Brazil Paper for the

Second Preparatory Meeting of the Proposed
United Nations Committee of Experts on
Global Geographic Information Management

on

Data Integration

Introduction

The quick development of contemporary geospatial technologies such as satellite imagery, aerial photography, Global Navigation Satellite Systems (GNSS) e.g. the Global Positioning System (GPS), hand-held computers, and geographic information systems (GIS) has opened unprecedented opportunities to use geographic information. These tools are increasingly used in the context of disaster management, environmental monitoring, natural resources protection, land use, utility services, etc., as they facilitate quick data collection and advanced data analysis, and allow for a flexible and integrated approach to information sharing and dissemination based on a spatial framework. The impact of these developments is also felt on official statistics at all the stages of population and housing censuses, where the efficiency in the pre-enumeration, enumeration and post-enumeration phases has been largely improved through the use of geographic information tools.

As a consequence, it is becoming increasingly clear that the applications of these technologies are of strategic importance. This raises the challenge related to how to address the integration of various geospatial information, including those of statistical nature.

This paper outlines the challenges posed by the integration of geographic data with socio-economic and other development data. It covers the specific issue of the integration of statistical and geographic data and shows that statistical systems gain to be integrated with geographic information systems for the building of a geospatial infrastructure in support of census and other statistical activities. The paper illustrates this through the national experience of Brazil.

I. Integration of statistical and geographic information

Statistical data, including many development data, such as economic or health data, concern human activities that can be geographically referenced. Geography is indeed increasingly recognized as key to virtually all national statistics, providing the structure for collecting, processing, storing and aggregating the data. The close integration of geographic information in statistical applications yields large benefits to national statistical offices as it reduces the cost and time required to collect, compile and distribute information, and leads to a greater number of services and a much wider use of statistical information, thereby considerably increasing the return on investment in data collection.

Mapping is one of the most critical activities of a census. The accuracy of the delineation of enumeration areas and the quality of their representation on a map have a crucial impact on the quality of the data collected. The increasing use of hand-held devices with GPS, and low-cost aerial and satellite imagery for spatial data collection and demarcation of statistical enumeration areas, as well as GIS for display of census information, have improved census mapping in fundamental ways. Geographic data are now more easily collected, disseminated, accessed and manipulated by multiple providers and users. The three main operational phases of any National Statistics Office (NSO) can be supported by GIS: integrating field data collection; processing of statistical data, disseminating data and supporting statistical surveys using maps that can be made available through the web for wide public access (Figure 1).

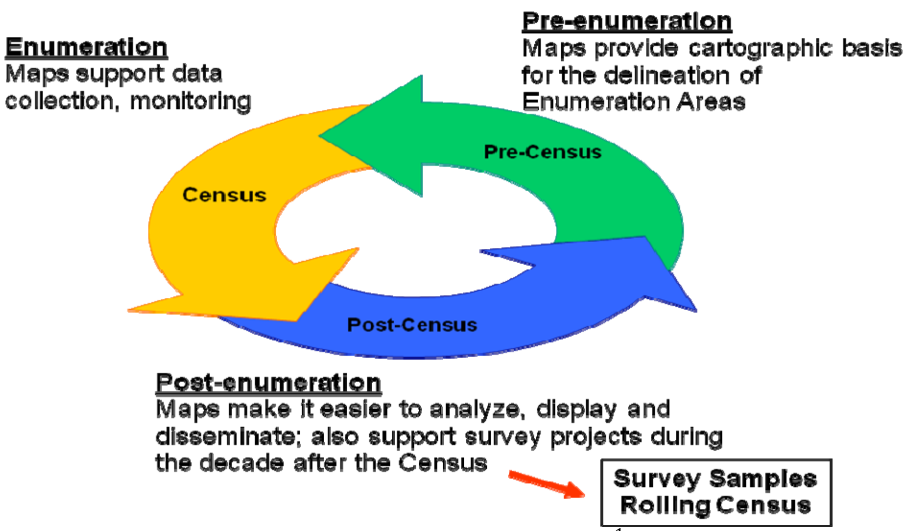

Figure 1 - Census Cycle[1]

The rapid integration of geospatial data with a variety of other data, including demographic and socio-economic data, and their analysis and modeling has increased the understanding of the dynamics of socio-economic and demographic structures and helped create more accurate, timely, and unbiased information for better decision-making. For example, this integration has proven to be critical to achieve improved operational readiness and responsiveness to disasters. By using satellite imaging, scientists and demographers can compare images and statistics taken before and after earthquakes to estimate the amount of aid to be allocated to populated areas. There are many such examples related to an increased use of geospatial data in socio-economic, demographic, and environmental analysis.

