage of mapping emerge from the super classes. These major classes are presented in Table 1.

TABLE I. MAIN CATEGORIES

Superclass	Class	Description
Afloramentos (Outcrops)	Descriptive data	General data of the outcrop, such as name and description
	Toponymy	Information on the location of the outcrop that help visually its correct positioning.
	Geographical location	Information on latitude and longitude of the outcrop. It can be understood as coordinates west and south in the case of planar coordinates.
	Structural measures	Measures related to tectonic structures or primary structures.
	Photos	Storing pictures of the outcrops in question.
Amostras (Sample)	Samples	Samples related to an outcrop.
Laminação (lamination)	Lamination	General data such as name of the blade, which is owned and outcrop description.
	Paragenesis	Data on the composition of the blade modal.
	Photos	Storing pictures of the blade.
Geocronologia (Geochronology)	Geochronology	Basic information such as responsible for analysis, sample origin and age.
	Dating method	Method used in the analysis
Geoquímica (Geochemistry)	Geochemistry	Information on sample origin, responsible for analysis and consideration.
	Analysis	Types of analysis performed on the sample
	Results	Results of sample analysis. Load the type of analysis, the sample, the result, the laboratory and all considerations.

The next step is to define all the spatial classes. Each one of these classes will be implemented in the physical model. Following are the identified spatial classes.

TABLE II. SPATIAL CLASSES

Spatial Class	Description	Geometry type
Afloramento	Outcrops	Point
Estruturas (structures)	Tectonic structures	Linestring
Unidade Geológica (Geological units)	Differential unit of the crust for their compositional characteristics, age and physical boundaries	Polygon
Projeto (Projects)	Representation of the project area.	Polygon
Grupos (Groups)	Represents the work area of each work group.	Polygon
Contatos (Contacts)	Geological contacts	Linestring

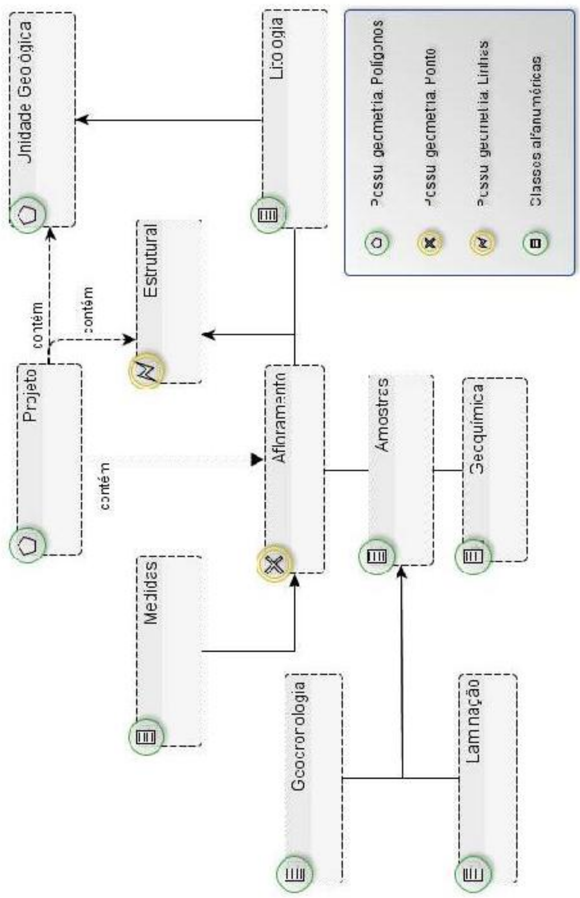

Figure 6. Model class indicating geometry
Legend: Table 1 and Table 2.

From the mapping of classes, a known model diagram class (Figure 6) was generated. This model was adapted to allow viewing the spatial types for each class. This adaptation allows us to clearly differentiate the non-spatial classes from others and display a preview of the spatial relationships that are present in the database.

Mainly to create the conceptual model of the geological database, each of the super classes have worked individually, especially in the quest for spatial delimitation of its features and possible relationships with other classes.

The first analyzed superclass, the spatial class project, aggregates all the others, functioning as the parent class in the model. It necessarily contains all other geometric classes of the database and represents the spatial delimitation of the mapping project. The class project represents a polygon geo-object, as already shown in Table 2.

The mapping project is formed from their groups. Groups are small areas that together, in the end, will aggregate to the final project area. Thus, all the mapping work is performed within a group. In practice, all classes are tied to the bank group, which is in turn aggregated by the project.

C. Data Management

The third step deals with the proposed availability of data. A tool able to manage its own projects and data was developed, allowing the publication of these data in other systems capable of utilizing WMS services.

1) Functional Requirements

- Access Control
- Management of users and permissions
- Project Management
- Creating projects and groups (subprojects)
- Definition of the project area and groups (subprojects)
- Outcrop Management
- Insert, edit and query outcrops
- Insert structural measures
- Photos
- Insert photo samples
- Entering, editing and deleting contacts in the sphere of geological group (subproject)
- Entering, editing and deleting slides related to the samples
- Management Isotope geochemistry samples held in
- The system must provide dynamic queries (spatial or not) and reports.

2) Non-functional Requirements

- User-friendly interface and simplified access to user data in a project,
- Light and simple software to compensate for the large volume of information.

3) Business Rules

- The user should be able to insert outcrops and contacts only within the boundaries of the area of his/her group,
- The user should be able to insert outcrops and contacts visually by map or via a form.

4) Mapping and Projects Management System

The main product developed, the management system of mapping projects, is an application that runs in a web environment and is able to manage since the creation of a new design until the closing of the geological map.

