FIFTH INTERNATIONAL WORKSHOP ON
BUSINESS DATA COLLECTION METHODOLOGY
19-21 SEPTEMBER 2018 – STATISTICS PORTUGAL, LISBON

USE OF COMPUTING MOBILE DEVICES IN THE ECONOMIC CENSUSES FOR UPDATING THE MEXICAN STATISTICAL BUSINESS REGISTER AND GEOREFERENCING ESTABLISHMENTS

Susana Pérez Cadena

México

INEGI

Background paper

Information obtained by Economic Censuses has several uses in public, private and social projects, since it is the only source in Mexico that presents, in high detail, the characteristics of the national economy, that is, it reports geographical data at different levels: country, state, municipality, locality, by basic geostatistical area, neighborhood, and even by sets of blocks; at sector level, it contains data of all the economic activities (except agricultural) in the country (around a thousand different activities), and thematically publishes around 1300 variables on economic units surveyed.

One of its most relevant uses is the updating of the Mexican Statistics Business Register, as well as the National Business Directory derived from it, the National Statistical Directory of Economic Units (DENUE, its acronym in Spanish). The DENUE is the most consulted product of INEGI (the National Institute of Statistics and Geography in Mexico), with more than 60 thousand accesses per month. It is a highly demanded product by users due to its characteristics:

- All businesses of the country are located there, uploaded in a GIS,
- Businesses are classified according to their economic activity, based on an international classification system,
- It contains data on every single business in the country concerning identification, location, contact, establishment's size, and the activity they carry out,
- Every business appears represented in the digital cartography, at the very precise place where it corresponds, identified by means of a dot. The dot appears in the block where it is located, but not only that, it is also in front of the corresponding block and closer to the corresponding place in that front.

With all these characteristics, the user is able to consult the exact subuniverse of his or her interest, by choosing the specific activities, specific sizes and/or the specific geographical zones he or she wishes to consult. And, since they are uploaded in a GIS, the user can both obtain the list and visualize them. That GIS also provides around 250 layers of information that help to complement the directory, such as layers of highways or railways, relief, bodies of water, population according to age range, handicapped population, households, among many others.

For making this possible, it is necessary to allocate each business with its "dot", and since data of Economic Censuses is the main source for updating the DENUE, during the collection of information the dot of each establishment is allocated (or updated).

This is done by taking a Mobile Computing Device (DCM its Spanish acronym) to the field, and using the digital cartography generated by INEGI. But the story of using a DCM it is not simple, and what comes next is a summary of such story.

The beginning of the story… Economic Censuses 2004

Economic Censuses are conducted every five years in Mexico, since 1930. 18 Economic Censuses have been carried out up to date, of which, only in the last three, the DCM has been used. Its utilization has had as a goal to provide the Economic Censuses with higher quality and efficiency, as well as higher timeliness for publishing results.

Experimentally, a DCM was first used in the census of 2004 for partially collecting data: 10% of the geographical areas were covered by means of a DCM. The DCM employed was a PDA (Personal Digital Assistant). That time, it was possible to compare the results obtained from data collected in paper versus data collected using a PDA, and improvements were observed in several aspects, mainly the fact that re-enquiries were done right at the moment of the interview since the system provided messages informing the interviewer if an incongruity had been found and that was immediately corrected or clarified with the respondent.

With this experiment, the project of the Mexican Economic Censuses became a pioneer in the use of Mobile Computing Devices for collecting census data. As of that project, other INEGI's censuses and surveys started to adopt the use of a DCM.

Evolution of the use of the DCM

Going back to 2004, the PDA had a very limited capacity that allowed to introduce only a few criteria for validation of the collected information, and it cartographically enabled to conduct only basic actions, such as selecting the block that each interviewer would cover, capturing the name of the roadways surrounding the block (not digitized then) and thus automatically allocate the block key and the name of the roadways to each census questionnaire, in a homogeneous way, for all establishments and dwellings in the block.