Institutional arrangements vary widely from country to country. In many cases, the collection of statistical data and the collection of spatial data, including the thematic map production are carried out by different organizations. This often results in lack of common standards and consequently incompatibility of data and in other quality problems, such as

[1] United Nations Statistics Division, 2007. *Digital Census Mapping Process: conceptual framework and different approaches*, United Nations Expert Group Meeting on Contemporary Practices in Census Mapping and Use of Geographical Information Systems 29 May - 1 June 2007 United Nations, New York

incomplete documentation of data. A close linkage between geography and statistics, of course, exists, in those countries, where the national statistical and mapping agencies are institutionally integrated under the same roof: This is precisely the case in Brazil and Mexico.

Independently of the institutional arrangements, countries are increasingly discovering that they can leverage the strengths of national statistical and mapping agencies through what it is referred to as a National Spatial Data Infrastructure (NSDI)[2], an institutional arrangement to permit data sharing and collaboration across government at a variety of levels including national, regional, and local. For example, foundational data layers such as demography and administrative boundaries that are produced by the NSO can be shared among many users, eliminating the cost of duplicative efforts. These two layers are important components of any National Spatial Data Infrastructure.

However, for many developing countries, SDI is still at an incipient stage of development. There is also a lack of public policies at different levels in relation to geographic as well as statistical information for development. It is therefore critical to sensitize politicians and policymakers about the usefulness and applications of spatial data coupled with statistical data.

II. National Spatial Data Infrastructure in Brazil

Geospatial data and information (GI) are largely produced, maintained and acquired by public sector organizations of all Brazilian governmental spheres. Notwithstanding, it remains difficult to the user to find what geospatial data sets are available out there, what their features are, where they can be found, who are their maintainers, and how one can access and eventually get them. In other words: GI do exist in large volumes in Brazil´s governmental organizations, but can hardly be found or accessed even by decision-makers of the public sector. Hence, the very first requirement for a Spatial Data Infrastructure (SDI) to be established in Brazil is that geospatial metadata should be made fully available for the existing GI assets or collections maintained by the public sector organizations.

The Presidential Decree 6666 of Nov 27th, 2008 has established the legal framework for the Brazilian NSDI´s enterprise, which has been named “INDE” (acronym for NSDI in Portuguese). More details can be found in http://unstats.un.org/unsd/geoinfo/9th-UNRCC-A/IP/IP 2013 Brazil SDI.pdf.

The launch of the INDE was carried out on April 8th, 2010, when the Brazilian Portal of Geospatial Data, also known as “SIG Brasil”, started to be accessible on http://www.inde.gov.br.

---

[2] NSDI is a combination of technology, policies, standards and human resources necessary to acquire, process, store, distribute, and improve utilization of geospatial data. (from GSDI Cookbook, 2000).

III. Brazilian practical experiences of the integration of geography and statistics

Integrating statistical data into NSDI opens new horizons, based on the possibility of correlating this data with all other data layers, like the ones related to natural resources and the environment. The new web-based technologies allow NSOs to produce Census Mapping in a full digital way, integrating maps, enumeration areas boundaries graphic and text files, as well as addresses files. The current availability of low cost high resolution orbital and aerial imagery can help updating the Census Maps.

IBGE has already been preparing the 2010 Demographic Census operation. The census maps, which will be used to plan and direct its operations, are being prepared with intensive use of satellite images. Therefore, they will constitute a rich collection for the construction of NSDI. Much has been done in the previous years to adapt internal processes to absorb the new technologies available and to improve the data collection, the monitoring support, and the results dissemination of the census. An important advance was achieved in the 2007 Census, when 80,000 Personal Digital Assistant devices (PDAs) were used to ensure the quality control at the time of survey and the real time monitoring of data collection in all municipalities. This procedure allowed that rural establishments (farms, schools, etc.) could be georreferenced, as PDAs were equipped with GPS. For 2007 Census, over 70,000 maps of census enumeration areas were converted to the new SIRGAS2000 geodetic system and, for 2010 census, it is expected that this figure will reach the value of 280,000 maps of census enumeration areas.

Another fundamental effort initiated in the scope of the 2007 Census operation was the preparation of the National Register of Addresses. For 2010 a further step is being taken, based on linking address records to block faces of the digital census maps. This will improve data collection, allowing the 2010 census interviewers to tap on each block face on the PDA screen to have access to addresses and corresponding questionnaires of that specific block face. In addition, this will increase data dissemination possibilities based on the association of census data to different portions of the territory. Figure 2 illustrates these new dissemination possibilities.

