Meeting of the Expert Group on Business Registers
Brussels, 21 – 23 September 2015

Name of author(s): Susana Patricia Pérez Cadena (susana.perez@inegi.org.mx)
y José Arturo Blancas Espejo (arturo.blancas@inegi.org.mx)

Organization: Instituto Nacional de Estadística y Geografía (INEGI), México

Sesión III. Use of geo-spatial information and other sources: practices and matching methods for improving the SBR

Establishment geo-referencing system for mobile devices used in the Mexican 2014 Economic Censuses and perspectives of use for geo-referencing economic units from administrative registers

Abstract

In the collection of the 2014 Economic Censuses, verification and updating of the National Directory of Economic Units (DENUE) and the census cartography were performed. To do this a system for mobile devices was designed, tested and used, during the course of the census interviews, allowing the location of both the establishments and houses in the exact point where they are found in the geographical space. This method gave better results than other geographic localization systems, such as the global positioning system (GPS), since it is more accurate in the positioning of the point where the establishments are located on digital cartography. Thus, using the system for mobile computing devices, a point was assigned to over 5 million surveyed establishments.

The application includes tools for incorporating new blocks, roads, neighborhoods, updating names and types of roads, exterior numbers and merging of blocks. This means it allows incorporating the changes that have occurred in settlements and referencing the economic units and houses in the geographic space with greater accuracy and, above all, with a greater degree of confidence in the positioning of the geographical points on digitized cartography.

On the other hand, the need for geo-referencing the establishments from administrative registers that lack complete geographic location data, provides the future possibility of applying this field method, modified to be used at the office, to find the geographical point of those establishments using information that is updated

by the censuses, such as the exterior numbers, digital cartography, polygon vectors of blocks, DENUE, Google tools, among other elements.

This paper describes the features of this application, its scope and the results obtained in the geo-location of the economic units captured by the 2014 Economic Censuses, as well as the outline of the method for geo-locating the point on cartography, which will be used to reference the establishments from the administrative register and incorporate them into DENUE.

Introduction

The collection of the 2014 Economic Censuses was based on The National Statistical Directory of Economic Units (DENUE), which in turn was the result of the previous censuses and the subsequent updates from administrative registers. Being based on the DENUE allowed a stricter control of the collection and enabled the possibility of observing the changes that occurred since the previous census record by record, point by point. Of course, among the census outcomes, a new fully updated version of the DENUE was produced including the exact location of every establishment.

For this, an automated system was developed to allow locating on cartography the almost exact point corresponding to the location of every establishment in the country. Moreover, the point was allocated not only to every establishment but also to every dwelling, empty premises, land under construction, empty lot, and park; in summary, to every urban land plot in every one of the 1.4 million blocks visited in the country.

The developed system ensured that “the point” was located not only in the correct block but also in the correct blockface and in the location corresponding to the establishment or very close to it, but always in the correct blockface.

Carrying out the digital geo-location at the blockface level for the establishments, dwellings and urban land plots was feasible because INEGI counts on a digitized National Geo-statistical Framework that keeps updating with every operation the institution performs. This infrastructure, the product of the coordinated and collaborative efforts of the areas that work in the production of statistical and geographic information, as well as the areas that develop computer applications, has made it possible to accurately geo-reference the observation units including their statistical information.

To meet the needs of collecting information, updating the geographic data, as well as geo-referencing the establishments, *the Mobile Computing Device Operating Routine and Capture System* was developed, which has several modules that, aside from allocating the point and updating geographic and statistical information,

allow the monitoring of progress and coverage from the planning up to the closing of data collection. This document will only address the purpose and function of one of its components, the Cartographic Module corresponding to the cartographic update and the allocation of the geographic point of establishments, dwellings and urban land plots. The results obtained from using this cartographic module were compared to another point allocation method using a GPS, through field tests.

Moreover, the need of solving in the office, the cartographic geo-location of some establishments that were not captured in the field by the census enumerator (for example, when dealing with branches whose information is provided by the parent), led to the use of the application called ***Establishment Allocation and Reallocation System*** (SARE). This application tangentially served for the reallocation of the geographic point to a small number of establishments that presented problems in the assignment of the geographic point, due to failures in the processing and transmission of files.