Even with the PDA limitations in capacity, **the objective of the experiment was achieved: to prove that the use of DCM was feasible and convenient for collecting census data.**

For 2009, once functioning was proven, the same PDA was used but now for all the data collection activities of the Economic Censuses and with a substantial improvement: the inclusion of the digitized cartography created by INEGI's own account. With this innovation, the interviewer was able to find his location in the field with the PDA and conduct the cartographic updating that was previously done on paper.

Probably the most important achievement of including the digital cartography inside the PDA was the possibility to allocate a dot for each establishment of the country in the cartography, representing its location. This way, user could have access to a product with the directory of all establishments in two perspectives: the list of establishments on one hand and their geographical representation on the other.

Geographic representation of an establishment required accuracy concerning being in the block where it is in reality and, once in the block, in the right front of street, and as close as possible from where it is exactly located in reality.

Before deciding to use digital cartography in the PDA for allocating the dot of each establishment, it was proven that this method provided better results than using a portable GPS: dot allocation using a portable GPS resulted in dots that were not necessarily located in the right block, but in front, or dots in the middle of the street, and it was uncertain if they corresponded to one block or the other. It also depended of not having high buildings that interfered with the adequate satellite signal transmission, or too many tress, among other problems.

Digitized cartography in the PDA, on the other hand, was self-sufficient for allocating the dot, since it contains all land features that allow the interviewer to have certainty about his or her location (line and key of the block, name of surrounding streets, median strips, parks, arbors, churches, schools, statues, position, among other features), and therefore allocating the dot with minimal error; it also has a development that does not allow to include a dot out of the digitized line that demarcates the block (in order for the dot not to be in the middle of the street, or in the middle of the block, for example).

All the above resulted in a high level of quality for geographical reference of establishments and dwellings, with a high confidence level about positioning of the dots in digitized cartography, much higher than the one that could have been obtained using a GPS.

This way, it was assured that dots were in the right block, in the right front of block, and in the right order in that front of block.

Additionally, using the PDA for the Economic Censuses 2009, allowed:

- That the interviewer visualized in the PDA the his or her whole corresponding responsibility area, therefore he or she did not need to have a map in paper
- To visualize the name of the streets that demarcated the block and, by clicking, selecting the name of the street where the establishment of interest was located (the name was already digitized)

- To conduct the cartographic updating directly in the digitized cartography,
- Allocating cartographic data to the questionnaires and counted dwellings, considering even the cartographic updates conducted, such as: name of roadway, block, neighborhood, zip code, locality, municipality, state,
- Registering the direction of the roadway.

The total of questionnaires was captured through this method (excepting Large Enterprises, that had the option to provide their answers to the questionnaires via internet), with huge success and with the aforementioned savings. Data quality did not discredit, on the contrary, the amount of required re-enquires decreased due to validating information at the very moment of the interview, although the number of validation was still small because of the limited capacity of the PDA.

For Economic Censuses 2014, with the development of new technologies, a new device for collecting information was used: a tablet-type device (which in reality was a subnotebook changeable to tablet) that added even more technological advancements.

INEGI aimed for higher working memory capacity, a bigger screen (although the DCM should have been light in weight) and characteristics for "heavy duty". The main technical characteristics of the DCM chosen in the Economic Censuses 2014 are:

- Central Processing Unit Intel Celeron 847 Dual Core with 2 Gb of RAM
- 1.72 kilograms of weight
- Battery with 6 hours of continuous operation
- Heavy duty physical characteristics (drop resistant case from a meter in height, portability, resistant to heat and weather humidity, discreet for avoiding temptations...)
- Power cord
- Light pen
- 11' touchscreen
- Windows 7

The characteristics of the DCM utilized in 2014 allowed improvements in capturing information, as well as in both data processing and transmission, that provided higher speed, agility, efficiency and quality when compared with the PDA from the previous censuses. Specifically, the characteristics of higher processing capacity and memory, and also a bigger screen size, enabled to incorporate the DENUE to the DCM, which allowed a follow up in field for each establishment included in the directory mentioned.

