

STATE OF THE ART OF THE BRAZILIAN GEOGRAPHICAL NAMES DATABASE

Graciosa Rainha Moreira
graciosa.moreira@ibge.gov.br

Instituto Brasileiro de Geografia e Estatística – IBGE
Brazilian Institute of Geography and Statistics
Computer Science Directorate
Coordination for Assistance and System Development
Geosciences Systems Management

Abstract

Since the establishment of mankind on earth, the social body always aimed to identify, by any available means, the most significant places, establishing its presence on the planet. All places with some relevance for humanity are known by a name that can only relate to it univocally. The geographical names hold a priceless heritage for a nation and must therefore be protected and divulged because they reveal patterns of occupation, identity and language diversity. The study of geographical names and their attributes contributes to the quality of cartographical information. The Brazilian Geographical Names Database (BNGB) begun with the establishment of a work group comprising several specialists of the Brazilian Geographical and Statistics Institute (IBGE), in the fields of cartography, geodesy, linguistics, geography, communication and computer science. The interdisciplinary work of this group made the creation of the Geographical Names Database possible with the main objective to provide support for the discussions and resolutions of the future authority in Geographical Names of Brazil, as well as, to allow access for the Brazilian community to the official geographical names with the proper spelling corrections and standardizations. The absence of standardization of geographical names entails, among other consequences, lack of bases for documentation and lawsuits in agrarian and territorial issues. This paper on the “State of the Art of the Brazilian Geographical Names Database” presents the principles that guide the development and implementation of the “Brazilian Geographical Names Database”, as well as an application developed for WEB research. BNGB considers the following aspects: geocartographical, historical – geographical and linguistical. The geocartographical aspect considers the name identification as a code, official geographical name (and variants), geographical coordinates, geometry, cartographical element classification, scale and documentation; the historical geocartographical aspect considers the territorial importance and historic inclusion, toponymic changes, historical name variants; while the linguistic aspect considers etymology, ethnolinguistics, orthography, toponymic motivation and grammatical classification. BNGB was established in PostgreSQL/Postgis and currently includes over 55,000 geographical names, derived from the digital version of the “Integrated Cartographic Base of Brazil” (in scale 1:1,000,000, versions 2.2.1 e 2.2/2007) and from the “Municipal Grid of Brazil

2005" with availability for consulting through the WEB, of the geocartographical aspect by use of the free software language PHP and MapServe. The approach to such aspects by BNGB makes it complete for supplying the Authority in Geographical Names but asserts again the need of creating a National Network of Geographical Names with the aim of enabling the inclusion of a much greater number of names, entailed by the linguistical, cultural and geographical diversity.

IBGE offers to the Brazilian community the "Brazilian Geographical Names Databank" (BNGB) to be used in the effort for standardization in the ambit of the National Cartographical System as well as externally, by establishing partnerships that will improve the National Toponymic Network, which will be supported by the expertise of its members in mapping science, geospatial information and data production, linguistics, philology and research in geosciences.

1. Introduction

Since the establishment of man on earth, the social body always aimed to identify, by any available means, the most significant places, thus establishing its presence on the planet. All places with some significance for humanity are known by a name that can only relate to it univocally. The geographical names hold a priceless heritage for a nation. They must therefore be protected and divulged because they reveal patterns of occupation, identity and language diversity.

The geographical names articulate concurrently: language, territorial policy and identity. To name and rename rivers, mountains, cities, neighborhoods and other geographical features includes a political and cultural meaning, involving ethnical or cultural groups, hegemonic or not.

Likewise, a geographical name is understood as the standardized toponym, including, in most of the occurrences, a specific name and a generic designation, added by attributes that characterize it as an ethnographic, etymologic, historical, geographically referenced group and inserted in a temporal context.

It is observed that the absence of standardization of geographical names generates, among other consequences, lack of bases for documentation and litigations in land occupation and territorial issues. And, also, that toponymy is a fundamental component in the composition of geospatial databases, in the structuring of the National Infrastructure of Space Data (INDE) and, mainly, for the quality of the benchmark mapping for reference of the Country.

The Brazilian Institute of Geography and Statistics (IBGE) resumed the Project of Brazilian Geographical Names according to UN recommendations, in 2005, having as main objective the modeling and implementation of the Brazilian Geographical Names Database (BNGB).

