Spatial Database to Store Years of Earth Observation Information Obtained from Field Expeditions in the Amazon

Gabriel Crivellaro Gonçalves[1], Lucas Augusto, Maria Isabel Sobral Escada[1], Silvana Amaral [1]

[1] Remote Sensing – National Institute for Space Research (INPE)
Av. dos Astronautas,1758, CEP 12227-010, São José dos Campos, SP, Brazil.

gabriel.goncalves@inpe.br, lucasaugusto@gmail.com, {isabel, silvana}@dpi.inpe.br

Abstract. *Earth observation information obtained through remote sensing and field expeditions generate large volumes of data. These information are usually stored on physical drives such as hard drives. However, this data must be available to all researchers and be stored at safe places. This paper presents the preliminary results of modeling a spatial database (SDB), and some tools for storing Amazonian field expedition information of land use and localities features. Now, an information subset is already being stored in the SDB using the web platform for management. This work contributes to highlights the relevance of database approach, making the management, access and query of this information easier to research purposes.*

1. Introduction

Spatial Databases (SDB) are computer systems created for store, manage, read and write information with spatial attributes in a database. Unlike a common database (DB), SDB is capable to model and store spatial attributes such as location, extent and shape of geographical features. SDB also enable different spatial operation with spatial data, like union, intersection, spatial cutout, overlay operation and others [WINSTANLEY, 2009]. These are capabilities very useful for different knowledge areas, such as land-use and land cover studies based on remote sensing and spatial data.

For the last 10 years, the Laboratory for the Investigation of Socio-Environmental Systems (LiSS) from National Institute for Space and Research (INPE) has been studying the processes involved in Land Use and Land Cover Changes (LULCC) through Remote Sensing (RS) in the Brazilian Amazon, specially at Pará state, as described in the respective research reports [AMARAL et al., 2012; DAL´ASTA et al., 2011; ESCADA et al., 2013]. Remote Sensing data requires validation *in loco* and additional information to understand the LULCC process.

A large amount of data has been generated and obtained for each field expedition: remote sensing images, vector files, spreadsheets, photos, audio recordings, videos and GPS points and tracks. A total of 700 gigabytes of information is already saved in physical store units like hard drives. However, this strategy of information storage is risky, once this hardware is defective, all information may be lost. On the other hand, since several researchers in the group need to access the data, is it essential to have an integrated platform from where everyone could access the information.

To cope with this problem, this paper describes the spatial database model developed to store field-acquired information and remote sensing data, and its initial results. The main objective was to build an SDB to store all the information in a practical and secure way, based on web service access for information insertion, query and download, also accessed via geographic information systems.

2. Data

A database for fieldwork purposes should handle raster data, as satellite images, vector files, such as location points and trajectories, and also documentary data such as spreadsheets, videos and photographs. Figure 1 describes the present composition of fieldwork data according its data type. For the final database, the data amount should be reduced, like photographs, by selecting only the most significant ones.
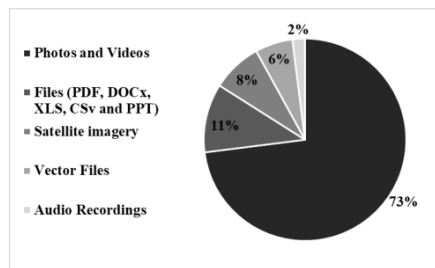

Figure 1.Type of fieldwork data available from 2008 to 2018.

The time and spatial reference of the data is central: information is referred to localities, towns, cities, institutions and trajectories, at different expeditions and dates. For each place visit on the field, indexed by a GPS coordinate, it is usual to have information from one or more questionnaire, with its audio records, and a physical description based on photographs and notes (Figure 2). After the fieldwork, all information should be organized in worksheets and standardized to compose the SDB.
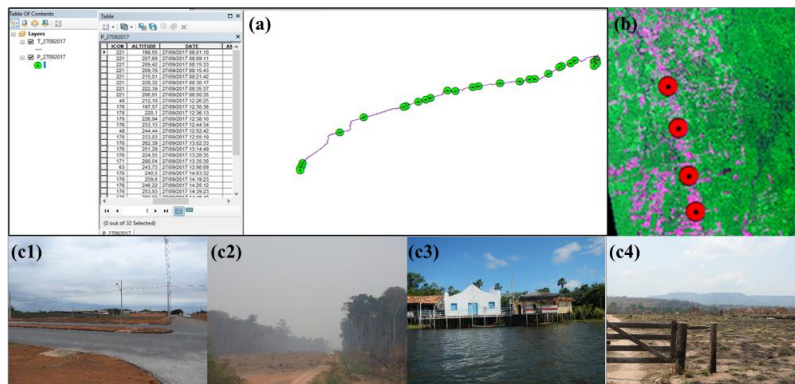

Figure 2. Fieldwork data examples: (a) worksheet with localities (points) information; (b) points to be checked over a reference Landsat image;(c) Land use and cover photos – urban area (c1), deforestation (c2), riverside community (c3) and pasture (c4).