In the first section, this document describes the outcomes of the tests performed for geo-referencing the establishments through GPS and the *Mobile Computing Device Operating Routine and Capture System*. The second section describes the most important features of the Cartographic Module, as well as some results obtained from the geo-referencing. In the third section, the main characteristics of SARE are delineated, as well as the results of its application. Finally, the paper addresses the problems faced and perspectives on the use of SARE for in-office geo-referencing of economic units from administrative registers that provide DENUE with information.

Field test performed for geo-locating observation units

Considering the territorial features of the country and of the observation units to be geo-referenced in cartography, the most convenient method for the cartographic location point of every establishment, dwelling and urban land plot had to be chosen, i.e. for the allocation of the geographic coordinates to each of these units. Therefore a test was done in March 2013, using a mobile computing device which integrated a global positioning system (GPS) that included digitized cartography, and a Cartographic Module pre-loaded with vector information on the collection area, which was developed by the National Institute of Statistics and Geography (INEGI).

The inputs used for the test were:

- Tablet type Mobile device.

- GPS USB device.
- ND-100S Software for capturing GPS data.
- Cartographic Module for the 2014 Economic Censuses.

Test results

After performing the test and comparing the results of both methods for geo-referencing the establishments, some disadvantages were observed in the use of GPS:

- Before initiating the geo-location of units, the availability of at least three satellite transmitters of GPS signal had to be checked.
- The ND-100S software used for receiving the geo-location data on the establishments occasionally presented problems: for example, it froze.
- The climatic and physical conditions of the area affected the allocation of coordinates because it diminished the quality and intensity of the signal; i.e. cloudy day, forested area, tall buildings, roofed sidewalks, etc., were factors that prevented proper signal reception and accuracy.
- There was no guarantee that the point was located on the correct block or blockface. To determine the displacement of the measured point to the coordinate obtained, the approximate location of the premises was established in a raster image and the displacement was measured in relation to the coordinate captured with the GPS USB device, finding a variability in a radius of 2.19 to 29.79 meters, which produced an outcome where the establishment could be referred to another block or to another blockface.

On the other hand, with INEGI´s Cartographic Module the following results were obtained:

- It guarantees that the establishments are geo-referenced in the correct block and blockface, with a margin of error that does not exceed 10 meters (always in the blockface in which the unit to be geo-referenced is actually located) since it has the vector information of the area of work and the blocks the census enumerator covers.
- The inaccuracies in the registration of the point in the blockface that could be generated by using the mobile computing device´s stylus to manually allocate

the point (due to the thickness of the stylus tip, interviewer´s pulse, among other factors), are corrected since the system has block vectors and when the census enumerator selects the blockface to cover, the system automatically places the points in that blockface and in no other.

- Its functionality and accuracy do not depend on the network or satellite signals, nor on the environmental factors because its functionality is local, i.e. it has the pre-charged digitized cartography, satellite images or orthophotos, as well as data on DENUE´s establishments at block level[1], and vector information on blocks and their blockfaces.

Based on the outcomes of this test, it was decided to use the Cartographic Module that was developed by INEGI, which is part of a more general purpose system, called the ***Mobile Computing Device Operating Routine and Capture System***.

Equipment for census work

In order to guarantee the good performance of the system, Meebox1 brand classmate mobile computing devices were used, with an Intel Celeron 847 Dual Core (1 Mb Cache, 1.10 Ghz, 32 bit) processor and a 2 GB DDR3 13333Nhz RAM; 180° bidirectional rotation screen, 10.1" LCD Touch Screen 10 point capacity and stylus compatible to the touch screen. The devices had 32 bit Windows 8 Pro in Spanish installed in the operating system.

These devices were resistant to outdoor environments, user-friendly, with outer shell resistant to shock, weighed 1.72 Kg and had a 6-cell battery lasting 6 hours in continuous use.

