



Administrative Records and Over-coverage in the Statistical Business Register at INEGI

Enrique de Alba*

INEGI, Mexico City, Mexico – enrique.dealba@inegi.org.mx

Abstract

In Mexico, the generation of statistical information is carried out by several agencies across the government and it is coordinated by the National Institute of Statistical and Geographical Information (INEGI), an entity determined by Constitutional amendment in 2006 with full technical and operational autonomy. INEGI is ruled by a Board of Governors consisting of 5 members. This legislative change also established the Statistical and Geographical Information System (SNIEG). SNIEG is ruled by its own law (LSNIEG) enacted in April 2008. The SNIEG is divided in 4 Information Subsystems: Socio-demographic, Economic, Environmental and Geographical, and Judiciary. The LSNIEG establishes that, the Economic Information Subsystem must produce a National Directorate of Economic Units (DENUE) as part of the Infrastructure of the SNIEG; it has been carried out in Mexico since 2010. The Business Register is an essential element of it. Information is collected directly from businesses in the Economic Censuses, as well as from administrative records of some major government agencies, notably the Federal Electricity Commission (CFE), the Mexican Social Security Institute (IMSS) and the Tax Revenue Administration (SAT). Each one of these institutions has administrative records that include a large part of the businesses in Mexico. Due to the increase of administrative records that can potentially be used for statistical purposes, INEGI has been exploring the possibility of exploiting a series of administrative records to supplement or substitute traditional Economic Censuses. In addition to the fact that they are not designed for statistical purposes, one of the main problems is that there is not a unique code; each one may use a different code which makes it very difficult to integrate them. There is usually a lack of consistency in their concepts. Their use can easily lead to businesses being enumerated more than once because a given enterprise can appear in two, or more, registers with an apparently different address, or to businesses being enumerated at two addresses, like subsidiaries of an enterprise. Nevertheless each record separately can be used to check consistency of census results, especially at the local or state level. Special care needs to be taken so as to avoid over-coverage. INEGI has developed the Business Statistical Key (CLEE) as a unique identifier, and regulation has been passed to enforce its use by all producers of information in the SNIEG. The business register of Mexico includes 5 million enterprise/establishments.

Keywords: Business Statistical Key; Census; Infrastructure.

1. Introduction. The Legal and Institutional Framework for Mexico's National Statistical and Geographical Information System

As in many other countries, the production of statistical information is distributed in several agencies across the government and in Mexico it is coordinated by the National Institute of Statistical and Geographical Information (INEGI), an entity determined by Constitutional amendment in 2006 with full technical and operational autonomy. INEGI is ruled by a Board of Governors consisting of 5 members appointed by the President and ratified by the Senate. INEGI faces great challenges as an autonomous entity coordinating the National System. In the last eight years, INEGI has been transformed from an organization within the Treasury Ministry to be an autonomous entity, where it is expected to assume a leadership role to connect all the statistical providers while guaranteeing the integrity and objectivity of the information produced.



This technical and operational autonomy must follow the principles of free access to information, transparency, objectivity and independence. All these principles are relevant, as long as the integrity of the information supports not only the decision making processes, but provides elements to evaluate the government's performance and to follow-up on public policies.

This legislative change also established the Statistical and Geographical Information System (SNIEG), linked it to the National Planning System and in 2014 linked also to the National Council responsible for the Assessment of Social Public Policy. SNIEG is ruled by its own law (LSNIEG) enacted in April 2008. The SNIEG is divided in 4 Information Subsystems: Socio-demographic, Economic, Environment and Geographical, and Judiciary. Its mandate is to provide timely, accurate, reliable and relevant statistical and geographical information for policy design, analysis, evaluation or international reports. One of the key instruments considered in the law that defines the scope of the SNIEG is a Standards Compilation System that integrates all the rules, standards, catalogues, classifications and guidelines mandatory for all agencies while generating, processing, publishing and disseminating statistical information. It also includes the System coordination rules.

