World Wide Access to Amazon Forest Inventories of Non-Timber Products

Felipe Costa[1], Sylvain Desmoulière [2], Oscar Farias[3]

[1] Dept of Computer Engineering, Rio de Janeiro's State University, Rio de Janeiro, Brazil, fsc7mail@gmail.com

[2] SIGLAB, National Institute for Amazonian Research, Manaus, Brazil, syltao@gmail.com

[3] Dept of Computer Engineering, Rio de Janeiro's State University, Rio de Janeiro, Brazil, fariasol@gmail.com

Abstract

The Amazon rainforest has gained importance due to its role in maintaining the balance of global climate and preserving its biodiversity. Thus, all initiatives that contribute for the sustainable development of the region are relevant. At the National Institute for Amazonian Research - INPA, one investigation being carried out uses data from forest inventories of RADAM, a project of the Brazilian government that mapped natural resources and geographical aspects of the region during the 70's. Using data from RADAM, the present research aims at providing information on non-timber forest products existing in the region, as a way of promoting the local development and business based on products that use renewable natural resources as raw material. The project involves the recovery, quality control and conversion of that information to a Postgis database, thus providing homogeneous data in interoperable formats for analysis, visualization of ready-to-think data layers and the analysis results on line. It also aims to distribute the results in a more attractive way to the user and to allow the possibility of spatial and non-spatial queries to the database. To meet these goals Open Geospatial Consortium - OGC web services WMS and WFS were generated. The project made available an interface enabling users to easily find the data they need using Geonetwork, an interface, called I3Geo, for viewing these geographic data on the web and Google Earth are also being used as platforms for dissemination of knowledge of biodiversity in the region.

1. Introduction

The concern with climate change has dramatically increased in the past few years. In this context, studies (Soares-Filho *et. Al*, 2005; Fearnside, 2006) show that the Amazon rainforest has great importance in maintaining biodiversity, cycles of water and carbon stocks. The deforestation and the burning contribute to the release of carbon dioxide, the major factor of greenhouse effect. Such activities occur in the vast majority due to pressures from exploratory economic activities, which will ultimately destroy the forest. Several initiatives have been taken so that the development of the region can be done in a sustainable way, using renewable resources and the enormous biodiversity of the region; however, society as a whole needs more information about the existing resources and how to develop itself making use of these resources without destroying the forest.

At SIGLAB, Thematic Geoprocessing Laboratory, at the INPA, investigations have been carried out on the Amazon region, using Geomatics tools. One of the research activities (Desmoulière and Carneiro Filho, 2007) of the lab uses data collected during the project RADAM as a way to promote sustainable development. One of the objectives of the project RADAM refers to a forest inventory, which was a field survey of 3131 observation areas within the Brazilian Amazon, where information was obtained from more than 147,000 trees of 771 different species resulting in the largest forest inventory ever created. Apart from the forest inventory data, were also part of Project RADAM the mapping of the vegetation, soils, geology and geomorphology of the Brazilian Amazon.

The technology currently available allows these data to be used for different types of users with different knowledge and goals. Through the provision of the data from the Project RADAM using the SOAP and OGC web services, catalogued in a metadata manager, and through outputs in multiple formats—KML, shapefiles, CSV and spreadsheets—it is expected that society may use the data on the location of non-timber species in the region to generate information that promote the sustainable use of these resources.

2. Data recovery

The database of the project RADAM went through a long process of development to achieve its current status. Initially, in the mid-70, the original data consisting of thousands of research field books generated several volumes of reports on paper. Gradually the data from the field research were scanned by the Brazilian Institute of Geography and Statistics, IBGE, and stored in various formats until it reached today's dual architecture of GIS with an Oracle relational database with several shapefiles associated, each with a field identifier linking textual information to geographic features .

Nevertheless, these data were not being well exploited because users outside the SIPAM-System of Protection of the Amazon-, the providing organ of such data, could not even manage to set up and import the data into an Oracle server, let alone relate the data with shapefiles. It is relevant to remind that potential users of such data are mostly biologists, environmentalists and forest engineers, professionals who are not usually trained to work with relational databases. The shapefiles themselves were useless because they did not have the information provided by the data richness of the relational database. They had only features without information.

The first step in the data recovering process was the installation of DBMS Oracle, the recovery of data from the dump files and export of data to PostgreSQL, which, being a robust database and having the Postgis spatial extent, would allow the spatial attributes to be related to the tabular data. The use of Oracle Spatial was discarded for economic reasons as well as MySql Spatial for issues of immaturity in attendance of the SF/SQL patterns.

