The 2010 Brazilian Population Census: innovations and impacts in data collection

Bianchini, Zélia

Brazilian Institute of Geography and Statistics, Survey Directorate
Av. República do Chile 500 - 10 floor
Rio de Janeiro, RJ, 20031-170, Brazil
E-mail: zelia.bianchini@ibge.gov.br

Introduction

The 2010 Brazilian Population Census produces information on the main characteristics of people and housing units for each of Brazilian municipalities. Population Census is the main challenge for a Statistical Office, specially in a country like Brazil, with 8,515,692 km^2 distributed in a heterogeneous and sometimes inaccessible territory, made of 27 Federation Units and 5,565 Municipalities, encompassing approximately 67 million housing units.

The 2010 Brazilian Census introduced several innovations in the data collection process. A digital Census mapping was developed and integrated with the National Address File for Statistical Purposes. In addition, the computer assisted personal interviews was carried out using handheld computers that are equipped with a Global Positioning System (GPS) receiver and an option for online data collection via Internet was also available. The use of handheld computers allowed the georeferencing of all units visited in the rural areas as well as the monitoring of the field operation geographic coverage.

This paper describes the innovations and highlights their positive impact on the required quality, the timeliness, and the relevance of the Census results, providing the users a more interactive and space-related data.

2010 Census Mapping

To attain the required quality of a Census, it is essential to guarantee proper coverage of the whole territory. Therefore, the planning of the 2010 Census operation started with the development of a Census Mapping Project that built an integrated Territorial Database to support the 2010 Census operation, from the planning to the collection and dissemination steps. A single digital Census mapping was created, integrating urban and rural areas of the municipalities and moved from the hybrid analog/digital format to a full digital format. Geometrical adjustment of urban and rural limits was done for each Municipality.

The country was divided into 316,574 enumeration areas (EA). All of these areas have a cartography geocoding. For each EA, an enumerator was designated to visit households and interview residents.

The enumeration area is the territorial unit delimited for the control of registered addresses. It is a continuous area and it is either urban or rural, with a specific area and a given number of households to be surveyed.

The National Address File for Statistical Purposes (CNEFE) was developed to be statistical frame to support data collection procedures in statistical operations. This work started in 2005 based on records of adresses of all Brazilian municipalities produced in the 2000 Population Census, and it was updated during the operation of the 2007 Censuses[1].

An important milestone was achieved during the 2007 Censuses, when 80,000 handheld computers were used in data collection and over 70,000 maps of census enumeration areas were converted to the new Geocentric Reference System of the Americas (SIRGAS2000). Another fundamental effort that was also initiated in 2007 Censuses operation was the preparation of

---

[1] The integrated operation of the 2007 Censuses encompasses the 2006 Census of Agriculture, the 2007 Population Count and the National Address File for Statistical Purposes (CNEFE).

CNEFE. For 2010, a further step was taken, to link the address records to block on digital census mapping. This step forward allowed the 2010 census enumerators to tap on each block on the handheld computer device screen so as to gain access to the addresses and corresponding questionnaires of that specific block.

The enhancements implemented in the 2010 Census Mapping enabled the construction of a geospatial database containing digital maps associated to the CNEFE file to the blocks on urban enumeration areas. As a result, the office extended the Census dissemination profile based on association census data with different portions of the territory always taking into account disclosure restrictions.

The Brazilian System for Census Mapping (SISMAP) was developed for municipal mapping in a single continuous spatial database. The input data were from several sources of vector and imagery data, like Geographic Information System (GIS), Global Positioning System (GPS), satellite imagery, digital and aerial photography.

Partnerships

In order to conduct an operation of the magnitude of the Population Census it was necessary to join efforts of external partners from private and public spheres.

An Advisory Commision composed of 12 members selected among renowned subject matter and methodologists was established. In addition to extensive consultation with the society, it is also worth mention the creation of the State Census Commissions (SCCs), the Municipal Commissions of Geography and Statistics (CMGEs) and some Census Local Commissions (CCLS). The creation of these commissions aimed to establish a channel of communication and dialogue between IBGE and local administrators, giving greater transparency to the work of the 2010 Census.

The Commisions (CMGEs) had as their main tasks the monitoring the Census Mapping update and later the opportunity to support and monitor the data collection operation, taking note of the operational planning of the Census Mapping, the municipal boundaries practiced and getting the first results shortly after data collection had been completed.