The LSNIEG establishes the components, objectives of SNIEG, the responsibilities of INEGI as coordinator of this system and as producer of key statistical and geographical information as well as the participation of other agencies. INEGI's Board of Governors had issued general rules and guidelines to enhance the coordination of the different agencies. Among its key elements the LSNIEG establishes, in its Article 23 that "SNIEG's National Economic Information Subsystem will have an information infrastructure containing at least a Geostatistical Framework and a National Directory of Economic Units (DENUE). The use of the DENUE, as well as the economic classifications that form part of it, is mandatory for the organization of administrative records, from which National Interest Information can be obtained."

Article 95 of LSNIEG complements the Art. 23 stating: "Individuals with business activities and legal entities, excluding those of the public sector, will be required to register in the National Directory of Economic Units and keep their registration updated, in accordance with the provisions issued by the INEGI.

Government units in charge of administrative records from which it is possible to obtain 'Information of National Interest' shall provide INEGI with the information it requires for the creation and updating of the National Directory of Economic Units."

2. The National Business Register of Mexico (RENEM)

In 2010 INEGI began the construction of the DENUE. It is intrinsically related to the RENEM. It contained statistical and geospatial information on all business establishments and firms obtained in the Economic Censuses of 2009. DENUE was originally made public in 2010 through the Geographical Information System, according to the geographical coordinates of business establishments. It was later extended as RENEM and published in 2015 according to international guidelines. Since then DENUE is the public version of RENEM.



Information is collected directly from businesses in the Economic Censuses and surveys, as well as from administrative records of some major government agencies, notably the Federal Electricity Commission (CFE), the Mexican Social Security Institute (IMSS) and the Tax Administration Service (SAT). Each one of these institutions has administrative records that include a large part of the businesses in Mexico. Due to the increase of administrative records that can potentially be used for statistical purposes, INEGI has been exploring the possibility of exploiting a series of administrative records to supplement or substitute traditional Economic Censuses and surveys. In addition to the fact that they are not designed for statistical purposes, one of the main problems is that there is not a unique code; each one may use a different code which makes it very difficult to integrate them. There is usually a lack of consistency in their concepts. Their use can easily lead to businesses being enumerated more than once because a given enterprise can appear in two, or more, registers with an apparently different address, or to businesses being enumerated at two addresses, like subsidiaries of an enterprise. Nevertheless each record separately can be used to check consistency of census results, especially at the local or state level. Special care needs to be taken so as to avoid over-coverage.

Mexico's Business Register. RENEM is the most important of economic statistics generation processes that contains information about the characteristics of the economic units located in the country. It is the only SBR that has a public component and one of the few that links statistical information to cartography; i.e. it is georeferenced.

Its goals are:

- a) To be the integrating element of the different sources of internal and external information, to enhance the statistical capacity as a whole, creating and taking advantage of synergies.
- b) To be the only updated source of information on economic units, to provide directories and sampling frames for the different economic statistics projects of INEGI and SNIEG.
- c) Follow up on each of the economic units, keeping a historical record of their changes, with the purpose of facilitating the conduct of longitudinal studies and analysis of business demography.
- d) Provide the database of active establishments for updating the DENUÉ and in the future directories from other agencies: the Federal Telecommunications Institute, the Ministries of the Environment or Energy.

It is to be the main element for organizing the generation of economic statistics and for linking information sources. It faces the great challenge of harmonizing and linking the data of the economic units of RENEM with administrative sources, see Figure 1.

RENEM has been developed and structured in accordance with the following international guidelines:

- a) UN Guidelines on Integrated Economic Statistics
- b) Eurostat-OCDE Manual on Business Demography Statistics.
- c) Eurostat Business Registers, Recommendations Manual.
- d) UNECE Guidelines on Statistical Business Registers.

