Use of administrative registers for strengthening the geostatistical framework of the Census of Agriculture in Mexico

Susana Pérez
INEGI, Dirección de Censos y Encuestas Agropecuarias.
Avenida José María Chávez 1913 Fraccionamiento Prados de Villa Asunción.
Aguascalientes, México.
susana.perez@inegi.org.mx

DOI: 10.1481/icasVII.2016.f32b

ABSTRACT

For a few years now the statistical use of administrative registers has been fostered, as they have been an important tool to reduce costs in producing information and the burden on respondents, for this reason the institutions responsible of producing statistics have boosted capacity building, in order to increase knowledge on the use of administrative sources and ensure the quality of the data produced.
In Mexico, INEGI has established agreements with various institutions to take advantage of the information produced by the administrative registers they produce, with the aim of producing demographic, economic and social statistics.
Regarding the agriculture and forestry sector, in Mexico some of the main sources of administrative registers have been identified, among which the following are worth mentioning: the Ministry of Agriculture, responsible of regulating the actions related to the agricultural and fishery activity; the National Forestry Commission, in charge of fostering the conservation and restoration of forests; and the National Agrarian Register, responsible of controlling and updating part of the rural property in the country.
In this context, there have been collaboration agreements with the Ministry of Agriculture and the National Agrarian Register, who provided inputs such as vector files of land plots and tables of producers from their respective programs, with the objective of analyzing them to determine their utility for strengthening the list and area frames of the Census of Agriculture.
In the beginning, incorporating the tabular and cartographic information into the geo-statistical framework of the Census of Agriculture, was thought to be simple, since it was easy to use digital files; however, the process has not turned out this way, given that there are many considerations to take into account for using the information obtained.
Thus, before incorporating the vector and tabular information to update the list and area frames of the Census of Agriculture, it is necessary to consider important aspects like: the projection in which the land files were produced, the geodetic reference system, as well as the standardization of the vector files obtained from different sources. Whereas for the tabular data, it is very important to obtain the minimum information necessary for the identification of both the producer and his address, independently from the structure that defines every agency.

Key words: administrative registers, census of agriculture, geostatistical frame, list frame, area frame.

1. Introduction

One of the viable alternatives for obtaining data with statistical purposes is the exploitation of administrative registers generated by federal agencies and private entities. In this regard, INEGI has reached agreements with public sector institutions that collect information from individuals and legal entities for using it as a data source to produce statistics with a geographic breakdown.

INEGI currently produces and publishes statistics from exploiting administrative registers in the following matters: economic (mining, foreign trade, and public administration), social (culture, labor relations, suicides, security and public order and health) and demographic (births, deaths and marriages).

Concerning the exploitation of administrative registers from the agricultural sector, INEGI has established alliances with some institutions in order to use their tabular and cartographic information, with the purpose of strengthen the permanent updating of the geostatistical framework of the Census of Agriculture.

2. Analysis of administrative registers from RAN and PROCAMPO to strengthen the geostatistical frame of the Agricultural Census

As a result of the 2007 Agricultural Census conducted in Mexico, the digitalization of approximately 70% of the land plots, where agricultural activities are held, was obtained. The remaining 30% of the land plots is not accurately demarcated since their specific boundaries are unknown, however, they are identified by INEGI in a "control area" (one section of the 187,326 parts that INEGI used to divide the national territory). Additionally, as a result of the Census of Agriculture, it was determined that 49% of the producers are linked with their land plots as production units, while the other 51% production units are not linked.

With the purpose of increasing this 49% of production units, several attempts have been made for incorporating information from administrative registers into the Agricultural Census' information; nevertheless, most of the registers of producers' associations were discarded, since they contained heterogeneous and incomplete information, and they could not be used for this aim.

During the review of different sources, it was identified that RAN and PROCAMPO registers' have three basic characteristics that allow a better and detailed analysis:

1. - They contain the producer's table (a list of several variables that includes the name and address of the producer and other information)
2. - The land plot of the producer is a located territory and has defined boundaries.
3. - Both elements are in digital format; the first one is in database file (.DBF) format and the second one is in shapefile (.SHP) format.

Considering the necessity of being provided with the largest number of rural land plots, properly identified and linked with its producer, INEGI glimpsed the possibility of strengthening the sampling frame of the Census of Agriculture, by exploiting RAN and PROCAMPO information.

In view of the above, it was necessary to train INEGI staff in the use of GIS. With these systems, spatial comparison of the vector files of the land plots from the Census of Agriculture and those from the mentioned sources could be made by using different tools. Besides, high resolution satellite images (with a spatial resolution of 2.5 and 1.5 meters) were used and added into the GIS, in order to use them as a background. Satellite images were used to support the correct location of the boundaries of land plots that normally coincide with roads, rivers, and some other physical boundaries. Additionally, batch processing of geographic information is allowed by the GIS, which reduces the required time to compare land plot files and its databases.