Figure 2: New ways of data dissemination

Another example of integrating geospatial information with statistics in Brazil is related to the Amazon deforestation monitoring. Due to its continental dimensions, it is not possible to measure the impacts of human actions in the Amazon without using satellite imagery. The Brazilian government has two permanent systems to monitor the deforestation in this region, named DETER and PRODES, developed by the National Institute for Space Research (INPE) (more information on http://www.obt.inpe.br/prodes/ and http://www.obt.inpe.br/deter/, respectively). The deforestation data can be combined with forest inventories and floristic survey, using geographic information systems, to provide indirect estimates of volume of wood extracted from the forest in a given area/time. An example of these estimates is shown for an Amazon municipality named Tailândia between 2001 and 2006 (Figure 4). PRODES informs that during this period of time an area of 633 Km^2 was deforested. Eight sample points of forest inventory near Tailândia with the same type of vegetation indicated a mean of 262.37 m^3/ha of wood in the area (Figure 5). By a simple operation the volume of wood extracted from this municipality can be computed as more than 16 million m^3 between 2001 and 2006. This kind of estimates could be very useful to build forest accounts.

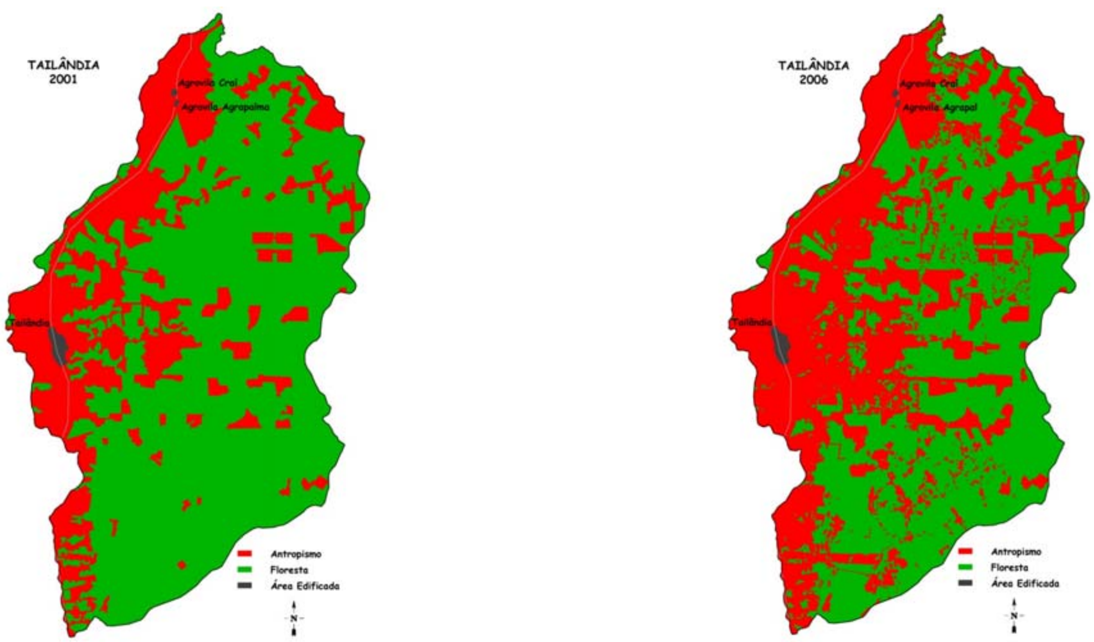

Figure 4 – Deforested area in Tailândia municipality from 2001 to 2006

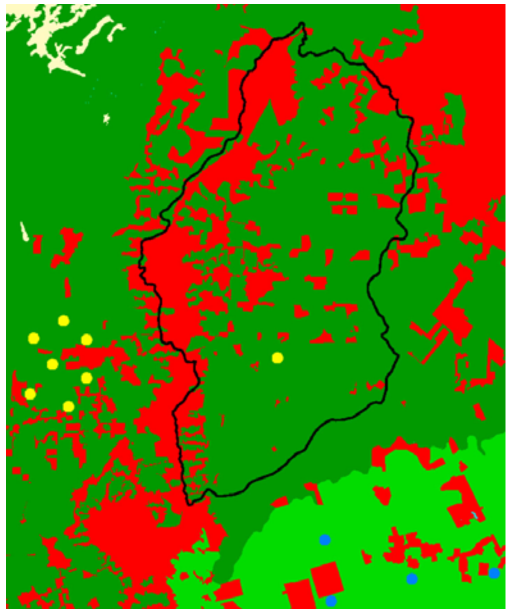

Figure 5 – Points of forest inventory sample near Tailândia municipality

With the availability of INDE – the Brazilian NSDI, the possibilities of integrating and analyzing many geospatial data is countless. Figure 6 gives an example of visualizing railroads from IBGE's 1:1,000,000 digital topographic database together with mineral resources extraction points from the Brazilian Geological Service (CPRM) database. The databases are physically situated in different locations and the SDI technology allows accessing and merging them in real time.