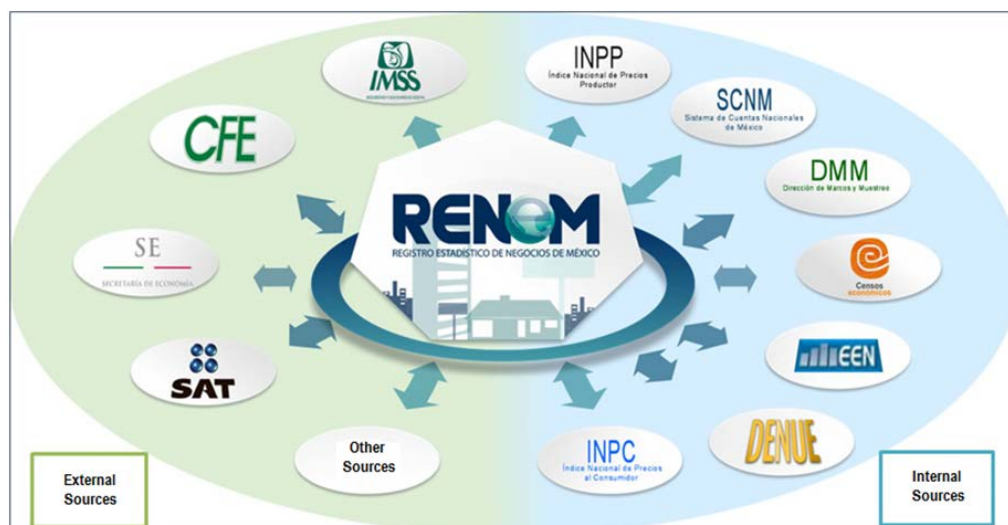


Figure 1

The main sources of over-coverage in the case of RENEM are the following:

- Different sources of information provide data for a single establishment, using different identifiers, which makes it necessary to carry out manual and automatic verifications using identification and location data to detect duplicities.
- Frequently, administrative records contain information on establishment that apparently are new, but that already exist in RENEM so that before incorporating their information it is necessary to carry out automatic and manual checks using their location and address.
- It also happens that administrative records from different external sources contain data on establishments that are not part of RENEM, and this must also be checked before including them

Based on LSNIEG, and in its role as central coordinating unit, INEGI has issued classifiers and regulation, called technical standards, that must be followed by the other agencies/institutions in the SNIEG. Their purpose is to standardize statistical and geographical concepts, in order to simplify and expedite the process of integrating RENEM, in particular to solve potential over-coverage problems. These make information exchange and comparisons possible. The main ones are:

- Agreement for the use of the North American Industrial Classification System (SCIAN) in the compilation, analysis and presentation of economic statistics.
- Technical Standard to incorporate and update information to the National Statistical Directory of Economic Units.
- Technical standard on geographical addresses.



3. Harmonization and linkage activities

In order to achieve the objectives set, INEGI works with the administrative sources on the following activities:

Homogenization: Conceptual and informatics.

Cleansing: Preparatory work for confrontation by name, address, colony, economic activity, etc.

Linkage: Association of RENEM information with data from external sources that refer to the same economic unit with the purpose of assigning a unique identifier in the different sources.

INEGI has developed the Business Statistical Key (CLEE) as a unique statistical identifier, and regulation has been passed to enforce its use by all producers of information in the SNIEG. It is a unique statistical identification key, assigned exclusively by INEGI to every one of the Mexican establishments and companies, which will be used to exchange information with the State Units.

CLEE Functions:

1. It uniquely identifies each establishment and company.
2. It links the establishments of the same company, indicating the type of economic unit (headquarters, branch, single).
3. Records changes in location, economic activity and size of data using a dynamic component
4. It allows control and monitoring movements of some variables useful for the business demography information.
5. It helps link information from other state units to allow information exchange

RENEM is updated completely every five years from the Economic Censuses and between them it is updated partially as follows:

- a) Monthly, using results and field work from national economic surveys.
- b) Annually, using administrative records from the Federal Electricity Commission, and information from the Tax Administration Service (SAT) on new establishments.
- c) Annually, by means of a field operative to update information on the largest establishments and firms in Mexico.