Although digital files are easy to use and integrate into other sources of data, it is important to consider the following aspects before the analysis, and exploitation of administrative data:

1. - Geographic information must be in the same geographic and reference systems and in the same cartographic projection; for all cartographic elements, it is important for the register to be generated with the reference geodetic system's parameters and with the projections that are officially managed in the country, as well as using the parameters that are normally published in official journals for public knowledge, in order for the matching to be as accurate as possible.

Figure 1: *Shift effect observed by the different geodetic reference systems in shape files.*

2. - Database information must have the same structure (size and content of fields must be normalized). Concerning tabular files, it is required that they have basic information needed for the identification of both the producer and his address, regardless of the structure defined by each agency. It is therefore recommended for the name of the producer to be split by the last name of the father, the last name of the mother, and the name itself, in separate fields as it is common for the full name to be included in a single field, which makes information difficult to handle. It is also recommended for the information on the address to be organized according to the technical rules generated for its standardization; these rules usually specify the minimum information on streets and surrounding streets that allow locating addresses; furthermore, localities must be properly identified with the name and geostatistical code that have been assigned to them. Normalization of the database structure will prevent any problem in the process of the information and it will also allow the interchange of data between databases.

3. Analysis of information from the RAN

The measurement and certification of land plots and plans from each "ejido" in México, (shared lands of social property), was done between 1993 and 2003 and backed up on magnetic media by the RAN. The technical works of measurement were carried out by INEGI until 2003; after that, RAN has kept on updating its information. Now INEGI considers important to use these administrative registers to update its information as well.

INEGI obtained the vector files of land plots and producer's tables from RAN, with the purpose of analyzing them and determine if they were able to be used for complementing the information produced by the 2007 Agricultural Census. The criteria set to analyze the information were the following:

1.- The land plots that spatially coincided with information that INEGI already had, were not used to add information to the area frame of the Census of Agriculture; they were only used to verify that INEGI's data is updated.

2.- RAN's land plots, located inside one of INEGI's land plots, were considered as a subdivision, as long as they coincided with its boundaries. In this case, RAN's land plots were considered as part of the updating of the area frame of the Census of Agriculture.

3.- RAN land plots that are subdivisions of a land plot from INEGI, but whose boundaries overlap INEGI's boundaries, will be visually analyzed in detail to determine if they will be used for the updating process.

From a total of 5 685 450 land plots received in shapefiles format, approximately 94% were part of the land plot shapefiles of the Agricultural Census. The high coincidence of land plots from this source is explained because INEGI initiated the social property land plot certification task in Mexico in 1993 and then, in 2003, the activity completely became RAN's responsibility. The remaining 6% corresponds to the land plots incorporated and surveyed by RAN from 2003 up to date and they are divided as follows:

3.31% of the remaining land plots correspond to lands that can be directly incorporated in the shapefile format of the Agriculture Census' land plots, or with minimal editing; these land plots corresponded to subdivisions of larger areas that were already defined in the Agricultural Census, or to common use spaces of the "ejidos" that were divided and distributed among the landowners.

Figures 2 and 3 show different examples of RAN's land plots that were incorporated within the boundaries of a large land plot of the Census of Agriculture, with the respective information of producers.

Figures 2 and 3: *Examples of information of RAN incorporated into Census frame*.

The remaining 2.71% corresponds to land plots whose configuration does not match exactly with the spaces bounded by the Agricultural Census, making it necessary to perform a detailed visual reviewing using the satellite image to determine whether the land plot is incorporated or omitted. In some cases even when land plots do not match, RAN's land plots have a better definition of boundaries (see Figure 4), so they are used to update INEGI's frame. INEGI is currently in the process of determining the land plots of this segment that could be incorporated in the land plot shape.

Figure 4: *Review of RAN's Land plots using satellite image*

Furthermore, the review of the tabular file associated to the shape, allowed identifying the name of the producer along with the information about the "ejido" he or she belongs to, as well as the number of the land plot; however, the tabular file lacks producer's address (street, number, name of locality, etc.), which is a fundamental data to locate the producers during a census or survey. In this regard, INEGI is in the process of comparing the names with the names included in the Agricultural Census in order to associate the corresponding addresses in case of matches. At the same time, INEGI is requesting RAN to complete producer's information.

