Paperless census: the Brazilian experience

Américo Vicente Silva de Miranda Júnior*
Brazilian Institute of Geography and Statistics, Rio de Janeiro, Brazil
americo.junior@ibge.gov.br
Luciano Tavares Duarte
Brazilian Institute of Geography and Statistics, Rio de Janeiro, Brazil
luciano.duarte@ibge.gov.br
Miriam Mattos da Silva Barbuda
Brazilian Institute of Geography and Statistics, Rio de Janeiro, Brazil
miriam.barbuda@ibge.gov.br
Wolney Cogoy de Menezes
Brazilian Institute of Geography and Statistics, Rio de Janeiro, Brazil
wolney.menezes@ibge.gov.br

Abstract

In 2010, the Brazilian Institute of Geography and Statistics (IBGE) carried out the first population census that was completely paperless, being collected through mobile capture devices (DMCs). Many technological and methodological innovations were introduced in different processes, from map preparation to data collection, control, verification and dissemination. The DMC made fieldwork easier and was employed in many stages of the census operation: training, pre-collection, data collection and supervision. The equipment enabled census mapping to be used in digital graphic files that were associated with a preliminary list of the addresses to be collected by the enumerator, making their job easier and enabling the monitoring and evaluation of field coverage. In addition to face-to-face electronic collection with handheld computers, for the first time the IBGE allowed the respondent to fill out the questionnaire through the Internet, using a website exclusively designed for the census. The use of the devices in data collection led to significant gains in time and quality, since they allowed the introduction of rules for jumps and editing rules during the interview, in addition to the pre-codification of most open-ended questions. The IBGE was able to follow the evolution of the collected data because microdata could be accessed as soon as they were transmitted, with no more need to wait for a data digitalization stage (through typing or scanning). Moreover, electronic collection enabled the creation of automatized systems to control and supervise data collection, improving and accelerating the quality control process of the enumerator's work. Electronic collection also made it possible to create a public website for society to follow census operations and preliminary information.

Keywords: mobile data capture devices; data collection; innovation.

1. Introduction

In Brazilian census operations, institutional mobilization exceeds its own human resources limits to improve information production processes, motivated by the goal of portraying the reality of a country with continental dimensions. The Brazilian Institute of Geography and Statistics (IBGE) always introduces new technologies, especially in census operations. However, the last operations were characterized by a significant transformation in the process and a paradigm shift, with digitalization replacing paper from the data collection stage until their release. Before 2007, the population surveys carried out by IBGE employed a paper questionnaire filled out by an enumerator on face-to-face interviews. This traditional method for collecting information required a considerable amount of time to complete the cycle of household visits, filling out questionnaires and converting them into digital media through typing or mark or character recognition.

With the introduction of mobile data capture devices (DMCs), the operation went through a decisive transformation since data was recorded in digital media while the questionnaire was being answered, decreasing the duration of the process.

This article aims at discussing the key operational stages of the census that were affected by the introduction of the electronic questionnaire and its implications in the process.

2. Planning the territorial database

The territorial database assists the IBGE in planning statistical surveys and censuses, constituting a system of cartographic information associated with records in a complementary relationship that offers statistical and geospatial information on the Brazilian territory and its dynamics.

The volume and costs of the data collection operation are measured through a characterization of the fundamental aspects of the national territory, allowing for a delimitation and description of the smallest collection units – the census sector, which ensures the enumerator recognizes their work area and avoids the occurrence of omissions and/or duplicities that could harm the territorial coverage of the survey.

The census sector started being the required level of information to fulfill its original purposes of aiding data collection and composing areas to help the planning of the government and the private sector.

2.1 Strategies and goals

For the 2010 Demographic Census, the geographic characteristics and the records of the territorial database were updated as in all census, but new technologies and inputs from remote sensing and GPS surveys were employed in geospatial databases designed for the Geographic Information System (GIS), constituting an entirely digital process.

The great challenge of the territorial database of the 2010 Census was to integrate mappings of urban and rural areas of different scales and sources. For cities, the mappings originated from cadastral scales created by city government, private companies and local government bodies, while for rural areas they were based on the existing systematic topographic mapping, with territorial coverage shared by the IBGE and the Brazilian Army Geographic Service (DSG).

The maps were produced and updated in a spatial database of the Mapping System (SISMAP), in a distributed environment with production being managed by workflow, which maps all its stages and the IBGE work units involved. The maps were transformed into digital files with enough quality to be used in the electronic collector. Modern information technology resources are employed for the following purposes:

- Associate the National Address File for Statistical Purposes (CNEFE) with blocks and block-faces in urban areas of census sectors;
- Adapt the geometry of urban and rural grids, and city and state limits, to elements of the inputs of remote sensing and cartography used for the construction of sector maps, city maps and city statistics;
- Design the territorial database to be continuously updated and used in surveys, increasing the potential of statistical data dissemination.

At the end of this diverse productive process involving many work areas, the inputs went through a quality control system in a central environment for the pre-collection and data collection stages, before being packaged and sent to more than 7,000 collection centers in the country.

