United Nations ECE/CES/GE.41/2018/19

Economic and Social Council

Distr.: General
18 July 2018

Original: English

Economic Commission for Europe

Conference of European Statisticians

Group of Experts on Population and Housing Censuses

Twentieth Meeting
Geneva, 26–28 September 2018
Item 5 of the provisional agenda
Technology

Towards the 2020 Population and Housing Census

Note by the National Institute of Statistics and Geography (INEGI) of Mexico

Summary

The National Institute of Statistics and Geography (INEGI), carries out the population and housing censuses of Mexico. INEGI, as part of the National System of Statistical and Geographic Information, must provide society and the State with quality, relevant, clear and timely information to contribute to Mexico development.

The purpose of this work is to make known the computer tools and technologies that have been used or that are scheduled to be used during the work of 2020 Population and Housing Census, from its design, planning, census taking to the results delivery.

GE.18-11933(E)
1811933

I. Introduction

1. The National Institute of Statistics and Geography (INEGI), based on the provisions of the United Mexican States Political Constitution in its Article 26, section B, and in the exercise of the powers that the Law of the National Statistical Information System and Geography confers it, carry out population and housing censuses. INEGI, as part of the National System of Statistical and Geographic Information, must provide society and the State with quality, relevant, clear and timely information to contribute to Mexico development.

2. The purpose of this work is to make known the computer tools and technologies that have been used or that are scheduled to be used during the work of 2020 Population and Housing Census, from its design, planning, census taking to the results delivery.

3. Defined the conceptual, technical, methodological aspects and the corresponding administrative supports, operation of capture and processing phases is carried out according to the requirements established in the design phases and it is finalized with results presentation phase, in the which products and services are offered with statistical data.

4. For increasing efficiency and control in the process described above, for 2020 Population and Housing Census (2020 Census), INEGI plans to make intensive use of communication and information technologies, since it has experience in statistical events previous; in addition to the use of mobile devices, a strategy of operating procedures has been implemented supported by a set of computer tools for the figures of the operational structure, i.e., analysts and decision makers, managing to automate several processes of planning stage, census taking , follow-up, and information processing.

5. The general planning work of 2020 Population and Housing Census started in 2017 with the Public Consultation from August 21 to November 30, which allowed to identify the information needs of users, in addition to other activities such as the performance of Operational Strategy Test 2017, which had as premise the incorporation of technologies and computer equipment in information collection process.

II. Main characteristics of 2020 Population and Housing Census

6. The census data are very useful, since the information they provide allows us to analyse the Mexico transformations and provides the opportunity to identify the social backwardness, vulnerable groups conditions, housing, education, health, and service needs of piped water, electricity and drainage, among other aspects that are important for decision-making in political matters and for preparing plans and programs that tend to combat poverty and improve the inhabitants living conditions.

7. The subject matter that will be part of the census, will seek to meet the information needs demanded by the three levels of government and different sectors of the population. Likewise, it is intended to have information that is required to report the Mexico progress in Sustainable Development Goals fulfillment, Government Plans and Programs, or for covering information needs for execution of actions established by laws that regulate the country development. It is important to highlight that the topics included take into consideration the international principles and recommendations, which allow the comparability of the national situation with that of other countries.

III. Methodological bases

8. The Census 2020, like its predecessors, will be a de jure census, so the population will be counted at its place of usual residence.

9. The observation units of the census are the usual residents of the country, the private and collective dwellings.

10. The proxy respondent will be the head of the dwelling, or in his/her absence, a person of 18 and over who is a usual resident of the dwelling and who knows the dwelling information and their residents.

11. The census taking will take place during March 2020.

IV. Collection method

12. The information collection will be mainly through direct interview with computer applications support in electronic devices. In addition, self-enumeration through Internet and telephone-assisted interview will be implemented, in cases that require it.

13. Regarding the information collection instruments, there will be 2 population and housing questionnaires, a short form, which will be applied in the total enumeration, consisting of around 30 questions, and a long form, for a probabilistic sample, with approximately 75 reagents.

14. In addition, 2020 Census will gather information on the urban environment of the blocks in localities of 5 thousand and more inhabitants, as well as the infrastructure and socioeconomic characteristics of localities with less than 5 thousand inhabitants. It is also scheduled to prepare the Social Assistance Accommodation Census (CAAS), which captures the characteristics of user population, people who work in them, buildings and services they provide.