Cartographic Module of the Mobile Computing Device Operating Routine and Capture System for the 2014 Economic Censuses

The Cartographic Module is a local/mobile application derived from the Digital Map of Mexico[2] System desktop computer, which allows geo-referencing the various

[1] The DENUE was divided up to block level to integrate the Directory of Establishments by Block (DEM) used in the census collection.

[2] It is a Geographical Information System (GIS) developed by INEGI, which integrates information on the natural and cultural elements that form the geographical environment of the country and allows relating them to statistical information. The desktop mode was designed for promoting and facilitating the

units (establishments, dwellings and urban land plots) performing specific operations of cartographic updating during the operations, contributing to the updating of the Unique Cartographic Base (BCU) of INEGI. Its design, development and implementation summarize the years of institutional experience in capturing and geo-referencing information.

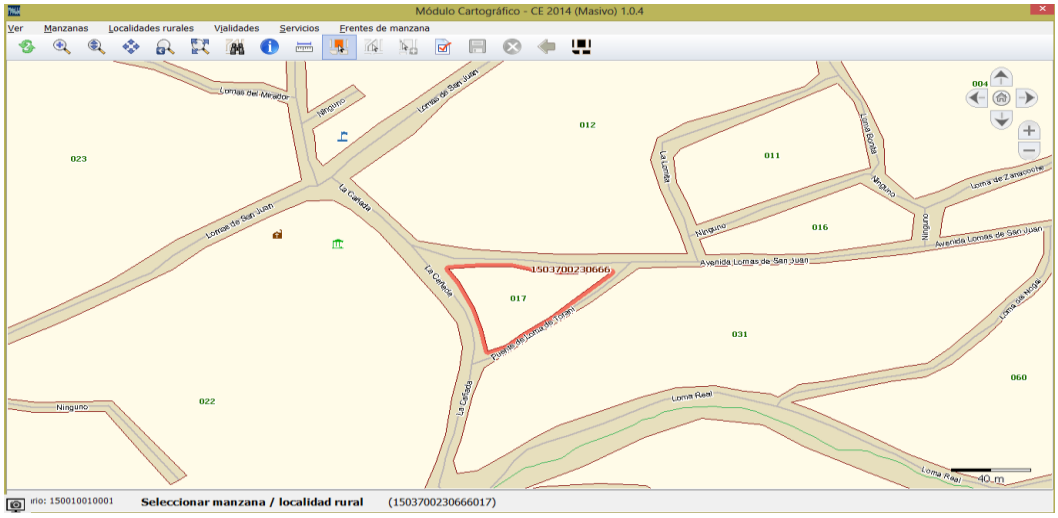

Illustration 1 Cartographic Module.

It features digitized cartography[3], with satellite images[4] and orthophotos[5], for enabling both cartographic updating work and tasks related to the geographic point allocation.

Main characteristic of the Cartographic Module

As mentioned before, the Cartographic Module represents a simple interface for managing cartographic information, which works locally and has tools for performing census tasks, allowing:

---

integration, use, interpretation and analysis of geographical and statistical information; it allows the connection to geographical databases and web map servers, incorporating information such as data tables and documentation, among other capabilities.
http://www.inegi.org.mx/geo/contenidos/mapadigital./default.aspx

[3] Lambert Conformal Conic Projection (CCL) Parallels type 29°30' (N) AND 17° (N). Origin latitud in the parallel 12° (N) = 0 meters in the "y" (ordered) and the origin longitude in the meridian 102°" (W) 2,5000,000 meters en the "x" (abscissa) en the reference system ITRF92.

[4] Satellite images from the Spot 6 satellite. Updating: 2013-2014 (Oct-April). Resolution: 1.5 m Panchromatic. Resolution: 6 m. Multispectral. National Coverage. Y, Geoeye. Resolution: 50 cm. Updating: 2011 to 2014. Partial Coverage.

[5] Based on aerial photographs: Resolution 1.5 and 2 m. Coverage: 99% of national territory.