To migrate data from Oracle to PostgreSQL in a massive fashion, for there were more than 90 shapefiles and more than 80 tables of data, functionality of specialized tools was tested in integration and conversion of data, using Spatial Data Integrator (SDI Talend 2.2.1) PostgreSQL and Oracle Import, Export & Convert Software 7.0. Neither of the tools exported relationships and SDI still needed each table to be exported individually, which would demanded a lot of time. Therefore, we opted to make the export manually adapting the generation queries of the Oracle tables for the PostgreSQL preserving the relationships between the tables followed by the import of data in CSV, the spatial data being converted from shape to Postgis using the shp2pgsql and ogr2ogr tools.

Data from the forest inventory basically consists of the tables below:

Table 1. Tables of the forest inventory

Name	Description
tree_inventorylayer	Spatial table containing the sampling points, and information on the environment at each point. This table originally did not contain spatial data from the shapefile that were added in Postgis.
tree_obs	Intermediate table that connects the sampling points with the found species.
es_dc_species	List the species found with gender, family and name of species.
es_dc_commonname	List the ordinary names of species.
es_dc_speciessynonym	List the scientific names that the same species may have.
tv_metod_amostrag	Table that describes the types of sampling.

Data from soil, geomorphology, hydrography and geology were also migrated, which are layers that have not suffered significant changes since the 70's and still consist of the most comprehensive data on the Amazon and reference to any research on the natural environment in the region.

From these tables a reverse engineering was done to draw a entity-relationship diagram from the data into PostgreSQL. To do that the OpenOffice Base software was used. From that diagram it is possible to reindex tables so that access to the database is optimized, see Figure 1.

From that standardized/normalized database, we are able to give specialized users, researchers and developers in general a model of the entity-relationship together with the data dictionary so that it will be aware of what exists in the database.

Another step is to meet the creation of pre-manufactured queries (views) and a way of visualization to attend non-specialized users.

3. Dissemination

The main data to be released in this early stage are forest inventories data that are basically sampling points with information on the fieldwork at the time and the relationship of species found at that point with its proper frequency. The collections were made in an area of 0.04 ha per 1 hectare, each tree within that area being registered.

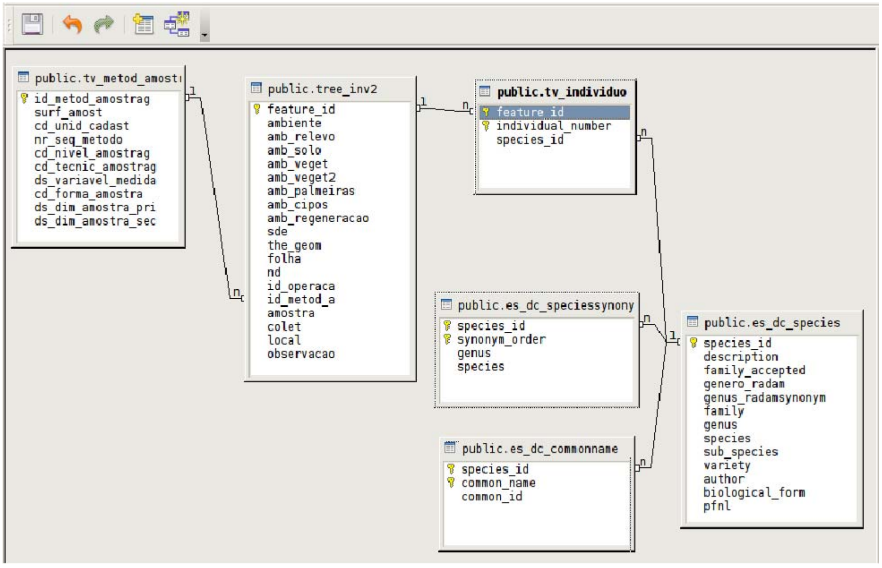

Figure 1. Diagram of entity-relationship refering to forest inventories.