15. In order to carry out a correct geo-reference of information collected during 2020 Census, an interviewer will list the dwellings and stick a label with a Quick Response code (QR code). The QR code, together with the geo-referenced location of the dwellings through the Global Positioning System (GPS), will allow updating the National Housing Inventory and will start the statistical registration of buildings in the country.

16. The characteristics of the DCM to be used are a Touch Screen with a minimum size of 6.4"and a maximum of 7", with a resolution of 1280x800 Mega Pixels, the operating system will be Android 6 or higher, with minimum specifications of 32 GB of internal storage, 2 GB of RAM, a battery life of at least 8 hours a day, 2 MP digital camera and functionality for USB OTG (On-The-Go). Other necessary components are GPS, ports for SIM and SD cards, as well as Wi-Fi and Bluetooth connectivity. The satellite location/navigation in most cases are GPS and GLONASS, the DCM to be used have GPS and A-GPS technology.

V. Use of Computer Tools in different stages of 2020 Census

A. Conceptual design

17. This look at activities through which the information needs are identified, which serve to determine and define the conceptual framework to which the data will be referred,

the results presentation schemes, instruments for its capture and validation criteria for review, as well as depuration of inconsistencies.

18. In order to support the conceptual design tasks, two web applications were developed: the first one called Public Consultation to 2020 Census Users, which aims to facilitate the task of identifying information needs of users. In it the user can expose, in a systemic and simple way, the objective and importance of their need for information, while, for the decision-makers of the Institute, it allows to follow-up on each of the requests made by society. The Visual Studio .Net development platform and the JavaScript AngularJS Framework were used for this system, and Oracle 12c was used to manage the database.

19. The second application, is the Conceptual Infrastructure of Census and Counts, which aims to be a repository of the conceptual frameworks for population and housing censuses, allowing in a simple way to review and update the conceptual and operative information on the variables that have been collected in census questionnaires, likewise allows to review the conceptual changes of variables over time. Visual Studio Net 2013 was also used for developing this application, with the SQL Server 2016 database.

B. Operative design

20. In this phase, the procedures and technical schemes for the data collection activities are established, as well as the administrative, organizational, control, and following-up aspects. Since the 1990 Census, INEGI has been acquiring experience in systems development for these tasks, for the year 2020 it is proposed to use three systems:

21. The Responsibility Areas Conformation System makes it possible to distribute equal workloads to the operative structure, automatically generating groups or segments of urban blocks or rural localities grouped by their dwellings number and communication routes, among other factors, thus generating operative responsibility areas and with an order of census taking by means of a simulated annealing algorithm. It allows to make adjustments manually to the conformations applying the experience in field and knowledge of operative personnel. This system was developed with Delphi XE4 using a MS SQL SERVER data server.

22. The Training System, developed in the JavaScript language under the Ecmascript6 standard, and with the SQLite database, aims to provide the operative staff with the necessary skills and tools for the execution of their functions through learning modules, in which the didactic materials can be found, remote advice can be received, the learning activities can be sent, and monitor the training advance. In order to generate the application for Windows, the Electron framework was used, and for the Android operating system version, the Apache Cordova.

23. On the other hand, the Opera System, which is used in several stages of information generation process, has made it possible to follow-up on the progress made in operative structure formation, in addition to systematizing the personnel recruitment activities, from the applicants registration via Internet up to their hiring, for monitoring the operative logistics by controlling the materials that are necessary for both field and management activities (vehicles, computers, questionnaires, brochures, among others). For its development, Visual Studio.Net was used and for the database management, the AngularJS JavaScript Framework with Oracle 12c.

C. Information collection

24. This phase covers the set of activities to obtain data of each element of the population, based on the program and established work procedures, with an operational structure and controls that ensure effectiveness in each of the actions. This implies the execution of the schemes for collecting previously designed data, such as the preparation and distribution of support materials (cartography, manuals, instructions, and catalogs), the integration of human resources, communication and agreement, under a detailed activities program, an organizational structure, and controls that ensure effectiveness in each of the actions.

25. For this stage, the Census Administrator is currently being designed in which functionality is sought to visualize and manage workloads and re-assign them, if necessary; it has a Capture Module, which is used for registration of Building List, Urban Environment Questionnaire, Locality Questionnaire and the Inhabited Housing Questionnaire; it allows generating specific progress and coverage reports, and integrates a Supervision Module in which areas sampling is implemented to review the buildings classification and dwellings housing condition; it also allows to select randomly dwellings for re-interviews by the supervisor of interviewers and through verification mechanisms review information quality recorded by the interviewer.