2. The Brazilian Geographical Names Database

The Brazilian Geographical Names Database began with the designation of a Work Group comprising a group of IBGE specialists, in Cartography, Geodesic, Literature, Geography, Communications and Information Technology. The interdisciplinary workgroup made the creation of BNGB possible, with main objective to support the discussions and resolutions of the future authority in Geographical Names of Brazil, as well as to allow the Brazilian society access to the official geographical names with the due orthographic corrections and standardizations.

The stage of survey and analysis of the requirements for the creation of BNGB signaled a heterogeneous environment in several aspects: nature and source of the data, users and technologies whose use is already consolidated in the institution.

As sources of data surveyed initially to compose the BNGB, the products of the IBGE Geosciences Management department were considered, as follows: Municipal Mapping, the Integrated Cartographic Base of Brazil in digital version and scale of 1:1,000,000; Topographical Mappings in the scales from 1:25,000 to 1:250,000, the products of Geographical Mappings; data originating from the Coordination of Natural Resources and from the Coordination of Geography of IBGE and, also, data collected in Censuses, data under responsibility of national and international organs, data produced in the teaching institutions and its researches, as well as literature on the theme.

The diversity of organs producing the information that feed BNGB also leads to a diversity of technologies used for the development and availability of that information. Thus, about 70% of the information was stored in the Access platform and in files in the shapefile and DGN format, and the remaining is distributed in Oracle, in electronic spreadsheets (Excel) and in analogical media. The diversity is also present in the application of Geographical Information Systems (SIG), with the use of ArcView, Arc/Info, MGE and Geomedia.

In the survey of requirements a need was observed for storing geo-referenced data in BNGB, defined under two aspects: the descriptive aspect, that allows the recording of the descriptive information concerning the objects in analysis and the alphanumeric representation; and the geometric aspects, that allow record the graphic representation (points, lines and polygons) of these objects.

For the BNGB implementation it was chosen to use an integrated architecture (Figure 1) that consists in recording, in a Database Management System (SGBD), the spatial and alphanumeric components (Casanova et al., 2005). This allows the use of the resources of an SGBD for control and manipulation of spatial objects, as transactions management, integrity control, concurrency and specific search languages.

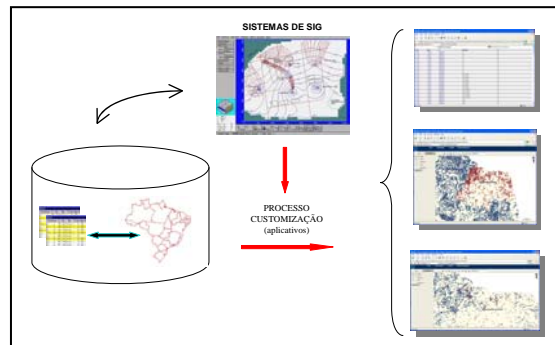


Figure 1. Integrated Architecture

Among the solutions for SGBDs that allow the implementation of the integrated architecture, stand out PostgreSQL and Oracle, which are object-oriented database management systems. The first is a free application with open code source, and the second is a commercial application, requiring license for its use. Both support spatial extensions for geographical data treatment in vector format. PostgreSQL requires PostGis extension and Oracle the Oracle Spatial extension (Casanova et al., 2005).

The conceptual project of BNGB began with the conception of the alphanumeric data model, being opted for the use of an Entity-Related Model for being a renown tool of easier use for alphanumeric BNGB modeling (Cougo, 1997).

The logical and physical specifications of the alphanumeric aspect of BNGB were initiated in mid 2005 and were based on the ORACLE platform for being the official SGBD used by IBGE. Oracle 9i Designer was the tool used for obtaining the Logical and Physical BNGB Models.

Concurrently, at the end of 2006, based on Federal guidelines based on the e-PING (Interoperability Standards of Electronic Government) for the use of free software in government institutions, the tests with the PostgreSQL/PostGis platform were initiated.

The implemented modeling sought to meet all the subjects indicated by the workgroup and resulted in an extensive model, that, for better understanding, is treated on three aspects: (1) Geocartographical aspects, that contemplate the identification of the geographical name as code, official geographical name (and variants), geographical coordinates, geometry and classification of the cartographic element; (2) Historical-geocartographical aspects that contemplate the history of the geographical names, a historical of territorial pertinence and cartographic “involution” of the geographical name and (3) Linguistic aspects, that contemplate etymology, ethnolinguistics, motivation, spelling, grammatical classification, dating and bibliographical reference of the geographical names.