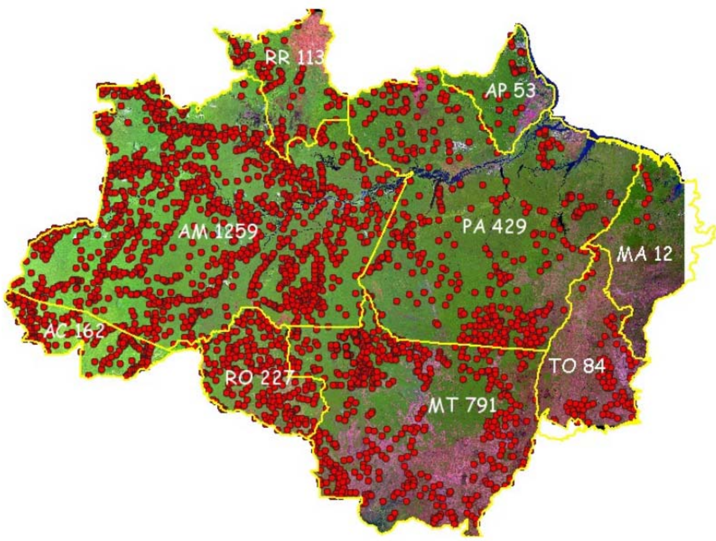

Figure 2. Distribution of sampling points of the forest inventory in the brazilian Amazon.

Although there is a strong deforestation in the south of the Amazon basin and a selective cutting of noble species throughout the region, making the data collected by the project RADAM in some areas no longer representative of the reality, such data do allow, for instance, to vizualize of the geographical distribution and abundance of a particular family or genus in the Amazon region, to know the average size of trees and to estimate the volume of timber in a given region. Through the identification of the type of vegetation from a particular point of sampling, for instance, one can generalize and estimate the carbon stock in a given region, the possibilities are numerous.

The release of tabular data for environmentalists and macroecologists allows the development of various scientific papers, but the main objective of this work is the dissemination of spatial data for technicians and professionals who contribute to the promotion of non-timber products, those who can support the creation of public policies for sustainable development for the region.

To make the data available, users were categorized into two classes: experts and non-specialists. Expert users can be both researchers with expertise in GIS and relational database as well as software developers that can use these web services to create applications. All they need in addition to the data available through web services properly catalogued is the type of entity and relationship and data dictionary; from then it is possible to develop a web application based on these data.

The OGC web services were available through the Geoserver software, it was chosed because its great ease of inclusion of new data as web services that this tool allows and the possibility of having different types of outputs configured automatically to the same spatial table, or view of PostgreSQL/Postgis, the use of software as Mapserver, FeatureServer andr Deegree were not discarded for each had its own potential (Carnevalle, 2008; Metacarta, 2008).

Geoserver allows that, once it has been configured to access the Postgis database as a data source, each table or view could be available as Web Map Service (WMS), Web Feature Server (WFS), Keyhole Markup Language (KML) and shapefile; besides, it is possible to make some changes of coordinate system, facilitate the creation of thematic maps for non-specialists, using Styled Layer Descriptor (SLD), which may be created using the uDig software.

Not all tables have spatial attributes but for those potential users fully benefit from the data it is necessary that they know the structure of the relational database. One possibility is the creation of views where the user has a pre-prepared query by the data maintainers and hence its WFS becomes available. It is theoretically unfeasible to create all possible combinations within a database having a considerable number of tables. It is much more interesting to allow users to make this combination of data, including the use of other table sources.

Non-geographical tables of the relational database are available using the W3C standard SOAP, with the services described by WSDL, the use and the relationship between geographical and non-geographical tables should be done now through applications. Some direct requests are being addressed through the creation of views, and available through SOAP and WFS. There is also the possibility of direct access to the database for users of the intranet who can retrieve spatial and non-spatial data, and run queries through Mezogis as software, Kosmo and GvSig.

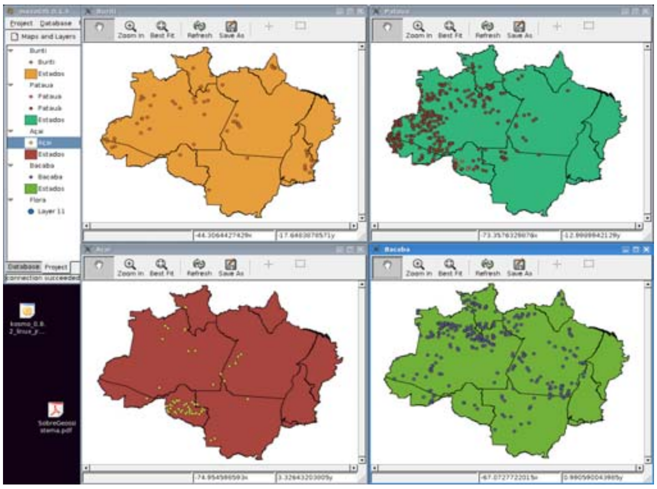

Figure 3. Analysis of spatial distribution of species by means of software Mezogis.