Pre-Census Operation

The Pre-Census operation was conducted from March to June 2010. Over 20 thousand supervisors, hired for the Census, listed addresses of residential and non-residential units of urban areas, and collected information on the characteristics of census blocks.

The Pre-Census operation was carried out in 224,402 urban Enumeration Areas (EA). The objective was to review the area limits, loading the maps, the list of streets and associated blocks, updating digital maps with handheld computer and associating addresses to digital maps, allowing the enumerators to get in advance the addresses they would have to visit during the data collection operation. The system was loaded in 30,000 handheld computers. It was developed in Windows Mobile, using SQLServer CE 2005, Framework 2.0 and Geopad.

The work unit was the block and the supervisors' duties were:
- For each street: confirm, delete, include, change the name and/or change name's spelling;
- For each block: include, exclude, confirm adresses or classify as "NAR" (no address in the block) and collect local infrastructure characteristics (existence of public lighting, paving, storm drains, landfills, hazardous waste landfill, open sewers, curb, sidewalk, wheelchair access);
- To track with GPS the shape of blocks not represented on the map.

Thematic Content

The Brazilian Population Censuses has used two questionnaires since 1960: a short form and a long form. The latter is administered to be applied a sample of occupied housing units or collective living quarters with residents. All the variables of the short form questionnaire are also included in the long form.

The definition of the thematic content investigated in the 2010 Census was defined in the planning phase, after a wide consultation with several representative segments of the society.

Several studies, tests and two pilot surveys were conducted to outline the proposed questionnaires for the Census Rehearsal. Based on the collection operation in the Census Rehearsal, some adjustments to the questionnaire content were still necessary.

The selection of topics to be investigated in the 2010 Census took into account several factors: consultations with Census users, review of traditionally investigated topics, reassessment of the need for keeping data series, evaluation of new data requirements, the existence of available alternatives data sources and also the international recommendations.

Fundamental criteria were also taken into consideration, such as relevance and applicability of the investigated variables. Priority was assigned to information demanded for the development of: population estimates and projections; local area public policies; rare subpopulations of political interest; and structural information of public interest.

For the 2010 Census, the short form had 24 questions for each housing unit and 13 for each person of the household, whereas the long form had 38 questions per housing unit and 69 for each person (including those questions of the short form). The long form questionnaire contained[2]:

- Housing characteristics: **species** (permanent private, improvised, collective living quarter), type (house, apartment, room); **ownership** (own, lent or rented), materials used for outside walls, number of rooms, sleeping rooms, **number of bathrooms**, **existence of piped water, public services** (sewage, water, waste, electricity), existence of durable goods (radio, refrigerator, washing machine, mobile and fixed phone, motorcycle, personal computer, Internet access, television sets and cars).
- Population general characteristics: **sex**, **age**, religion, **color or race**, **indigenous (ethnic group and language)**, **birth certificate**, nuptiality (type of union, marital status).
- Household composition: **shared responsibility**, **relationship with household reference person**, mother alive, identification of spouse or partner.
- Internal and international migration, **international emigration**.
- Fertility and **mortality** related questions.
- Disability (seeing, hearing, mobility, mental or intelectual).
- Education (**literacy**, school or nursery attendance, public or private school, education level, kind of college degree, masters or doctorate.
- Labour market status and **income**.
- Commuting to study and to work.

Information and Communication Technology (ICT) for 2010 Census

Data collection was carried out using handheld computers in a face to face mode interview (CAPI) and concurrently the National Address File for Statistical Purposes (CNEFE) was updated. Two types of handhelds were used: LG-750Q smartphones locked (Phone, 3G) used solely as data collection equipment and programs certified by the IBGE or MIO-P550B, already used in 2007 Censuses.

IBGE staff was responsible for the development the Information and Communication Technology (ICT) for 2010 Census. There was no paper questionnaire. The handheld computer ran system to collect and store the information. Each handheld computer was equiped with Enumeration Area Map; list of addresses developed during pre-census operation (for the EA that this operation was done); list of collective living quarters; short and long form questionnaires; GPS; and summary information collected to facilitate monitoring enumerator's work.

Regarding the technological advances, the experience obtained by IBGE with the 2007 Censuses, particularly in the use of handheld computer equipped with GPS for data collection, was extremely important for the adequate planning of the 2010 Population Census. The following table presents some figures related to the 2010 Census operation.

