Spatial data for scientific research in the Amazon – consideration from a user's point of view.

Silvana Amaral
Researcher – INPE. silvana@dpi.inpe.br

The Brazilian Amazon occupies 5 million km^2 and comprises the largest preserved contiguous tropical rain forest in the world: a forested area of approximately 4 million km^2. In the last decades, the region has experienced intense transformation processes in both physical and human aspects. Deforestation studies estimated a rate of 11,532 km^2 of forest conversion between 2006 and 2007[1]. Official census data estimated 14,623,316 habitants living in the states of the Brazilian north region in 2007, which comprises 7.9% of the Brazilian population, in contrast to 6.6% verified in 1970[2].

General policy based on sustainable development has to be adopted to counterbalance biodiversity conservation and needs of people living in the region. In order to discuss and monitoring the regional development, several research initiatives are focused in Amazonia. Biophysical or human sciences research requires and generates a significative amount of spatial data. However, making data available is not simple and different solutions usually imply in non-standardized and non-interchangeable databases.

This paper presents an overview of spatial data available for research in biodiversity and human dimensions for the Amazon region, considering different geographical scales and data access. From this general perspective, some suggestions are presented to improve data interchange and availability.

First of all, more than technical constraints or computational challenges, spatial data sharing in Brazil has a cultural problem – researchers are worried about making their data public. This is especially true for biodiversity research, where collecting primary data is very costly in time and money and data are, in general, very sensitive – related to the genetic potential of Brazilian natural resources.

Mapping agencies provide up to 1:250,000 scale data to anyone who intends to work in the Legal Amazon. However, as explained later, some effort is required to gather and integrate all data in a spatial database before any scientific research can actually begin. Information is spread out between several institutions, data format varies, and there are different interfaces for data access.

Basic geographical data, such as geopolitical limits, geology, geomorphology, soils, and vegetation are available at the Brazilian Institute for Geography and Statistics (IBGE) website[3]. Most of the information from surveys (Census) is available in table format, aggregated by some territorial limit such as census-sector or municipality. Some detailed

[1] http://www.obt.inpe.br/prodes/prodes_1988_2007.htm.
[2] http://www.ibge.gov.br/home/estatistica/populacao/censodem/tab202.shtm
[3] http://www.ibge.gov.br/home/

data are accessible only by associated institutions. Shapefiles of municipal or census-sectors limits can be downloaded from a list of files names, divided by municipalities or states. The census data have to be linked to the corresponding spatial information (municipality, census-sector, etc) by the user. IBGE provides maps for common people, students, and GIS users[4]. Unfortunately, some data still remain in pdf format (Thematic Maps), which one has to select and download from an ftp site, without previously knowing the geographical region of those maps. Planimetric maps can be accessed at scales ranging from 1:25,000 to 1:100,000, in different formats (pdf, dgn, shp, tiff).

Geographical data can also be displayed and browsed at an interactive map server[5] from where one can download data. These are the essential data sets, but it still necessary to do some work to gather all the data needed to create a particular database for a specific application or research.

Some institutions organize geographical data from different sources and make data available along with the respective metadata references and/or displaying the geographical data using map servers tools. The Ministry of the Environment (MMA) present different strategies, based on open source tools, for geographical data access[6]. The MMA-Geonetwork[7] allows to research and, knowing the metadata and the source, to request data from other institutions. They also provide data download (from an ftp file list), statistical data access, satellite images from a catalogue, and web services clients.

For environmental data and conservation strategies, a system used to integrate its branches, The SisCom, was developed by IBAMA[8]. Several environmental data can be downloaded, at different file format (shapefiles from an ftp list, or images from a catalogue). The Amazonian Protection System, SIPAM, also provides an interface where data sets from different institutions can be selected and downloaded[9].

Eventually, building those systems implicated in duplicate efforts to provide official data, especially considering updating needs and support to ensure integrity and continuity. Interactive maps providers were developed for IBGE, MMA and SIPAM.

For data sets acquired at more detailed scales, as from 1:250,000 to 1:2,000, geographical data are scarcer and usually collected by specific scientific projects or are provided by municipalities. Collected geographical data do exist, but accessing them is not easy.

Taking biodiversity data as an example, it is easier to get Brazilian plant information from the international databases, as from the New York Botanical Garden[10] or from the Missouri Botanical Garden[11] than from the Brazilian herbarium of the north region. In the Brazilian

[4] http://www.ibge.gov.br/mapas_ibge/.
[5] http://mapas.ibge.gov.br/divisao/viewer.htm
[6] http://www.mma.gov.br/index.php?ido=conteudo.monta&idEstrutura=41
[7] http://mapas.mma.gov.br/geonetwork/srv/br/main.home
[8] http://siscom.ibama.gov.br/
[9] http://www.sipam.gov.br/en/content/view/41/50/
[10] http://sciweb.nybg.org/science2/VirtualHerbarium.asp
[11] http://www.tropicos.org/

institutes of the Amazon, data are still being digitized, as in the state of Para (MPEG)[12], and there are sensitive data not published for general users, as in the state of Amazonia (INPA)[13].

Despite the general scientific production, data from surveys are hard to find, gather, and integrate. For studies of forest biomass and dynamic, the Amazon Forest Inventory Network proposes data sharing among collaborative researchers[14]. For other non-participant users, the data have to be requested to the researchers, following a protocol where there is a co-authoring commitment for any publication. Even data from surveys supported by the Brazilian government, as the Research Program in Biodiversity - PPBIO[15], are shared only among collaborative researchers, following the same distribution protocol.

Data used specifically in research activities in the Amazon region, at several scales, can be reached from the LBA project database[16]. Some data are available by ftp file list[17], and most of them can be reached with the metadata. For LBA participants the access is free, but for general users, the data "owner" has to be personally contacted. It is an important source of Amazonian data, especially for biogeochemical cycles and climatology data, at finer scales.

Besides these examples, several initiatives are studying and generating relevant data at detailed scale for the Amazon region. Non-governmental institutions, governmental projects, universities, and even municipal governmental secretaries, play very important roles and provide data for planning purposes. But these data are neither systematized nor easy to access.

A spatial data structure to study the Amazonia should first consider the amount and the diversity of data available. A general list (or a "catalogue") of geographical data available could centralize the data research. This list should inform the scale of mapping and the data type, addressing the data from the source institution. It should differentiate data visualization (download of figures, map servers or as web services clients), from data access (original geospatial data as grids, tiffs and shapefiles, or as WCS access).

For detailed scales, valuable geographical data are spread over projects and local initiatives. To make those data compatible with spatial structure, eventually some edition and integration with the appropriate spatial reference may be needed. Additionally, free tools to integrate and share the available data should be developed to help disseminate geographical data for such heterogeneous region as Amazonia.

[12] http://marte.museu-goeldi.br/herbario/
[13] http://angelim.inpa.gov.br/
[14] http://www.geog.leeds.ac.uk/projects/rainfor/
[15] http://ppbio.inpa.gov.br/Eng/dadosinvent/
[16] http://lba.cptec.inpe.br/beija-flor/
[17] https://daac.ornl.gov/new_user.shtml