3. The National Address File for Statistical Purposes (CNEFE)

In all censuses carried out by IBGE, a list of the visited addresses was created with the purpose of following the operation and assessing its level of coverage. However, for a long time the information was used only for operational control and not as a final product of the operation. Thematic information, for instance, was converted into digital media for a long time, but it was only after the 2000 Census that the list of addresses (known as collection sheet at the time) went through a digitalization process with scanning and optical character recognition. With this first list of addresses, an initial version of the address file was created for regular urban areas.

In 2007, the large-scale incorporation of mobile data capture devices allowed each DMC used in urban areas to provide the list of addresses identified by the file. Thus, the enumerator could confirm the address or include/exclude data to reflect the observed reality at the moment of the collection. The existence of this preliminary list accelerated the address filing process and guided the enumerator in his work route.

In 2010, the IBGE made a significant transformation in how enumerators used maps and the list of addresses by creating an association between both. For urban areas, the DMC started having a digital map in which each block-face was associated with the addresses located in it. The process was arranged so that enumerators could access addresses by pointing the screen of the DMC to the block-face, allowing them to use the information and possibly conduct an interview.

Nevertheless, for the entirely digital operation to be possible, the thousands of block-faces of the urban maps of each municipality had to be associated with their respective addresses. For that purpose, a pre-collection operation was conducted before the census, in which about 30,000 supervisors covered approximately 220,000 regular urban sectors, performing the required updates on the maps and then filing their addresses.

In rural areas, which are not organized by blocks and block-faces, the 2010 Census collected the coordinates of each surveyed address, enabling the points on the digital maps and the addresses filed in the CNEFE to be associated. As a result, almost all of the 78,000,000 addresses collected in the operation were associated with maps – whether to a block-face or coordinates.

The address file produced in 2010 has been partially updated, aiding the household sample surveys of the IBGE. The information was made public and has been widely used by society in many different applications.

4. Training

The training program of the 2010 Census innovated by employing self-learning through distance education in two modalities: online with a tutor and offline with a monitored forum, allowing trainees to arrive at the classroom course with some knowledge of the technical content.

With the inclusion of the DMCs in classroom training, participants were able to practice handling the tool before going to the field. With this purpose, training applications in the DMCs provided exercises and simulations during the training offered to trainees, as well as the final test and its correction.

Additionally, the application provided an evaluation of the training program that was filled by the trainees.

5. Data collection

The investigation model of the 2010 Census has been used in Brazilian demographic censuses since 1960, with a sample questionnaire and a basic questionnaire. All variables of the basic questionnaire are part of the sample questionnaire, and the common variables are investigated for all households and persons. The device randomly selects a household for sampling, according to the sampling fraction of the municipality.

In order to assist in the operational work of the collection process of the 2010 Census, transitory work units called collection centers were established in each municipality, district, neighborhood or locality. In these centers, enumerators received information, learned which sectors were to be covered, asked questions and transmitted collected data to the data storage centers every week, using the Management Information System of the Collection Center (SIGPC) installed in center computers.

Although the numbers of households and persons have increased over the last decade, the introduction of DMCs in the collection process enabled IBGE to employ a similar-sized staff in both the 2010 Census and the 2000 Census. Moreover, the electronic questionnaire was designed with mechanisms that made data collection more dynamic, allowing for a higher number of questionnaire items in comparison with the 2000 Census without compromising operational deadlines.

6. Census by Internet

For the 2010 Census, the possibility of collecting data through the Internet was made available with the purpose of ensuring interviews would be made when it was difficult for enumerators to conduct face-to-face interviews, such as the case of areas with limited access or when it was hard to obtain answers from residents. Although it was a significant improvement in census procedures, the enumerator was advised to be cautious when opting for the Internet because it was possible the resident would not fill out the questionnaire.

After choosing to answer the census through the Internet, the resident received an envelope containing all orientations for the access and the questionnaire, such as that the deadline to fill it out was five days. With the access code provided by the letter, the resident accessed the website informed by the IBGE whenever it suited them best, typing the code printed inside the envelope (the e-ticket) and the security password required, according to the corresponding number in a password table.

It is important to mention that the responsibility of the enumerator did not end with the delivery of the envelope. If the questionnaire was not filled out within the deadline or lacked information, contact with the resident was regained to request the answers. Firstly, the reminder was made through the call center of the IBGE, with an automatic phone call remembering the resident to fill it out. If the questionnaire still needed to be filled after 48 hours, the call center employed active telemarketing, with the attendant contacting the resident to remind them of the questionnaire. If the questionnaire was not filled once again, the enumerator returned to the household and conducted a face-to-face interview.

With this procedure, approximately 38,000 questionnaires were obtained through the Internet. It is worth observing that the computational infrastructure acquired for Internet use was entirely employed in the remaining surveys at the end of the 2010 Census. In addition, the application worked properly and the Internet questionnaires were filled out well. Moreover, in spite of the low number of Internet respondents, their profiles were characterized by economic and educational levels that were higher than average, as it was expected.

7. Evaluating and controlling the census operation

As with the 2000 Census, throughout the entire collection operation of the 2010 Census, the IBGE employed on a national level the Data Collection Management Indicators System (SIGC), a computerized tool that assists in the control of the collection work, accelerating even the payment of enumerators according to their effective production.