---

[2] In bold typeface are the questions that are included on the short form.

Table – Some figures related to the 2010 Census operation

Description	Figures
Handheld computers for data collection and supervision	**220,000**
LG-750Q (smartphones locked - Phone, 3G)	150,000
MIO-P550B (already used in 2007)	70,000
Laptops	8,700
Printers	7,000
Regional Coordination	
State Units	27
Operational Area Coordinators for each subject (technical; operational; administrative and ICT)	220
Subarea Coordinators	1,281
Data Collection Stations (spread over 5,565 municipalities)	7,000

Each Data Collection Station was equipped with notebooks (1 per 5 supervisors) to operate autonomously, connected or not, to the Internet. The interaction of handhelds was made solely on the notebooks of the Data Collection Station, regardless of a network of synchronous communication with the Central Processing. In Data Collection Stations without Internet connection, communication with the Central Processing was done through USB drives (flash drives) physically carrying lots of information to some other point of Internet access. The 220 ICT Area Coordinators provided all ICT technical assistance (software, hardware and communication). One printer and WiFi router were installed in each Data Collection Station.

The 7,000 Collection Stations were supported and supervised by 1,281 Subarea Coordinators. One of the duties of these coordinators was to visit Local Offices under their supervision to provide quality control of collected data and organization. Each Subarea was equipped with one notebook per coordinator, one desktop computer, one router, one multifunctional printer, broadband Internet and one 3G Modem per coordinator.

The Management System for Data Collection Station (SIGPC) was run on the notebook in order to support all decentralized operations of administrative and operational tasks and the communication with the Central System. This system was essential for the registration of temporary enumerators, to associate enumerator's workload, to load the program and Enumeration Area data into handheld computer, to download collected data from the handheld computer, to transfer data to Central Processing and to summarize the quantitative of data collected.

Another important task in the 2010 Census operation was the distribution and updating ICT subjects. This task involved the planning and controling ICT logistics distribution of supplies, the distribution of supplies itself (laptops, printers, wireless routers and handhelds) and the maintenance and operational control.

The distribution of supplies was done in two-phases: in the Pre-Census operation for 224,000 enumeration areas using 10 thousand flash drives; and in the data collection for 316,000 enumeration areas using 1,500 flash drives and website download. To perform backups of local database 28,500 flash drivers were distributed for the maintenance and operational control. Management reports were prepared to control versions, updates and data transmission to the Central Processing. Website was used to download, to update and to perform developed program distribution.

The communication resources were the following:
- 2 links of 155 Mb/s for the exclusive use for the Census 2010 in Rio de Janeiro (Headquarter); 2 links of 20 Mb/s in São Paulo and Minas Gerais; and 2 links of 12 Mb/s in Bahia and Rio Grande do Sul.
- Subarea Coordination and Data Census Collection Stations: 3,783 mini modems 2G/3G; 139

antennas VSAT (Hughes); 40 antennas BGAN (Tesacom); 4,000 ADSL connections (leased, or by courtesy of others); 1,000 Radio connections; and 300 other types of connections.

Data Collection

The Census operation involved about 230 thousand people all over the country, including those hired on a temporary basis by means of selection processes, and employees from the permanent staff of IBGE. In order to fulfill this enormous demand, professionals who were taking part in the 2010 Census participated in a specific training chain, from a small group of instructors to the thousand of enumerators.

Advances aimed at the production of more effective results were made in the 2010 Census training methodology, and the main change was the inclusion of distance learning in addition to traditional methods adopted in such programs.

Among the several teaching resources available for in-class training, a highlight was the use of video classes, which guaranteed uniformity in the transmission of concepts and procedures during the several steps of the training process.

Trainees made extensive use of handheld computers, being able to develop skills related to use of such equipment as a tool. Besides, all data collection units in the 2010 Census had digitized processing, which eventually contributed to the digital inclusion of thousands of enumerators and Census agents, since all of them had access to microcomputers and handheld computers.

The data collection was conducted from August until October 2010. All 56.5 million occupied housing units were surveyed to collect characteristics of the dwellings and people living there. All households were visited and any resident capable of answering the questionnaire could respond on behalf of all the residents of that household. All people that usually lived in the household on the reference date, August 1st, 2010, should have been enumerated.