26. The Census Administrator has a Cartographic Module whose purpose is to register the cartographic updates at the block level, locality, and road, allowing to digitize the polygons in new areas, to cancel those that are no longer in field and also to graphically represent changes such as mergers and divisions of areas allowing immediate updating of cartography and planning in information collection.

27. For the 2020 Census, a Module for Coverage Verification Operative will be implemented, whose objective is to facilitate the review activities carried out by the operational structure - which is independent of the interviewers- and carry out a supervision of a sample of areas, and in case if necessary, the missing population information will have to be collected in order to improve coverage in dwellings enumeration and people. On the other hand, it will also contain the adaptation of the necessary computer tools for Post-enumeration Survey execution and a Module for matching of the data obtained by enumeration structure and post-enumeration structure.

28. The Census Administrator is the means by which transfers of information from interviewer to supervisor of interviewers are generated and from this one to the Central Database managed by Opera, which has a module that allows monitoring the integration of the data collected by each operative figure and with it the analysis and the follow-up of certain indicators of coverage, speed, and census taking productivity; which are important for management decision making during operative of enumeration. The Census Administrator will be used in the Enumeration operation as well as in the Verification and Post-enumeration, and it was developed with Embarcadero RAD Studio 10.2.2 Tokyo, Architect version, a native multi-platform and multi-device Windows / Android / iOS development tool that includes the FireMonkey module which allows compiling the same programming code for Windows and Android platforms.

29. Also, for next account of population is intended to significantly boost the self-enumeration via internet and call center, therefore, an application for the population will be available, where you can answer the questions of the short form, it should be mentioned that this system will be adapted for collecting information of collective housing and personnel of Mexican Foreign Service. For the self-enumeration, a web tool was developed with Visual Studio 2017 and the Oracle 12c database.

D. QR use

30. To support the correct dwelling coverage, a unique folio label is designed by means of a Quick Response code (QR), which will be stick on all buildings. This instrument will serve as the basis for supervision in field and verification operation.

31. Also, it will be implemented in Internet self-enumeration strategy or assisted by telephone, that is, people will be able to answer the census questionnaire, without the help of an interviewer, but controlling the coverage and the correct geo-referencing of the information collected, since an interviewer will list the dwellings and stick the label with the QR code, which in turn will be associated with a foliated invitation, which contains a unique username and password for each dwelling, as well as the necessary instructions for Internet registration and the number of help phone, this user will be the same for both options.

E. GPS uses

32. The QR code, together with the geo-referenced location of dwellings through the Global Positioning System (GPS), will allow updating the National Housing Inventory and will start the statistical registration of buildings in the country. In the urban area, it is programmed establish a geographical coordinate for each dwelling, while in a rural area it will be done at some identifiable point in the locality.

F. Information processing

33. In this process, the data files are prepared, making sure they are congruent and ordered for their statistical use; in the case of information collection through mobile devices, it consists of coding activities, validation, updating of geographical keys, and database release. Strategies for information processing are also determined, including the relevant systems design of coding, validation, and results exploitation with their respective quality controls.

34. For processing, the Processing Following-up and Control System will be used, which integrates automatic coding by algorithms that recognize the text of the open variables and assign a key to the predefined catalogs, later assigning the workloads for assisted coding. Once keys has been assigned, the validation of the records is made. This system also includes testers of validation criteria in order to ensure the correct application of the conceptual guidelines.

35. During each stage of the previous processes, the application provides in real time, progress reports on the processes, information quality, and creates a log of the changes in the data. In parallel, assignment of geographical keys of the National Geo-statistical Framework is followed-up1.

36. Finally, it makes available the database for Data Release activity, which consists of a statistical review of the historical changes of socio-demographic indicators at the state, municipal level, and by locality size. For this activity, the Libera System will be used, which allows to take review control of each indicator, but also serves as a repository for documentation that supports changes found in some statistics.

[1] INEGI. Marco Geoestadístico Nacional.
http://www.inegi.org.mx/geo/contenidos/geoestadistica/m_geoestadistico.aspx

The Libera and the Processing Monitoring and Control System were also developed with Visual Studio .Net 2017 and Oracle 12c.