To make the Brazilian Geographical Names Database model available to Latin America countries, the model initially elaborated was converted using the Oracle Designer tool of the DBDesigner 4 free software, available at www.fabforce.net. The use of this tool allowed obtaining a Logical Model (Figure 2), instead of a Conceptual Model. This fact justified the names of attributes following a standard adopted in the physical specification of a database.

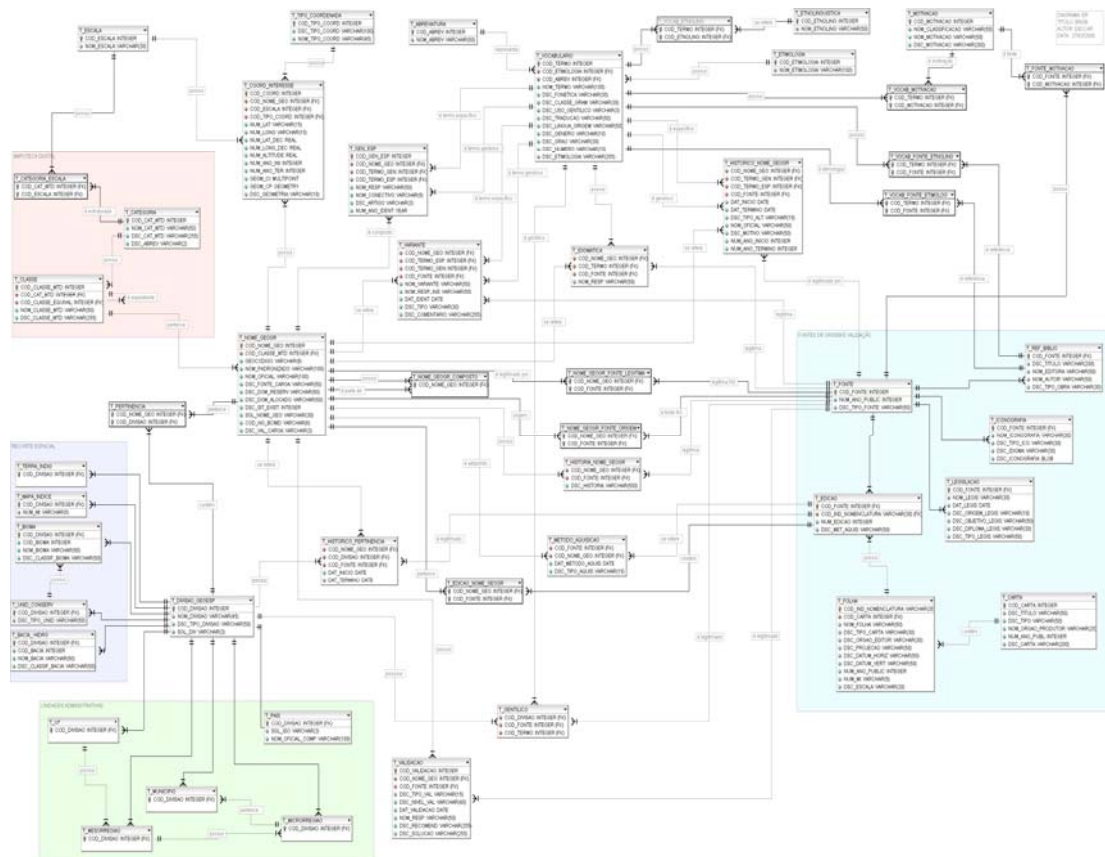


Figure 2. Conceptual BNGB Model

Considering that the data are produced by different IBGE departments and also by several external organs, the information generated were stored in separated databases in different formats, representing a great challenge for the process of reconciliation in the stage of data-base storing.

The feeding of data into BNGB cannot be considered as a trivial task, considering the aspects approached by the database, as well as the data sources, besides, of course, of the types of alphanumeric data and spatial data (point, line or polygon). Leaning on the expertise and on the wealth of IBGE Geosciences' Management, the initial BNGB loading is concentrated on the geocartographical aspects.

For each input, methodologies for acquisition of the geographical names are being elaborated, considering its geometry, according to the source of the data considering the geocartographical aspects: vectorial with or without database, old raster charts in the TIF format, printed lists – old gazetteers, academic works and data from Censuses.