Enumerators were wearing a waistcoat, a badge and were carrying a handheld computer to collect information by means of a face-to-face interview with questions listed in the questionnaire form, selected to be applied in each occupied housing unit. Enumerators to collect data used about 190,000 handheld computers equipped with GPS receivers.

Replacing the paper questionnaire by electronic model developed for the handheld computer allowed the introduction of editing routines that were implemented at the exact moment of completing the questionnaire, thus contributing to the quality assurance process in data collection.

The use of handheld computers equipped with a GPS receiver allowed the georeferencing of all units visited in the rural areas as well as the monitoring of the field operation conducted by the enumerators.

The choice of which questionnaire (short or long form) for each household was made by random selection during the data collection operation, according to the sampling fraction defined for each municipality.

In order to provide more precise information for small municipalities, the 2010 Census applied five different sampling fractions for the occupied housing units, according to the number of inhabitants estimated of each municipality. The sampling fractions varied from 5% (municipalities with more then 500,000 inhabitants) to 50% (less then 2,500 inhabitants). The overall average sampling fraction was 11% (6.2 millions of occupied housing units and 20.6 million people filled in the long form questionnaire).

After completing an interview, data were encrypted and could only be transferred to the IBGE through a secure network.

Online Data Collection

The Online Data Collection was developed as a complementary system, which was another option for gathering information, especially in households where it was difficult to contact the residents at a reasonable time, since many people work outside the house all day long.

The option of answering the questionnaire via the Internet was given especially to those

households, that, after being visited by the enumerator, did not respond directly to the enumerator at that time. In this case, an envelope containing e-tickets (access codes) was delivered to allow accessing the questionnaire (long or short form) on a specific site. The site was security-certified by the certifying authority, guaranteeing the encryption and confidentiality of information provided.

The reply system via the Internet permited the 2010 Census questionnaires to be answered through the web. In order to suit the resident's time availability to complete the questionnaire, it could be answered in parts, because the system automatically saved the answers. Therefore, the resident could interrupt the task of completing the form whenever he/she desired and resume it afterwards, at a more adequate moment, within a pre-determinated deadline.

If the household that had chosen to respond via Internet did not fill in the questionnaire until the pre-determined deadline, the monitoring system detected the case and then the census enumerator had to return to try once more to conduct the interview. In the 2010 Census, approximately 35,000 questionnaires (102,000 person) were collected via Internet.

Collection Monitoring and Supervision

The Data Collection Management Indicators System (SIGC) was used for monitoring of fieldwork progress through processing the information transmitted by the data collection stations across the Management System for Data Collection Station (SIGPC). Field information was transmited to a system of indicators, generating reports with data collection status in the enumeration areas and the quantities of households and people enumerated. Thus, the information could be aggregated at different management levels of the Census (Enumeration Area, Data Collection Station, municipal, regional, state and national).

Real time analysis enabled decision for related adjustments on operational aspects making of collection process. Web application running in the main Datacenter created management reports, summaries, indicators and cartograms.

During the data collection the following management indicators were assessed: people counted per day; average of people per housing unit; rate of men on the total counted; number of occupied private housing units; rate of occupied private housing units in the sample; difference between average of people in housing unit in and out of the sample; rate of people who informed the month and year of birth was informed; rate of people under 3 and over 69 years old on the total of counted people.

On the other hand, for those enumeration areas in which the data collection was completed, the following additional management indicators were assessed: occupied private housing units; number of units registered (residential or not); rate of occupied private housing unit without interview and vacant; rate of dwellings of "occasional use"; results of supervision: interviews, units confirmed, inclusions and omissions.

The SIGC performed immediate editing through these indicators that identified possible coverage failures. This system also indicated automatically the completion of the collection operation in the municipalities and the enumerator's payment autorization, producing preliminary results as soon as the collection was finished in each municipality.

A supervision system was developed with a set of automated procedures, providing a tool for evaluating the enumerators' work quality. This task was performed by the Supervising Census Agent, responsible for checking during the collection period, possible omissions or inappropriate inclusion of people, or erroneous classifications of residential and non-residential units, and also for assessing the correct use of concepts and definitions of the questionnaires' content by the enumerators.

The introduction of new technologies in the 2010 Census brought significant gains to the supervision step. A software was developed to run in the supervisor's collection equipment with a set of automated procedures. This software was periodically fed by collection-produced information.