IV. Conclusions

There is an increasing recognition that geography and statistics are interconnected and spatial distribution of social, economic and environmental indicators guide policy decisions on regional development, service provision, resource planning and many other areas. As an additional example of the benefits of such integration, it can be mentioned the possibility of correlating households to territorial areas in agricultural and rural statistics. The experience of Brazil has shown the capabilities and potential offered by the geospatial technologies to support and improve census mapping operations, and underscored the fact that geography is central to census activities. It also shows that implementing geographic databases to support the whole census process was facilitated by the fact that geography and statistics activities were managed under the same institutional framework. For many other countries, national practices in building up a robust basis for census geography are relying on symbiotic efforts within the national statistical authorities and outsourced contractors, as well as of partnership with national geographic authorities.

Based on what has been presented in this paper, the establishment of Spatial Data Infrastructures represents the basis for integrating various types of geospatial data and information, which can physically be stored in different institutions across a country, a region or even the globe.

Considering the various actors that need to be involved, institutional collaboration is crucial for the success of Spatial Data Infrastructure enterprises. The resulting data integration maximizes the value of the information, improving the quality of the decision making.

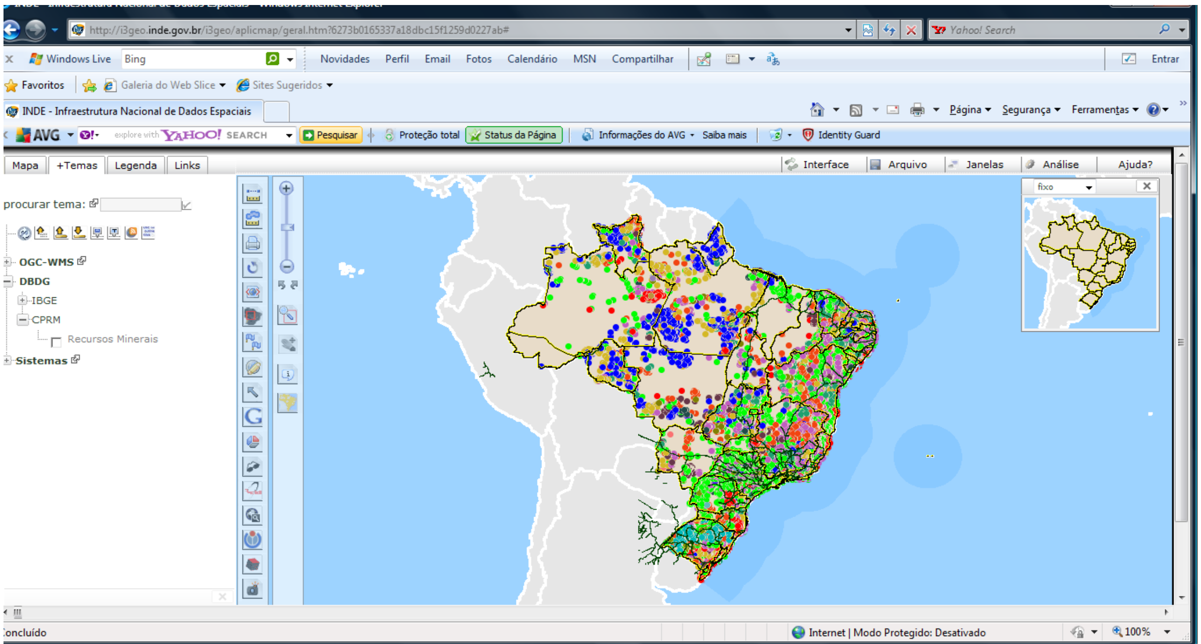

Figure 6 – Visualizing geospatial data layers using the Brazilian Portal of Geospatial Data. Railroads are shown in green and color dots represent mineral resources extraction points