In the annual field update there are different activities for internal and external sources. For the first group includes the following:

1. Major establishments and companies in the country, measured by their total income, according to the Economic Census.
2. Visit to single establishment businesses and in the case of multi-establishment businesses to the companies' head office where their information is updated (births, deaths and changes of addresses of the branches).
3. New establishments (births) and casualties (deaths) from the National Economic Surveys and Special Surveys.

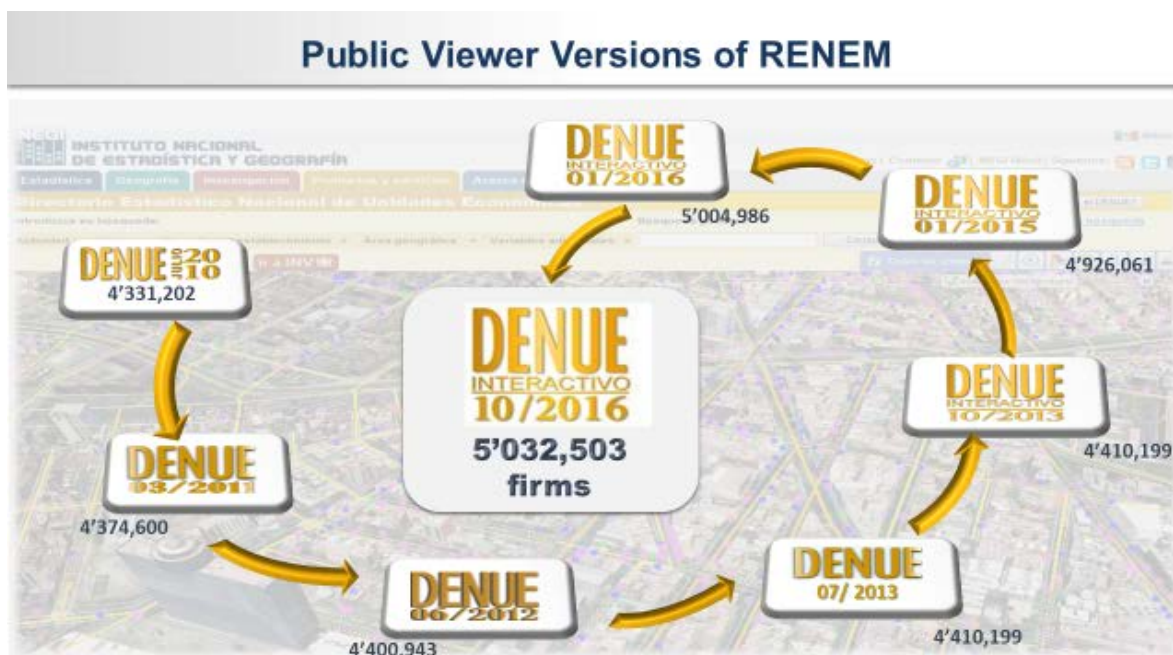


4. For the establishments and companies that are part of the samples of the National Economic Surveys, data is taken from the annual questionnaires provided by the respondent.

For the external sources:

1. Agencies are requested to update the RENEM databases and to carry out conceptual and informatics homologation activities.
2. Databases are received, then INEGI carries out information cleaning activities and confrontations with RENEM are carried out to detect new establishments.
3. The largest establishments are selected according to some characteristic of the economic unit, for example:
 - a. In the case of the CFE, those who hire a high voltage service and with high electricity consumption are considered.
 - b. For the SAT case, only legal entities are considered.

The business register of Mexico includes 5 million establishments/firms.



Variables contained in the database:

- Identification (Name and business name of the establishment, company to which it belongs)
- Location (Address and coordinates)
- Economic (Personnel, Income, Economic Activity, etc.)
- Demographic (Year of commencement of activities; for deaths: closure date, historical record of changes of: address, business name, company to which it belongs)
- Control (update date and source)