At first the point elements of the Brazilian Integrated Cartographic Base were loaded in the digital version, in the scale 1:1,000,000 (versions 2.2.1 and 2.2 2007) and the Municipal Grid 2005, with respective geometries and, following, the load of the line and polygon elements, that were acquired based on the procedures presented below, according to the developed methodology:

- The main coordinates of the lineal elements should coincide with the medium point of the feature, and those of the elements with geometry of the polygon type should coincide with the centroid of the feature.
- For all the lineal or polygonal elements, coordinates of interest should be collected along the feature obtaining at least:
 - One pair of coordinates for each sector in the scale 1:25,000 through which the feature extends.
 - One pair of coordinates in each municipality through which the feature extends.

Presently BNGB is implemented in PostgreSQL/Postgis and holds over 55,000 geographical names, originated from the Integrated Cartographic Base of Brazil in the digital version, in the scale of 1:1,000,000 (versions 2.2.1 and 2.2 2007) and the Municipal Grid 2005.

A geographical names database has an ever growing gamut of applications, such as: identification of places by coordinates, measurement of distances between places, classification of names of a certain ethnic group, classification of elements under federal, state, or local administration, association of geographical elements by category, construction of gazetteers (index of geographical names with the correspondent location) by mapping scales or areas, elaboration of glossaries, automation of searches and consultations on the internet.

3. Consultation System

The system for consultation to the Brazilian Geographical Names Database is being implemented in PHP, a free-code programming language very much used to generate dynamic contents on the World Wide Web, and Mapserver - a free software much used in the construction of spatial applications for the internet, with visualization maps, images and vectorial data.

This applicative was developed considering the three main aspects approached firstly in the discussions on the geographical names that guided the construction of BNGB - the geocartographical, historical-geocartographical and linguistic aspects. According to the UN, these three aspects are very significant as support for the Authority in Geographical Names in the standardization task, because the first aspect locates the geographical name in the territory, the second locates the element in the temporal context where the tradition of the use of geographical names can be observed and the third aspect, which is the linguistic aspect, informs on the cultural context of this name.

For the geocartographical aspects (Figure 3) one can search by categories, coordinates and geographical name, in order to obtain the disaggregated information in the following levels: category of information (second Digital Topographical Map Collection), class and state. The categories of information are place, hydrography, hypsography, borders, transport system and general economical activities. The classes are subgroups of the categories and aggregate the elements according to their representative feature, as for instance, water course in hydrography and city in locality.



Figure 3. Disaggregated information selection by category, class and state

For each selected geographical name the following data will be informed: code, official geographical name, standardized geographical name, variants of the geographical names, geographical coordinates, pertinence to states and in the future per municipalities, special areas, hydrographic basins, among other geospatial sectors, geographical feature classification, scales of the representations, load sources and cartographic products where they are represented. This page also presents a picture related to the geographical name and its geospatial representation (Figure 4).

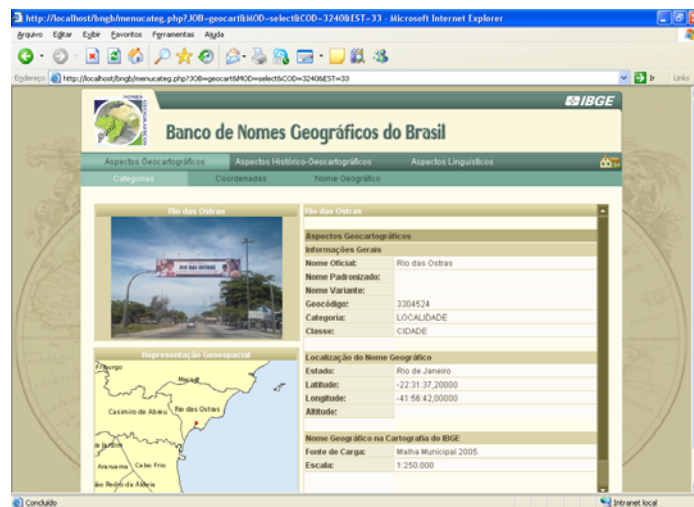


Figure 4. Window with a search result

For the historical-geocartographical aspects, the history of the geographical names can be consulted, historical variants (old names of the same feature) and historical of pertinence (dynamics of the territorial division). These information can also be disaggregated by category, class and state, as well as, by geographical name (Figure 5).