The programmed supervision automatic selection of housing units was run on the handheld computer. The supervisor did up to 3 requests per Enumeration Area, depending on the type of the enumeration area (1 to 3 in urban area and one at the end of data collection in rural area), the management indicators and the results of the previous visits. Each request complied:

check the path and inclusion of all units on the way on a sample of units; check the classification of the unities according to the type (occupied, vacant, unoccupied, etc.); reinterview people in up to 3 dwellings; interview people in special form (selected questions: name, sex, age, migration, literacy and school frequency, employment, live births, mortality).

A complementary supervision also was done. Through the access code on SIGPC, the enumeration area was chosen and data collected by the enumerator printed out: address, type of unit, type of questionnaires (long or short form) and people by sex. The supervisor returned to the area to check the information.

The data collected by the supervisor were compared with the data collected by the enumerator. In case of divergence, the information was corrected and added on the handheld computer allocated to the enumerator.

It should be noted that during the data collection the enumerated population in the country by state and municipality was daily updated on the 2010 Census website.

Final remarks

The main objectives that guided the planning of the 2010 Brazilian Population Census include the pursuit of excellence of the coverage, ie, that all persons residing in the country at the reference date were investigated, and the data quality.

The experience of the integrated operation, conducted in 2007, the Agricultural Census, Population Count and National Address File for Statistics Purposes (CNEFE), was fundamental for the success achieved in the 2010 Census, particularly regarding the use of handheld computers equipped with GPS receiver for data collection and computerization of about 7000 collection stations installed to manage the 2010 Census in all 5,565 Brazilian municipalities.

The Census 2010 counted just over 67.5 million housing units, which 56.5 million were occupied. The Brazil's population reached the milestone of 190,755,799 inhabitants.

The technological innovations implemented in the 2010 Census significantly affected the organization of the processes involved in this gigantic operation. Besides, the updating of the National Address File for Statistical Purposes - CNEFE and its integration with cartographic information and with the elaborated digital territorial base will have a substantial effect on the activities of the Brazilian statistical and geographic information systems, mainly, the system of household surveys, that is currently under revision.

Another important point to highlight is that the technological and methodological innovations introduced in the 2010 Census have a significant influence and constitute a step forward on the institute's future activities, through the increase of analytical possibilities, provided by georeferencing and geospatial data visualisation features, as well as the unlimited possibilities of the graphical interface.

REFERENCES

BEVILAQUA, J.S. and NUNES, E.P. Information and Communication Technologies for the Brazilian 2010 Census. Presented at Forum Unesco / Netexplorateur 2011. Paris, February 2011.

CONFERENCE of EUROPEAN STATISTICIANS, 1., 2006, Geneva. Recommendations for the 2010 censuses of population and housing. Geneva: United Nations, 2006. 200 p. Available at: <http://webrzs.stat.gov.rs/axd/POPIS2011/preporukeE.pdf>.

IBGE. 2007 Censuses. Innovations and impacts on the Brazilian statistical and geographic information systems. Rio de Janeiro, 2008. Available at: <http://censos2007.ibge.gov.br/Censos2007_Inovacoes_ingles_web.pdf>

IBGE. Report of the National Statistical and Geographical Institute of Brazil on global geographic information management: towards better global coordination and integration with statistical information. 41st Session of the United Nations Statiscal Commission, February 2010. Available at: < http://unstats.un.org/unsd/statcom/doc10/2010-13-Brazil-GGIM-E.pdf>

IBGE. 2010 Census. Summary of Survey Steps. Rio de Janeiro, September 2010. Available at: <http://www.censo2010.ibge.gov.br/download/sintese/sintese_censo2010_ingles.pdf>

NUNES, E.P. IT in the Brazilian 2010 Census. Seminar on New Technologies in Population and Housing Censuses: Country experiences. 42nd Session of the United Nations Statistical Commission – Side Events. February, 2011. Available at: <http://unstats.un.org/unsd/statcom/statcom_2011/Seminars/NewTechnologies/default.html>

PRINCIPLES and recommendations for population and housing censuses: revision 2. New York: United Nations, Statistics Division, 2007. 507 p. (Statistical papers. Series M, n. 67). Available at: <http://unstats.un.org/unsd/demographic/sources/census/docs/P&R_Rev2.pdf>.