G. Results Presentation

37. This activity specifies the presentation of results previously defined in the conceptual design, bringing to an end the products of the editorial program, in such a way that the diverse needs of the users are met in the best way and at the lowest cost.

38. Traditionally, printed tabulations were considered, however, since 2010 Census, they are no longer produced and instead were made available to users, tabulations in spreadsheets, thereby providing the public with more information processed, surpassing the limitations that paper has.

39. Likewise, since that event, three web systems that allow to more experienced users to access information in greater detail have also been published, these are: the System for Census Information Consultation, which is based on the MxSIG institutional platform, which includes the use of robust free software and open source components such as Tomcat for the publication of services, R for statistics for statistical analysis, PostgreSQL for storage (BD) and PostGIS for geospatial management, allows to associate the census statistical information with geographic space at which belongs, to facilitate the interpretation of sociodemographic phenomena, through the generation of thematic maps; the Consultation System of Territorial Integration and Locality, developed in Visual Studio .Net 2017 and Oracle 12c, which provides, in a dynamic way, the historical information of each inhabited locality of the country; and the Socio-demographic Panorama of Mexico, which is a dynamic consultation system that integrates, as a synthesis, relevant data to know the basic demographic, social, and economic characteristics of population and housing in Mexico.

40. For 2020 Census it is considered to update these consultation systems, besides publishing a Comprehensive Census Information Consultation System with the purpose of making available to the users the greatest amount of information related to the variables included in census projects, from its conceptualization up to statistical results obtained going through the operative design, information treatment and evaluations carried out.

VI. Conclusions on technological architecture and information security

41. In tests the 2020 Census, a strategy has been implemented in order to transfer information in which tools are used to generate packaged files with security passwords and internal integration with the bases of data encrypted with the encryption algorithm of the AES256 method (Advanced Encryption Standard with 256 bit keys) which is one of the safest and most widely used algorithms for public use, but for military security purposes. These encrypted data packets are deposited in USB memories through a function to transfer information integrated in the Census Administrator and are sent through the Information Integration Module of the Opera System, which is housed in a secure institutional website, subsequently deposited in mass storage servers, and finally they are integrated to the Oracle server of central databases.

42. Computer telecommunications services, corporate network and application servers, databases, and mass storage (NAS) inside the corporate network (DMZ and Intranet) are provided by computer infrastructure areas, which remain available these servers. The service availability scheme is configured through a load balancer of requests through the Opera web domain that equitably redirects each data processing request to the application

and database servers. Different security standards and service availability tests allow the planned stability in the whole set of services.

43. For Web applications development, current technological platforms are used, suffering even complete re-engineering, which contributes to a better performance in their operation. The main development tool is Visual Studio .NET, which is a complete set of tools for the generation of web applications, in addition to the AngularJS development framework that provides techniques for the creation of SPA applications (Single Page Application), that benefits performance by making requests to the server lighter by only requiring specific parts of the page. On the other hand, with the use of JWT (JSON Web Token), a more secure communication is implemented and it facilitates the administration of users, assigning roles and permissions. Oracle 12c is used for managing the database, which provides greater security and availability of information.

44. With respect to computing tools for collecting information in mobile computing devices, mainly Javascript and HTML5 are used as programming languages. The technologies used allow the manipulation of digital cartography in GeoJSON format using the Leaflet map handler, which provides special functions for mobile devices. Routines are used to generate questionnaires loaded in SQLite, lightweight mobile database, and through an application server programmed in Delphi Tokyo of Embarcadero Rad Studio, they are compiled for Android and for Windows, obtaining multiplatform responsive interfaces. In addition, frameworks such as jQuery and Bootstrap are used to improve the performance and appearance of these applications.

References

Cámara de Diputados del H. Congreso de la Unión 2017 - Constitución Política de los Estados Unidos Mexicanos

Cámara de Diputados del H. Congreso de la Unión 2008 - Ley del Sistema Nacional de Información Estadística y Geográfica

Instituto Nacional de Estadística y Geografía (INEGI) 2017 - Aspectos Generales del Censo de Población y Vivienda 2020. Consulta Pública 2017. Aguascalientes, México, INEGI. (Internal document)

Instituto Nacional de Estadística y Geografía (INEGI) 2010 - Censo de Población y Vivienda 2010. Aguascalientes, México, INEGI