135

3. Spatial Database

From fieldwork data and the needs presented by the users, a conceptual model for the SDB was proposed (Figure 3), considering as data categories: expedition (date), places visited (name and geographical coordinate), expedition raster (RS images), route (trajectory lines), expedition documents, questionnaires (applied to persons or institutions), questionnaire documents (authorization of rights of use of images, and any documents received from the interviewee), questions and answers from questionnaires, communities (name and geographical coordinate of each locality), images (photographs) and interviewed.

All information is organized by expedition, having a start and final date. Any answer for a question in the questionnaire is associated to a location, an expedition, and an interviewee. The photographs, information and documents obtained by the interviewee are also linked to the answer. This way, queries in the web platform can be performed consulting by date, expedition, locality, interviewee, type pf questionnaire, questions or answers.
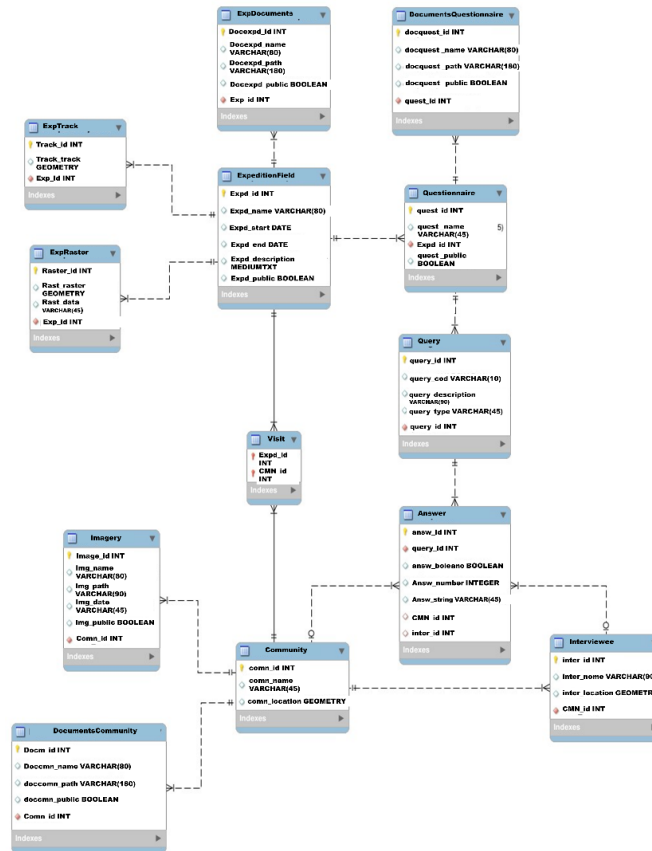

Figure 3. Conceptual model of the database.

From the SDB project specifications, the system should manage all the fieldwork data and provide additionally tools for analysis (spatial and attributes queries) and easy insertion of new data. First, the PostgresSQL was chosen as DBMS (Database Management System) because it is open source, stable, has great community support, and it contains the PostGIS extension, to manipulate geographic objects within an

object-relational database. The extension supports from storing multiple geometries to advanced operations of intersections and polygon joins [QUEIROZ & FERREIRA, 2006].

Then, a web system was proposed as user interface for the SDB. This interface was developed in Java language, using the Web Servlet Container. Such a system has a Restful API for data access for analysis, and a data management Web interface, which allows the CRUD (Create Read Update Delete) of objects to be made [PILGRIM, 2013].

Another feature of the system is easy-insertion of the data. The spreadsheets are loaded into the system, which extracts the originated entities, validates if the spreadsheet information is correct, and finally enters the database. All system persistence is managed by the Hibernate Object Relational Mapper (ORM). In it, it maps the objects to the respective tables in the database [IRELAND et al., 2009].

4. Web service platform and SDB management

In the web platform developed for the SDB management it is possible to insert, consult and visualize the fieldwork information. There is an initial user management interface to control and authorize the access by login and password. Once logged in, the page "management" enables the access to seven tabs: Expeditions, Questionnaires, Communities, Interviewees, Questions, Answers and Documents (Figure 4). These pages enable access to all the information in the SDB.

The "Expedition" tab enables insertion and consulting of expeditions, containing information about starting and ending date, and a brief description. The "Questionnaire" tab enables insertion and consulting of applied questionnaires, and associates each one to an expedition. The "Communities" tab enables insertion and consulting the locations in the database, including the name of the place, latitude, longitude and a brief description. The "Interviewed" tab contains some reference about the institution or person interviewed, like name, contact, latitude and longitude of the interview site and a brief description. The "Questions" tab contains the questions asked in the questionnaires. The "Answers" tab contains the answers, enabling consulting. And the "Documents" tab enables insertion and consulting of photographs, video, PDF files, MSOffice files, satellite images and vector files.