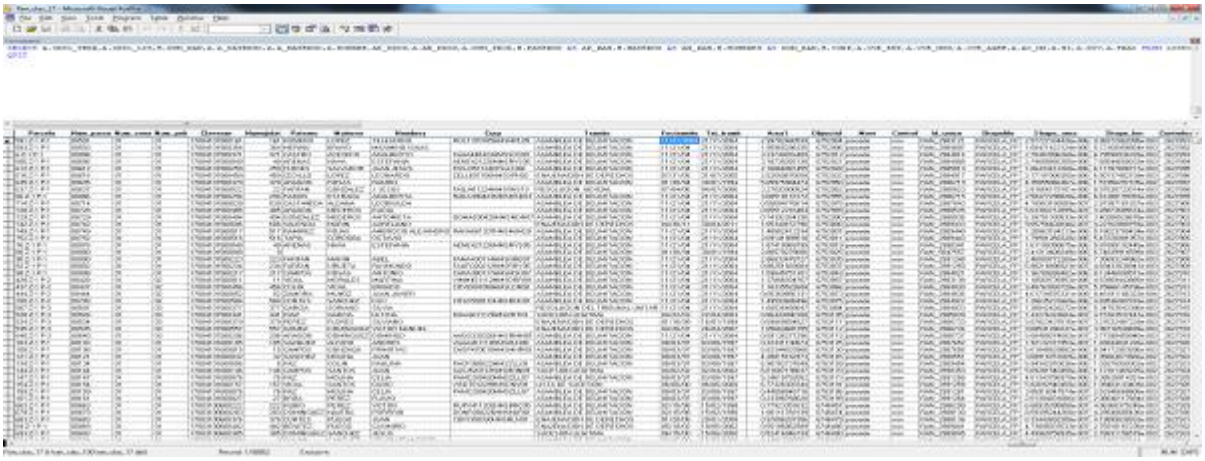


Table 1: *RAN´s information structure, without address data*

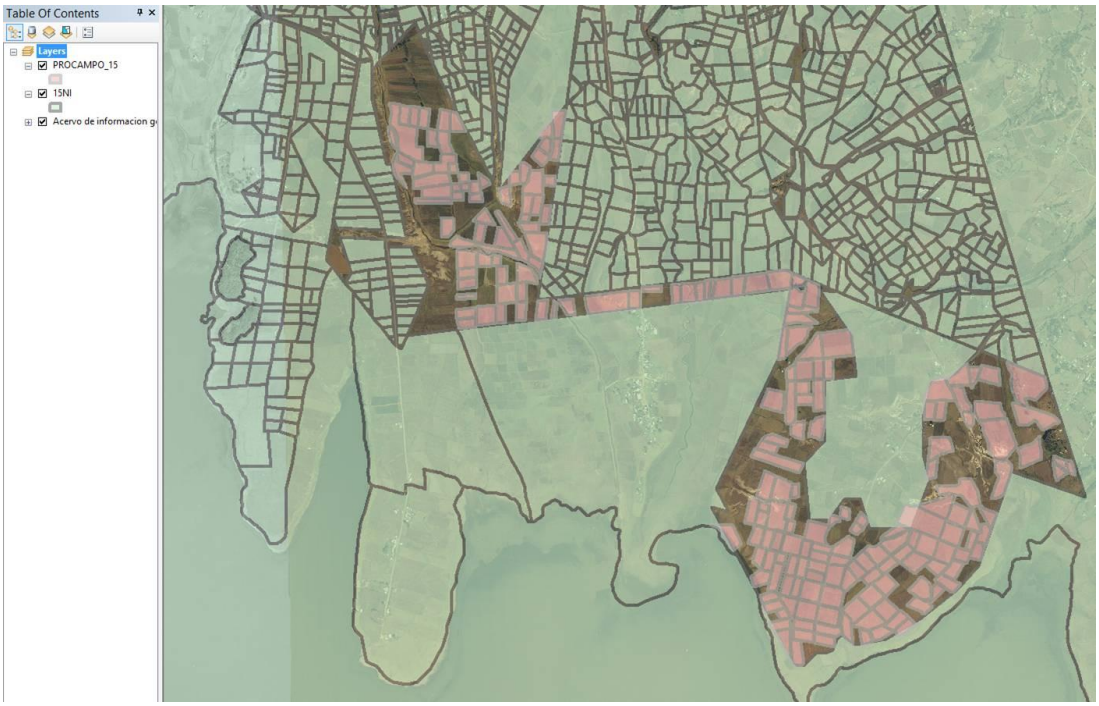
4. Analysis of information from PROCAMPO

Information generated by SAGARPA was also considered to be used, specifically data coming from the support programs for farmers that cultivate specific products in their lands. This is the case of PROCAMPO; to keep this register, SAGARPA identifies the land plot being supported; then determines its area and obtains the producer's information to whom the economic resources are being allocated to; the tabulate and cartographic information collected were also backed up in magnetic media, and it is also considered potential information for updating.

The review of the cartographic information revealed some differences in the configuration of the land plots; the most important was that data from PROCAMPO included land plots with boundaries classified in the Agricultural Census as "big land plots", without an inner subdivision. PROCAMPO's boundaries, in most cases, can be considered divisions of the Agricultural Census' original land plots, however there are some cases where PROCAMPOs' land plots corresponded only to the cultivated area supported economically by the program. In those cases, where minimal edition is required, will be reviewed before making the decision of integrating them or not.