In this way, the tools added to the DCM were:

- Responsibility area for each interviewer
- Directory of economic units
- Questionnaires
- Digitized cartography
- Satellite images
- Catalogues of products
- Operative manuals
- Helping tools for the interviewer such as the possibility to open a touch keyboard; help regarding concepts in the questionnaire; calculator: option for backing-up transferring information, warning messages for the interviewer to verify the process being carried out
- Registration of advance of blocks and establishments
- Cartographic module system, operative routine, questionnaire's validation, economic classification, as well as data security's protocols for encrypted data, regulated by INEGI

With these benefits, the interviewer could locate, in a systematic and ordered way, the block he should walk by (inside the digital cartography), to obtain automatically geographic reference data such as state, municipality, locality, neighborhood, roadway, and allocate them to the questionnaire, also automatically.

Concerning the interview, the capturing system of the questionnaire enabled a set of filters according to the answers of the respondent, which helped the interview to be quickly and fluently developed, which avoided unnecessary or out of place inquires.

In regards to validation of economic information, the validation system was much more complete (due to higher memory capacity of the DCM), to review consistency and integrity of collected information. If inconsistency appeared or if information was missing, the screen of the DCM showed messages asking for some clarification right at the moment of the interview, which avoided future re-enquiries and annoyance for the respondent.

The operative control and advances were pretty detailed, both in geographical terms (from block up to national level) and per interviewer (head of field, zone, state or national). With this, a timely and accurate follow up to census data collection was given.

It was also possible to update the cartography from all the modifications found in the field (division, fusion, creation or elimination of blocks, changes in names and direction of roadways, openings and closings of roadways, etc.) in the digital cartography. In addition, all the lands with no establishments or dwellings were registered in the system and classified according to their use or what was found there.

All the above allowed georeferencing more than 5 million establishments that form the productive plant of the country, as well as almost 26 million of dwellings, and all the empty lands, during the Economic Censuses of 2014.

Resulting implications of using DCM (complications and savings)

Coming to this scenario had, of course, important considerations regarding the traditional way of collecting information in paper, that can be seen as disadvantages. The most important one is that the change from paper to DCM **implied to define and design, with a lot of time in advance, everything**: the questionnaire, validation criteria, operative strategy, logistics for collection, detailed planning, logistics of supervision and following up, training strategy, and all the systems. For example, when collecting in paper, the follow up systems can be finalized the very same day in which data collection starts, while in DCM, such systems should be developed along with the capturing system of the questionnaire, that is, highly in advance to the collection itself.

However, it clearly represents getting savings in the censuses projects, that have to do with a series of topics:

- Saving paper (and tress), and in printing questionnaires, as well as transporting them,
- Saving wages of persons that validate information in the field, since they would not be required,
- Saving wages of typists since they would not be needed,
- Saving wages of supervisors of capturing
- Savings in rents of spaces for capturing; and buying computers for that activity
- Savings in storing questionnaires for five years at least,
- Reducing the number of re visits for the interviewer, since re-inquiries can be done right at the moment of the interview, by using the validation criteria incorporated in the DCM,
- Among others.

When comparing all these savings against the cost that represents buying all the necessary DCMs, net savings are of 20%, not considering that the DCM bought are used later for many other projects during several years.

In addition to savings, and based on the census experiences described, there are other benefits that have to do more with the improvement of collecting processes and that result in higher data timeliness and quality. The following are highlighted:

- ✓ Homogeneity in the development of the interview
- ✓ Increasing speed and accuracy in obtaining and updating data
- ✓ Immediate transmission of information
- ✓ Higher timeliness for the treatment and subsequent publishment of information
- ✓ Facilitating control and following up of the data collection
- ✓ Facilitating staff training

Conclusion

The use of Computing Mobile Devices is wider each time, but using them in a census has been essential for saving resources and improving data quality, control of data collection, treatment of information, training, timeliness of results, as well as for generating new products such as the directory of establishments in the digitized cartography.

Changing from paper to DCM is a complex process but it worths it in all aspects regarding data collection of the significance of a census.