Figure 4. Overview of user interface of the Web Service access to the Spatial Data Base

An interface for automatic insertion (CRUD Wizard) was created to facilitate the data insertion into the database (Figure 5). For this option, data should be first summarized on three files: geolocations, questions of the questionnaires, and answers. The first worksheet (Figure 6a) creates the communities and assigns a file ID to each community with its geographic coordinates. The second worksheet (Figure 6b) resumes

the questions asked in that questionnaire with the code created for them, and the third worksheet (Figure 6c) contains the answers, with the question ID and the FID of the community. At CRUD Wizard page, a zip file containing the data template is available for download.
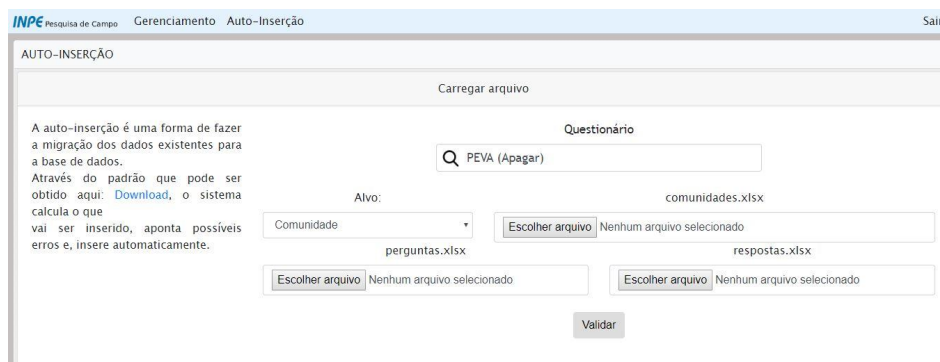

Figure 5. Interface for automatic data insertion into the database (CRUD Wizard).

To use CRUD Wizard, the tabular data must be standardized as shown in Figure 6. Due to the limitation of characters in the header of DBF worksheets, file with the information in shapefiles, it is necessary to create codes for the questions. The questions are in the columns and the communities in the lines, so an FID is assigned for each location (a) and code for each question (b). In the response worksheet, there will be the question code and the community FID (c).

Figure 6. Examples of standardize spreadsheets to automatic insertion into the SDB (a) geolocations, (b) questions and (c) answers.

Currently the system is online, the web platform and SDB. There is still some functionality to implement such as Web Map Service (WMS) and Web Features Service (WFS). Also, the creation of a page to make data that can be made public remains. About 200 gb have already been cataloged and are being inserted into the system. Expeditions of 2013,2015 and 2016 have already been cataloged. Only 2015 is already in the database.

5. Final Remarks

This work presents the spatial geographic database created to store data from fieldwork expeditions in the Amazon, carried out by INPE researchers from the LiSS. The database used PostgresSQL management platform, which allows the use of PostGIS. To make database management easier, a web platform was created for inserting information and querying in the SDB. The variety of original tabular information led the project to propose spreadsheet standardization for data insertion into the system.

The logical and physical structure of the SDB is already accomplished. The web platform for management is online and at final test phase. From the 700 gigabytes of the original fieldwork data, 200 GB was already cataloged, organized and it is being gradually inserted in the SDB. The spatial analysis and consulting will be possible when the web map service (WMS) and web feature service (WFS) is completed, and then, data will be accessed straightforward from a GIS.

At the moment, the SDB finally organizes data collected from a large time of providing data storage and data access to the researches, in a safe and integrated environment. This is an essential tool for researches a great ease in accessing the information generated by the group, making the storage more practical and safer since all information is standardized and stored on a server. The SDB is an important achievement, considering the possibility of further make part of these fieldwork information available for public.

Acknowledgment

References

AMARAL, S. ; BRIGATTI, N. ; DAL'ASTA, A. P. ; ESCADA, M. I. S. ; SOARES, F. R. . "Tem fofoca na currutela" Núcleos urbanizados e uso da terra de Alta Floresta (MT) ao Crepurizão (PA) na Transgarimpeira. São José dos Campos: INPE, 2012.

DAL'ASTA, A. P. ; ESCADA, M. I. S. ; BRIGATTI, N. ; AMARAL, S. . Núcleos de ocupação humana e usos da terra entre Santarém e Novo Progresso, ao longo da BR-163 (PA). São José dos Campos: INPE, 2011.

ESCADA, MARIA ISABEL SOBRAL ; DAL&APOS ; SOARES, F. R. ; ANDRADE, P. ; PINHO, C. M. D. ; MEDEIROS, L. C. C. ; CAMILOTTI, V. L. ; FERREIRA, V. C. ; AMARAL, S. . Infraestrutura, Serviços E Conectividade Das Comunidades Ribeirinhas Do Arapiuns, PA. São José dos Campos, SP: INPE, 2013.

IRELAND, CHRISTOPHER & BOWERS, DAVID & NEWTON, MIKE & WAUGH, KEVIN. Understanding object-relational mapping: A framework-based approach. Int J Adv Softw. 2. 2009.

PILGRIM, P. A. "The lifecycle of Java Servlets". Java EE 7 Developer Handbook. Professional expertise distilled. Packt Publishing Ltd. ISBN 9781849687959. Retrieved 2016-06-16. Java Servlets are governed by a web container (a Servlet container). 2013.

QUEIROZ, G. R. FERREIRA, K. R. Tutorial sobre Bancos de Dados Geográficos. GeoBrasil, 2006.

WINSTANLEY, A. C. Spatial Database. International Encyclopedia of Human Geography. P 345-347. 2009.