- Updating of the sketch: create, merge, and split blocks and blockfaces; create roads, localities, and services; modify names and roads, etc.
- Selection of blocks, blockfaces and the allocation of the point to geo-reference the establishments, dwellings or urban land plots that are in scope for the census.
- Use on mobile devices: laptops and tablets.
- Processing of information stored in digital format in Geographic Information Systems (SIG), and allowing its use in government offices or society in general.
- Working only with information related to certain geographical areas, minimizing errors due to a poor location of field personnel.
- Keeping a permanent update of the cartography of blocks, streets, external numbers, services, dwellings and urban land plots in the urban and rural environments.
- Minimizing the use of paper sketches and the consequent need for physical storage.

Procedure for geo-referencing

To capture the information, the census enumerator had to travel the block selected from his workload three times. The first trip involved updating the cartography; the second included unit geo-referencing (establishments, dwellings and urban land plots) and data collection; and the third implied geo-referencing and data collection of the semi-fixed establishments found on the block.

In updating cartography, according to spatial reality, the census enumerator could:

- update the outline of the blocks to identify cases of partial or total closing of roads, block merges, full opening of roads or block subdivision, deletion of blocks or creation of new blocks;

- register the movements generated in the urban sketch, and create or eliminate any public service (plazas, temples, parks or gardens, etc.);

- register changes in terms of roads, updating of street names or avenues, human settlement, postal code and exterior number of every urban land plot; and

- eliminate or create blockfaces, in which the points would subsequently be placed for geo-referencing identified establishments, dwellings or urban land plots.

Following the cartographic update, the census enumerator selected the blockface for visiting and initiated the geo-location of units and data capture. Once the blockface was selected, the location point of every establishment, dwelling and urban land plot found there was registered using the mobile computing device´s stylus.

After registering the unit´s geographic point, the system displayed a text asking for the external number and, if applicable, the internal number. Once the data was registered, another text appeared asking for the confirmation of the captured data.

If the exterior number had already been registered, the system sent an alert of duplication and asked if it belonged to the same building; if so, data was stored and the geographic location was displayed with a star. Whenever the exterior number was repeated, the system asked if the wish was to geo-reference another establishment or dwelling in the building being visited.

If the system detected an important variation while recording an exterior number regarding the previously registered number in the same blockface, a text was displayed asking if the exterior number was correct, giving a correction option in case of an error.

When data capture was concluded and the storage of the location was confirmed through the “save” option, the system automatically associated the rest of the address and geographic location data to every unit: street name, neighborhood, postal code, locality, municipality and state name and code.

Once the establishments and dwellings were located, the system displayed a screen for selecting the type of unit and action associated to it: to survey the establishment, count the dwelling or take note of the urban land plot without an observation unit.

If the registered unit was an establishment, the system displayed the questionnaire for capturing the business economic information.

If the registered building was a dwelling, the system required the confirmation of being inhabited or uninhabited. If it was inhabited, it always displayed the option of

registering whether there was economic activity or not. In case of being a dwelling with economic activity, then the system proceeded as if it were an establishment.

In case of urban land plots, the system opened a dialog box for selecting the option of empty plot, empty premises, land with building under construction, crop field or barnyards, etc.

Results and Problems

The automation of the process of unit geo-location and data capture was in general successful; it presented only a few problems, mainly in the transmission of files, causing damage or loss of some of them and having to reallocate the geographic point in the office. In this case 17,364 establishments were found, which represented 0.34% of the total of 5,023,077 geo-referenced establishments in the collection, from which 4,903,381 were geo-referenced by the system.

Geo-location of establishments, dwellings and urban land plots through the Cartographic Module resulted in 43,404,529 observation units geo-referenced on field, of which 11.3% corresponded to establishments[6], as seen in the following table.

Type of unit[7]	Total
Establishments	4,903,381
Dwellings[8]	33,164,790
Urban land plots	4,832,226
Total	43,404,529

This information will be used in different projects at INEGI as input for the production of statistical information, such as economic, socio-demographic and geographic studies.

[6] Includes 504,132 dwellings in which economic activity is performed, that were excluded from dwelling´s total data.

[7] Does not include the establishments collected in the rural sample of the 2014 Economic Censuses.