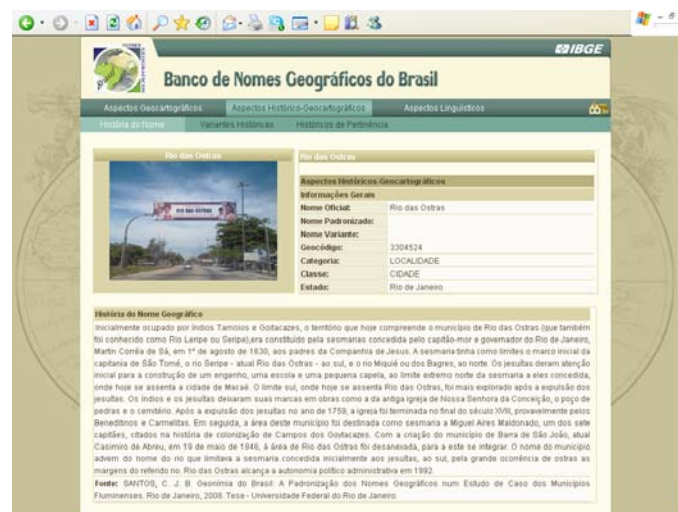


Figure 5. Result of a search of a geographic name history

In relation to the linguistic aspects, it may be stated that the studies and the surveys required for the perfect modeling of BNGB are still incipient regarding that context. Some changes will be required in the database modeling in order to allow the insertion of the etymology of the geographical names, as well as the ethno-linguistic aspects and the toponymic motivation, with the respective validation sources.

4. Conclusion

The Brazilian Geographical Names Database, due to its importance, will be fundamental for the evolution of research in geographical names in Brazil, because it contemplates the main information sources, associated legislation, validity date, motivation, etymology, historical, orthographic variants, historical and alternative names, indication of protected intellectual property, validations, positioning, relevant notes of its representation, among other attributes.

The approach of these aspects by BNGB makes it complete to provide support for the Authority in Geographical Names, however it reaffirms the need for creation of a National Net of Geographical Names in the sense of making the load of a significant number of registrations possible, considering the linguistic, cultural and geographical diversities of Brazil.

The Brazilian Institute of Geography and Statistics will offer the Brazilian Geographical Names Database to the Brazilian community to be used in standardization efforts, both in the internal sphere of the National Cartographic System as in the external sphere, establishing partnerships that will strengthen the National Toponymic Net, which will lean on the expertise of its components in mapping sciences, data production and geospatial, historic and linguistic information, as well as in Geosciences research.

References

- CASANOVA, M.A.; DAVIS, C.; VINHAS, L.; QUEIROZ, G.R.; CÂMARA, G. **Banco de Dados Geográficos**. São Paulo: MundoGEO, 2005.
- COUGO, Paulo. **Modelagem Conceitual e Projeto de Banco de Dados**. Rio de Janeiro: Campus, 1997.
- E-PING. **Padrões de Interoperabilidade de Governo Eletrônico**. Available at: <https://www.governoeletronico.gov.br/acoes-e-projetos/e-ping-padroes-de-interoperabilidade/versoes-do-documento-da-e-ping> . Access in: July /2009.
- MAPSERVER. **Documentation for the MapServer project**. Available at: http://ms.gis.umn.edu:8081/ms_plone/docs . Access in: July /2009.
- PEREIRA NETO, Álvaro. **PostgreSQL: Técnicas Avançadas: Versões Open Source: Soluções para Desenvolvedores e Administradores de Banco de Dados**. São Paulo: Editora Érica, 2003.
- PHP. **Manual de PHP**. Available at: http://br.php.net/manual/pt_BR/index.php . Access in: July/2009.
- POSTGIS. **Manual PostGIS**. Available at: <http://www.webgis.com.br/postgis/> . Access in: July/2009.
- POSTGRESQL. **PostgreSQL 8.4.0 Documentation**. Available at: <http://www.postgresql.org/docs/8.4/interactive/index.html> . Access in: July/2009.

SOARES, Wallace. **PHP 5: Conceitos, Programação e Integração com Banco de Dados**. São Paulo: Editora Érica, 2004.

UNITED NATIONS. **Technical reference manual for the standardization of geographical names**. United Nations Group of Experts on Geographical Names. New York: United Nations, 2007.