During the pre-collection process, many management reports were available so the work could be followed. As for data collection, in addition to management reports, the SIGC also provided the results of the data collection management indicators, which assessed if the collected information differed or not from the expected pattern through parameters defined by statistical methods, allowing for the sectors with the highest probability for the occurrence of errors to be identified.

For all the management levels of the census structure, the SIGC provided fast online access to reports on the progress of data collection, giving reliable and updated information that helped managers make decisions whenever necessary. The data were updated periodically and made available on the intranet. Moreover, the SIGC provided relevant information for the meeting of deadlines and alerted technical coordinators to the possibility of flaws in population enumeration.

Still aiming at controlling and evaluating the operation, a supervision plan composed of a set of tools that systemized procedures and actions was elaborated to guide supervisors in managing and controlling the work of enumerators, generating information that helped them detect possible flaws in the coverage or application of the concepts of the 2010 Census.

The supervision system encompassed a set of reports and indicators and an application installed in the DMC of the supervisor, which was fed by the SIGPC. The fastness of the handheld computer enabled the development of a module in the system called supervision request, which generated samples of units registered by the enumerator that needed to be checked in the sector. The samples were selected in different moments of the data collection process.

At the end of the application of each supervision request, the supervisor finished the request and the system compared the registers of the supervisor and the enumerator, generating a report that enabled divergences to be observed. It is worth mentioning that all information regarding supervision requests and the divergences obtained were made available at the SIGPC and the SIGC, being also part of the protocol for ending data collection in the municipality.

To follow and control the work of the enumerator, the supervisor also used management reports contained in the supervision application, which were updated in their handheld computers after being connected to the SIGPC.

The new supervision methodology introduced to the 2010 Census led to significant improvements in terms of following the collection process. During data collection, at least one supervision request was made for 307,705 census sectors (98% of total sectors), with 6,415,090 addresses being verified (7.7% of the total addresses visited during data collection). Follow-up interviews were conducted in a total of 1,658,771 households (5,813,994 persons), accounting for almost 3% of Brazilian households.

8. Dissemination

The dissemination of the results of the 2010 Census was improved especially by the production and information consultation technologies created over the last decade. As a consequence, the results could be presented as georeferenced information databases, microdata and aggregated data files, digital grids of census sectors and web applications.

In November 2010, the first results of 2010 Census were released in the website of the IBGE. Then, details were increasingly added to the information, made available in different medias and in many spatial cutouts with the purpose of reaching different segments of the population.

In the 2010 Census, the Internet became the main channel for disseminating results for users, providing access to information that varied from data on a statistical chart to complex tabulations in multidimensional databases. In this sense, a comprehensive set of digital files and systems containing the results of the census were presented in the website.

In the hot site created to publish the 2010 Census (http://www.censo2010.ibge.gov.br), consultation tools and web applications enabled results to be disseminated in the most appropriate manner for each user segment. The results updated thematic channels of the website and complemented public access systems. It is worth mentioning that all IBGE products, whether in print or digital files, were made available in the website, consolidating the Internet as the main channel for disseminating the results of the 2010 Census.

9. Conclusion

After the 2010 Census, cartographic reference information and its integration with the CNEFE expanded the use of this technology in current statistical surveys. Nowadays, data collection is entirely conducted with the use of DMCs to fill out questionnaires, with GPSs updating addresses through waypoints and carrying out field surveys by tracking each new block-face and road. The continuous improvement of the systems and the technological advances allow for the information collected through the DMCs to be directly transmitted through the Internet to a central environment.

The DMC has indubitably increased the efficiency of the entire data collection and verification process, allowing the IBGE to use a similar-sized staff in the two most recent censuses (2000 and 2010) due to the improvements in many stages of the survey. Thus, the paperless census is a significant and irreversible step towards excellence in the production of Brazilian public statistics.

References

Barbuda, M. M. da S. (2014). A Experiencia de Países na atualização dos Bancos de Dados Geográficos para a rodada dos Censos 2020, 3º Seminário de Metodologia do IBGE – SMI2014. Available at: <http://eventos.ibge.gov.br/images/smi2014/apresentacoes/SO2%20A%20Experiência%20de%20País es%20na%20atualização%20dos%20Bancos%20de%20Dados_MIRIAM.pdf> [Accessed on April 2015].

Bianchini, Z. (2011). The 2010 Brazilian Population Census: innovations and impacts in data collection. ISI2011 STS29. Available at: <http://isi2011.congressplanner.eu/pdfs/650048.pdf>

IBGE (2013). Metodologia do censo demográfico 2010. Relatórios metodológicos, ISSN 0101-2843. Rio de Janeiro. Available at: <http://biblioteca.ibge.gov.br/visualizacao/livros/liv81634.pdf> pp.352-390. [Accessed on April 2015].

United Nations, Statistics Division (2007). Principles and recommendations for population and housing censuses: revision 2. Statistical papers. Series M, n.67. New York. Available at: <http://unstats.un.org/unsd/demographic/sources/census/docs/P&R_ Rev2.pdf>