To allow the user to access spatial and non-spatial tables through web services, and be able to make the relationship between them within a GIS, it is necessary that the software has the capability to consider web services as data sources; however, no GIS Desktop has this capability today. To circumvent that obstacle in terms of data integration from a WFS and data from one or more tables available through SOAP web services, we plan to create mechanisms that allow GIS programs to see SOAP web services as data sources.

There are some alternatives for development, the first would be the construction of a plugin for a GIS. This plugin could read a WSDL of a Web service and the software could connect to its endpoint and download the data and, hence, consider the sets of data received as a table from which it may be associated with any given geographical through the key attributes. This approach has the advantage of being able to connect to any web service, though the plugin developed can only be used in a specific GIS.

The second way is to create a layer between the web service and a source of information available to various software, masking the web service as a database for example. This technique (Carroll and Calvo, 2004) allows web services to be accessed and queried using SQL as if they were tables in a database. Briefly, the technique is to create a function in a PostgreSQL database which loads a program in C/C++ generated from a gSOAP tool (Engelen, 2008) that allows a client to create a web service using SOAP protocol. This function makes available the recovered results as if they were tables in the database. This approach, although we must have installed PostgreSQL on a client with the access functions to each table, it allows the expert user to access the data from any application that connects to PostgreSQL, and allows the use of the language power of SQL to explore the database.

Some non-specialized users in web services and relational databases need the raw data or data views. For these users, data in spreadsheets are available, allowing them to make use of these data in statistical software among others, in some cases the dump file of the database can also be made available.

4. Metadata

So that the users can easilly access the data, users must rely on an interface that allows them to hold queries, either by keywords, subjects or by geographical region and find the information they seek. For this, a survey on geographic metadata manager software for the web that met the ISO 19115 standards. Among the projects the only free software that meets the requirements of having a friendly interface and being developed for the web was Geonetwork (GSDI, 2006), which has the possibility of doing research (harvesting) on data from other servers, in addition to providing Catalogue Service for the Web (CSW).

The documentation on the database of the forest inventory is available in the metadata catalogue of the Geonetwork, and consists of the entity-relationship diagram, data dictionary and links to web services. Also listed are the descriptions of spatial and non-spatial data, the link for the various available web services, WMS, WFS or the WSDL of the Web service. Although the Geonetwork is not a UDDI server (OASIS, 2004), which is a much more complex protocol of web service catalogues, it can catalogue SOAP web services through generic schemes of metadata. From the Gvsig one can connect to the metadata catalogue using CSW standard, and automatically add the web services available in the map view area of the program.

5. Interfaces

Currently we have several free software available for desktop with the capability to access web services[1], but the non-specialized users hardly have such software installed on their computers. For them to have access to those data a web interface would be made available. Among the various interfaces for webmapping available today[2], we chose to use the I3Geo, a Brazilian software based on the Mapserver software for interactive view of geographic data on the Internet developed with PHP and Javascript. The choice is due to the fact that it has an interface for the addition of web services such as WMS information layers and allows users to upload their own geographic files as well as allows users to save their projects and recover them in the next session and be able to generate some simple analyses, for instance, creation of buffer.

As an example of application using these data a web interface was developed that allows users to choose a specific species and visualize its spatial distribution and abundance showing the results either in I3Geo or in Google Earth, figure 4.

Interfaces between Postgis and GE through PHP scripts have been developed to produce well organized KML files according to user demands. To do this, a master KML document containing PHP scripts as <networklink> that produce and show KML layers when users require an information layer. Depending on the complexity of SQL queries, CPU consumption and frequency of data update these PHP KML layers are saved as KML or remain dynamically loaded by PHP scripts.