[8] Dwelling data excludes the allocated point to the 504,132 dwellings with economic activity that were accounted for in the number of geo-referenced establishments.

In-office geo-referencing

The Cartographic Module, as a derived application of the Digital Map of Mexico, for the allocation of the geographic point on digital cartography, was generalized for the allocation of the point in the office, since, as previously mentioned, the allocation of the geographic point to some establishments[9] had to be solved, through computer developments that allowed locating the geographic point On digital cartography with the address data. For this, the Cartographic Module for geo-referencing in the field was generalized as a tool for allocating and reallocating the points to the establishments, and it was called Establishment Allocation and Reallocation System (SARE).

In contrast to the Cartographic Module, this application only has the objective of performing the allocation of the geographic point in the office; it has access to internet, it contains the layer of exterior numbers and it has access to street view images. These are elements of great utility for increasing certainty in the allocation of the geographic point in the office.

SARE was used for allocating or reallocating the point to establishments in the office (replotting), for only 2.4% of the total geo-referenced establishments.

Establishment Allocation and Relocation System (SARE)		
Allocation	**Reallocation[10] (replotting)**	**Total**
3,847	115,849	119,696

[9] The point of the establishments had to be digitized, in those areas in which for security reasons the work of cartographic updating and data capture had to be done on paper, and tangentially solve some geo-referencing issues that appeared during the transmission of the point information of some establishments.

[10] Includes the 17,364 geo-referenced establishments on field with the Cartographic Module.

The following table summarizes the total of the economic units where the geographic point was allocated to digital cartography through the two systems developed for geo-referencing the information collected by the economic censuses. This information will serve to update the DENUE and other databases.

System	Total of economic units with a point
Cartographic Module	4,903,381
SARE	119,696
Total	5,023,077

Perspectives of using the system for geo-referencing establishments from administrative registers

Mandated by the Law of the National Statistical and Geographic Information System, DENUE also updates its information with data from the State´s administrative registers, such as the tax registry, the electric power supply service and social security. However, these registers generally lack geographical coordinate data for locating the economic units that is essential for the release of the Directory´s information.

Aside from the need to verify data provided by these registers, it is essential to capture data corresponding to the geographic point of the establishment´s location in order to incorporate them into DENUE. These tasks, when performed in the field increase costs, reduce the timeliness of data and raise the response burden on respondents.

Before the development of the in-office modality of the Cartographic Module for the allocation of the geographic points —el SARE—, data verification work in office was conducted by telephone or through enterprises´ websites, and the allocation of the geographic point was semi-automated, thus requiring more time and not guaranteeing data integrity. With SARE the tasks of geo-referencing establishments lead to the reduction of costs and response burden on respondents, as well as better timeliness for updating the Directory´s data and its release.

SARE opens a new possibility for linking administrative registers, because it could be the tool used in government agencies, which are providers of information for updating statistical registers, so the respondents can locate their establishment´s point on digital cartography at the moment of registration in the tax system, social security or electric power supply service. Having full information on the establishment´s geographical location will facilitate linking administrative registers with different statistical registers through the Enterprise Statistical Code.

Conclusions

The Cartographic Module proved to be a very accurate tool for geo-referencing the various observation units, being a means of cartographic updating, it reflects spatial reality on the maps; however, being a field instrument, its use is limited to census or survey periods.

In this sense, SARE represents an important opportunity for its use in the inter-census periods for geo-referencing administrative register data, which provide information for various statistical projects. Its use by government agencies would be an important element for linking conceptually different registers, through the establishment´s location data and for the reduction of response burden on respondents of the National Statistical and Geographic System.

In this context, the development of these geo-referencing, statistical and geographic data updating systems —which is INEGI´s work, resulting from the coordination, experience and efforts of various areas within the Institute—

contribute to the purpose of capturing and producing statistical and geographic information with higher quality and better timeliness, increasing the efficiency of work in the Institution; these are factors that also positively impact the fulfillment of the objectives of the National Statistical and Geographic Information System as providers of timely information for the development and evaluation of public policies.