Besides, there are land plots whose shape differs from the shape in the Agricultural Census, so a one to one detailed review is needed in order to decide which land plots can be incorporated and which are eliminated. This could be done by using the satellite image as background. Figure 5 shows a group of land plots from PROCAMPO that fill a "big land plot" of the census frame with detailed information.

Using the same criteria applied to RAN's information, from a total of 3 308 038 land plots received from PROCAMPO, 83.9% were already in the land plot layer of the Census of Agriculture; 8.3% corresponded to land plots that could be included with a few modifications; and finally, 7.8% were land plots that must be reviewed, one by one, and then decide upon their incorporation.



Figures 5: *Land plots from PROCAMPO (pink) to be incorporated into Census frame.*

In the case of the tabular information of PROCAMPO, the review reflected lack of data, although the directory contains the names of the producers that are supported by the program, and, in most cases, the street name and number. It also lacks the locality and the address, so INEGI is working with SAGARPA in order to obtain the complete information.

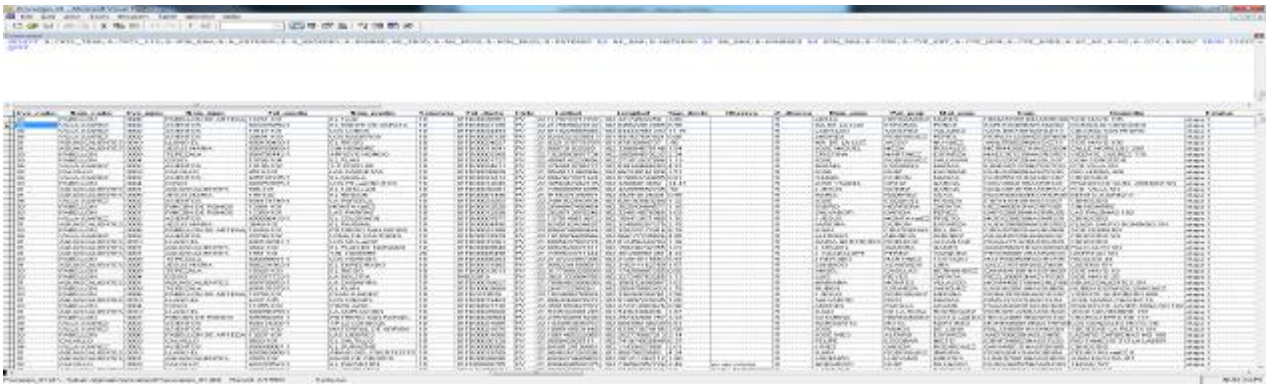


Table 2: *PROCAMPO´s information structure, without accurate address data*

Even though producers' names from PROCAMPO are not included in INEGI's information, they were used to update the information of the Census of Agriculture, since the information of the census corresponds to 2007 and PROCAMPO information is updated to 2012-2013.

5. Conclusions

The use of specialized software of GIS and high resolution satellite images, represents a huge advantage for analyzing, exploiting and managing cartographic and database information produced by institutions in the public and private sectors. Those advantages have been exploited by INEGI to compare vector data from different sources and to use them for strengthen the geostatistical framework of the Census of Agriculture.

Exploiting information from public institutions, as RAN and SAGARPA, represents a great opportunity for INEGI to update the list frame and the area frame of the Census of Agriculture; however, it is very important that information increasingly meets established standards in order to ease this task. The differences of the structure of the databases and the different reference systems of geographic and cartographic information are the main things to keep in mind while working with this kind of data.

With this work, it was possible to increase the percentage of production units (producers linked with their land plots) from 49% to 59%. Concerning the amount of land plots with delimited boundaries, the 295, 155 land plots that RAN added to the shape of the Census of Agriculture, compared with the 2 719 500 land plots that have not been delimited, corresponds to 11%.

Concerning PROCAMPO' data, the 189 853 land plots that were identified for possible addition in the area frame, are still in the process of detailed reviewing to decide whether they will be used to complement the shape of land plots of the Census of Agriculture or not. This amount represents 7% of possible land plots to be added.

REFERENCES

CONAPO (2010). Projections of the population in Mexico, 2010-2050

INEGI (2015), Results of the Intercensal Survey 2015-Sociodemographic Scope of Mexico 2015.

INEGI (2007), Glossary of de terms of the Agricultural Census 2007, http://www3.inegi.org.mx/sistemas/glosario/default.aspx?clvglo=cagf2007&c=12896&s=est

INEGI (2014), Statistical yearbook of the Mexican United States 2014.

http://www.inegi.org.mx/

http://www.snieg.mx/#top