---

1 Among the most known free Desktop GIS currently we have Qgis, GvSig, OpenJump e uDig.

2 Openlayers, Mapguide, Mapbuilder e Cartoweb are some of the free software to present data on the web.

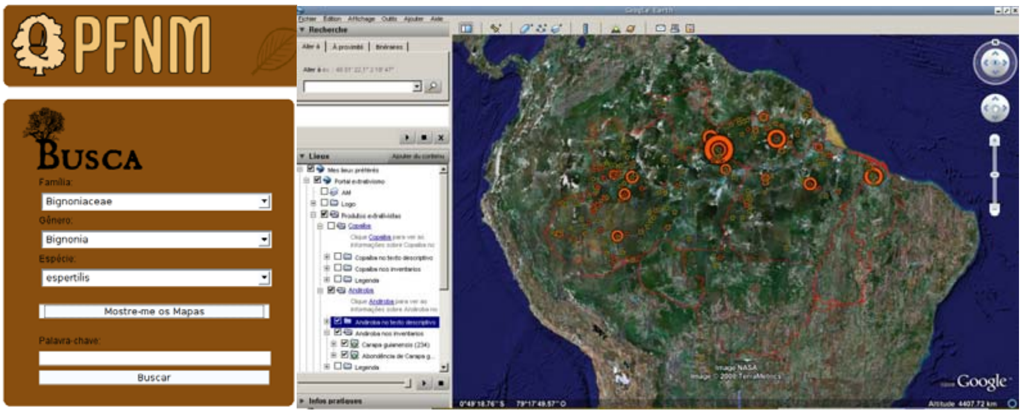

Figure 4. Interface for users to chose a specie and visualize its spatial distribution and abundance

6. Conclusion

Although the use of heterogeneous spatial data integration tools through ontologies (Azevedo *et. Al.*, 2006) has not been discussed in this paper, it is relevant to mention that it allows more complex queries using data from different sources.

Web services, SOAP and OGC combined with the power of Spatial Relational database as Postgres/Postgis offer efficient and free access for experimentation of new ways to manage and distribute spatial data.

If some features are not available in some applications, for instance, SOAP web service access from free GIS software, there is still the possibility of integrating data using database functions. With Postgres/Postgis as a background, interoperability gives the possibility to jump from one aplication to another to reach a specific objective.

Spatial biological data remain complex: metadata are essential to understand how field work and digitalization were done to know real biological meaning. Geomatics offers many possibilities to disseminate data but non-specialized users may easily be lost in huge amount of information.

The combination of free software, web services, geobrowsers and webmapping certainly is the door to the widespread of geographical information, allowing the society to better know and manage its natural resources.

7. References

Azevedo, VHM *et. Al.* 2006. 'Interoperabilidade entre Objetos Geográficos Heterogêneos'. *In: Geoinfo 2006*, viewed 20 september 2007, <http://www.geoinfo.info/geoinfo2006/papers/p50.pdf>.

Carnevalle, MT 2008. *Infra-Estrutura para Informações Espaciais*. UERJ, Rio de Janeiro.

Carroll, LN, Calvo RA 2004. 'Querying Data from Distributed Heterogeneous Database Systems through Web Services'. *In Proceedings of The Tenth Australian World Wide Web Conference (AUSWEB'04)*, viewed 7 april 2008, <http://ausweb.scu.edu.au/aw04/papers/refereed/carroll/paper.html>.

Desmoulière, SJM, Carneiro Filho, A 2007. 'Bases de dados digitais de árvores na amazônia: distribuição biogeográfica e planejamento da conservação'. *59th Anual Reunion of the SBPC*, viewed 15 March 2008, <http://www.servicos.sbpcnet.org.br/sbpc/59ra/senior/livroeletronico/resumos/R4225-1.html>.

Engelen, RV 2008. *gSOAP 2.7.10 User Guide*, viewed 22 april 2008, <http://www.cs.fsu.edu/~engelen/soapdoc2.html>.

FAO, WFP, UNEP and UN-OCHA GeoNetwork. Spatial Information Management Development 2006, GSDI, viewed 10 Jan 2007, <www.gsdi.org/SDIA/docs2006/jun06links/GeoNetwork-Opensource.pdf>

Fearnside, PM 2006. 'Deforestation in Amazonia: dynamics, impacts and control'. *Acta Amaz*, vol.36, no.3.

FeatureServer, 2008, FeatureServer -- RESTful Geographic Feature Storage, Metacarta, viewed 9 May 2008, <http://featureserver.org/>

Soares-Filho, BS *et. Al* 2005. 'Cenários de desmatamento para a Amazônia'. *Estud. av*. vol.19, no.54.

UDDI Version 3.0.2 2004. OASIS, viewes 15 Jun 2008, <http://www.oasis-open.org/committees/uddi-spec/doc/spec/v3/uddi-v3.0.2-20041019.htm>