To expedite the development of this application and ensure easy maintenance in the code, we used the Zend framework [6]. The choice of this framework is given for the following reasons:

- This design pattern isolates the application logic from presentation logic. It allows them to be tested and/ or modified separately, reducing development time and enabling better reuse of code;
- Object-oriented library that allows easy extension and reuse of code;
- Abstraction of the database, allows for the automatation of a common operations database as well as decrease the impact of changes in the model;
- Automating operations AJAX and JSON, facilitating the task of integrating with the API for manipulating spatial data OpenLayers [7];
- Framework is maintained by the same company responsible for the PHP language, which makes it very popular, facilitating future maintenance.

In addition to the framework used in the development of the PHP application, some other open source APIs were used in order to improve user navigation. They are the following:

Openlayers [7]: Set of open source tools, available as a JavaScript API for viewing and manipulating spatial data based on OGC standards.

JQuery [8]: This is one of the most powerful JavaScript API today. The main objective of JQuery[8] is to provide a web browsing experience and practice based on modern concepts of AJAX (Asynchronous JavaScript and XML)

5) Restricted Access and Security

To ensure data security and proper control of projects, we implemented a system of user verification. The user is identified by his/her e-mail address and password (Figure 11). Immediately upon entering the system, the user will be prompted to choose which project to work on. Thus, he/she can work on any project that is registered.

 ISBN: 978-1-61208-178-6

Figure 7. Screen to access the system.

The definition of access permissions, registration of users, projects and subprojects in the system itself is made by an admin user. The system allows a user to be registered on several projects, however, the user must belong to a single subproject.

Once registered, the data remain in the system. This means that even if someone has access to a user account and makes lots of changes in it, or even delete some data, these data can be retrieved by the system administrator through a restitution version.

It is important to note that all project data is nested in its subprojects and always related to a registered user in a subproject. If you do not have a subproject it should be created so that the work can be started.

6) *Features*

The system is designed in modules, that is, it is possible that some features are only available to a user according to his/her profile. Modularity also facilitates system maintenance and code reuse for other future projects.

The modules developed so far include the major classes of database:

- Outcrops: Responsible for managing user outcrops in a particular subproject. The basic function is to list all the outcrops of the user on a given project and allow him/her to add, modify or delete outcrops. From the list of outcrops, the user can view all information related to an outcrop, including its visualization on the map.

  In addition to the basic features of the outcrop, several sub-modules that add functionality to the system were developed. They are:
  - Measurements: Stores all measurements taken at the outcrop. This module not only allows the listing and input measures, but measures of export in text format to be opened in structural geology software.
  - Samples: Relates physical samples to the outcrop.
  - Petrography: Stores blades from the outcrop samples.
  - Geochronology: Responsible for keeping the data for isotopic analysis performed on samples.
- Map: Module that aggregates spatial data and allows viewing and editing data.
- Structures: Displays all the structures of the related group and allows creating and editing new structures and contacts.
- Topology and closing contact: Validation responsible for implementing topological rules and close contacts between groups. Such procedures are necessary before generating the geometry of geological units, which are automatically created through spatial operations.

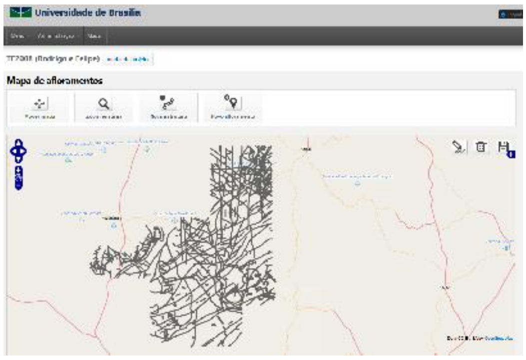

Figure 8. Module map showing some cartographic tools

IV. CONCLUSION AND FUTURE WORK

The application used for deploying PostgreSQL is sufficiently mature and stable to include large databases like the one here proposed. Its spatial component, PostGIS, responds very well to spatial analysis, demanding, however, even more studies and tests on their topological functions in order to overcome some difficulties that persist.

It was hoped an algorithm that would allow the automatic generation, at runtime, of geological units. However, none of the generated algorithms were able to properly turn contacts into areas and receive individual attributes of geological units correctly. This feature was only partially implemented and needs further studies.

Applications developed for the management of geological data responded positively to the tests in which they were submitted. However, there are security-related points that remain to be discussed further, before using these tools on a large scale.

REFERENCES

[1] G. Kösters, "GIS-Application development with GeoOOA". Int. Journal of GIS, 11, 1997

[2] C. Parent, "Spatio-temporal conceptual models: data structures + space + time". In Proc.7th ACM GIS, Kansas City, 1999.

[3] J. L. Filho, A. C. Costa, and C. Iochpe, "Projeto de banco de dados geográficos: mapeando esquemas GeoFrame para o SIG Spring". In Proc. GEOINFO – 1st Brazilian Workshop on geoinformatics, Campinas, 1999.

[4] K. A. V. Borges, C. D. Davis Jr and A.H.F. Laender. "OMT-G: an object-oriented data model for geographic applications". GeoInformatica, 5, 2001.

[5] M. Harrower, "A look at the history and future of animated maps". In: Cartographica n. 39: 2004. pp. 33-42.

[6] http://framework.zend.com (retrieved: August, 2011)

[7] http://openlayers.org (retrieved: November, 2011)

[8] http://jquery.org (retrieved: December, 2011)

[9] http://ogc.org (retrieved: December, 2011)

[10] https://eegrid.csiro.au (retrieved: November, 2011)

[11] http://iugs.org (retrieved: November, 2011)

[12] http://bgs.ac.uk (retrieved: November, 2011)

[13] http://geoserver.org (retrieved: December, 2011)

[14] http://apache.org (retrieved: December, 2011)

[15] http://opengeo.org (retrieved: December, 2011)

 ISBN: 978-1-61208-178-6