INF.2
12 April 2013

UNITED NATIONS
ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS

EUROPEAN COMMISSION
STATISTICAL OFFICE OF THE
EUROPEAN UNION (EUROSTAT)

ORGANISATION FOR ECONOMIC COOPERATION
AND DEVELOPMENT (OECD)
STATISTICS DIRECTORATE

Work Session on Statistical Metadata
(Washington, D.C. 8-10 May 2013)

Topic (i): Metadata standards and models

Geo-referencing statistical data: Geographical standards used in Mexico

Invited Paper

Prepared by Juan Muñoz López, National Institute of Statistics and Geography of Mexico (INEGI)

I. Introduction

INEGI is the coordinator of a National Statistical and Geographical Information System (SNIEG) that must provide valuable information to the Mexican society. In this role INEGI recognizes that he not the sole producer of information in the country, other public offices produce official information too. As all this information becomes congruent, comparable and interoperable its value raises and the nation can get more tools to take better decisions.

In the next section this paper examines the geographical standards used at SNIEG with emphasis in those that are used to geo-reference statistical data. In the road some basic concepts used in the field of geography are defined to provide a better context of the topic. In the following section, the document includes some examples of how those standards are applied to develop systems that shows information from a geographical perspective, giving an added value to the decision taker. Finally some conclusions are included, not with the aim of being a prescription of how to use georeferenced information but trying to open a path to create new ideas about the potential of this way of thinking about statistical information.

Note: The Mexican law which regulates the SNIEG uses the concept of "Technical Rules", but for the purposes of this review this is similar to the term of "Technical Standards", so they are used as synonyms in this document.

II. Geographic Standards Used in the Mexican System of Statistical and Geographical Information

A. The National Geostatistical Framework

When we georeference something, we define where is located in the physical space. To do that we use a system of coordinates or a map projection. In a coordinate system we use numbers called coordinates to uniquely determine the position of a point or another geometric element on a topological space. A map projection is a systematic transformation of the latitudes and longitudes of locations on the surface of a sphere or an ellipsoid into locations on a plane.

The "National Geostatistical Framework" (MGN) is a system used by INEGI for referencing information from censuses and statistical surveys to geographical locations using geographical coordinates. It divides

Mexican territory into parts with identifiable boundaries called "geo statistical areas". The framework contains three levels of disaggregation: State (AGEE), Municipal (AGEM) and Basic (AGEB), the last one is characterized as urban or rural.

An AGEB contain localities and blocks. A locality is an area which could have one or more dwellings and is recognized by a name officially or unofficially given. Localities are recognized as urban if they have 2,500 or more inhabitants or if they are heads of a Municipal area otherwise they are recognized as rural. A block is a geographical space with a polygonal shape of variable extent.

AGEBs are classified also as urban or rural. An urban AGEB is composed only of urban localities, in general terms they are conformed for sets of 1 to 50 blocks, and land use is residential, commercial, service, etc. A rural AGEB has a variable territorial extent is composed of urban localities, rural localities and extensions like swamps, lakes, deserts, etc., land use is mainly destined to agriculture and forestry activities.

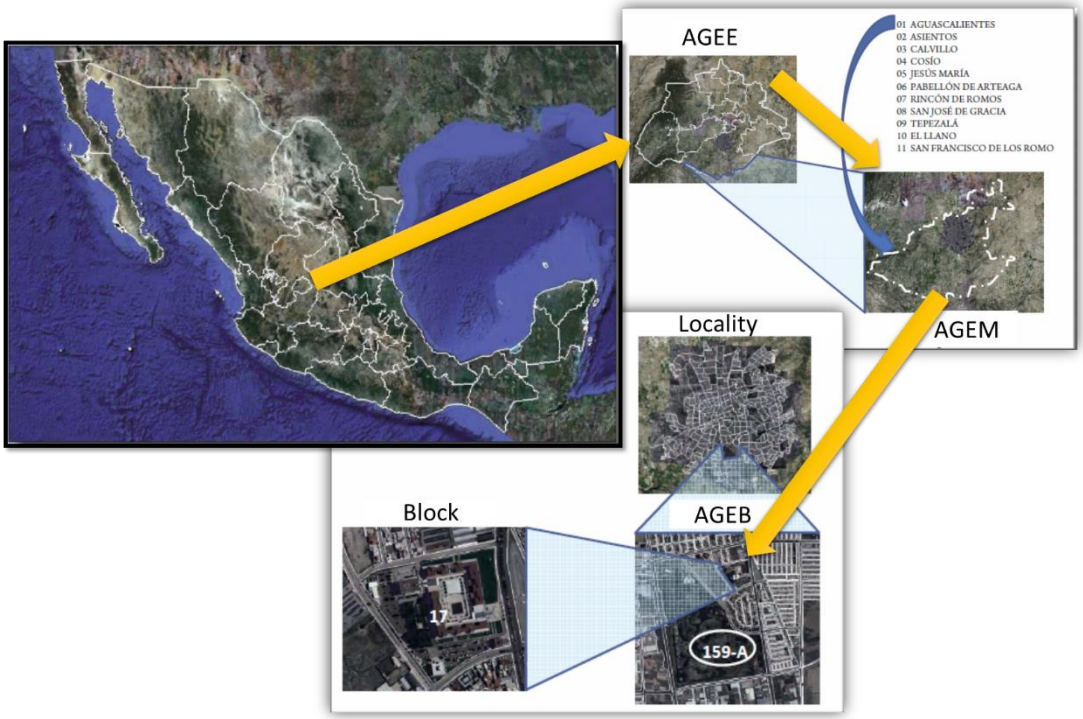

Figure 1; Disaggregation levels of the National Geostatistical Framework (MGN)

Geostatistical Framework information is an aid in the distinction between entities and municipalities, especially in places where political and administrative boundaries are undefined.

Each entity of the Framework has a unique numeric identifier that is assigned following a hierarchical path. Codification of urban and rural localities is made in the following way:

- Urban locality: **EE**+**MMM**+**LLLL**+**AAA-A**+**BBB**
- Rural locality: **EE**+**MMM**+**AAA-A**+**LLLL**+**BBB**

Where:

- EE, AGEE (State) represented with two digits
- MMM, AGEM (Municipality) represented with three digits
- AAA-A, AGEB represented with three digits a hyphen and a verification digit
- LLLL, Locality represented with four digits
- BBB, Block represented with three digits

MGN contains code lists and polygons for each disaggregation level. In Figure 2; Example of codes and Polygons of municipal areas from the MGN a map for the Municipal level is showed along with some codes of the municipalities. Each municipality has its own polygon, one of them is presented in red to show the municipality of "Ensenada" which is identified by the code "02-001".

The polygons of the MGN are stored in shape format (shp). Each shape file contains several physical files that will store the geometric, alphanumeric, datum and projection information.

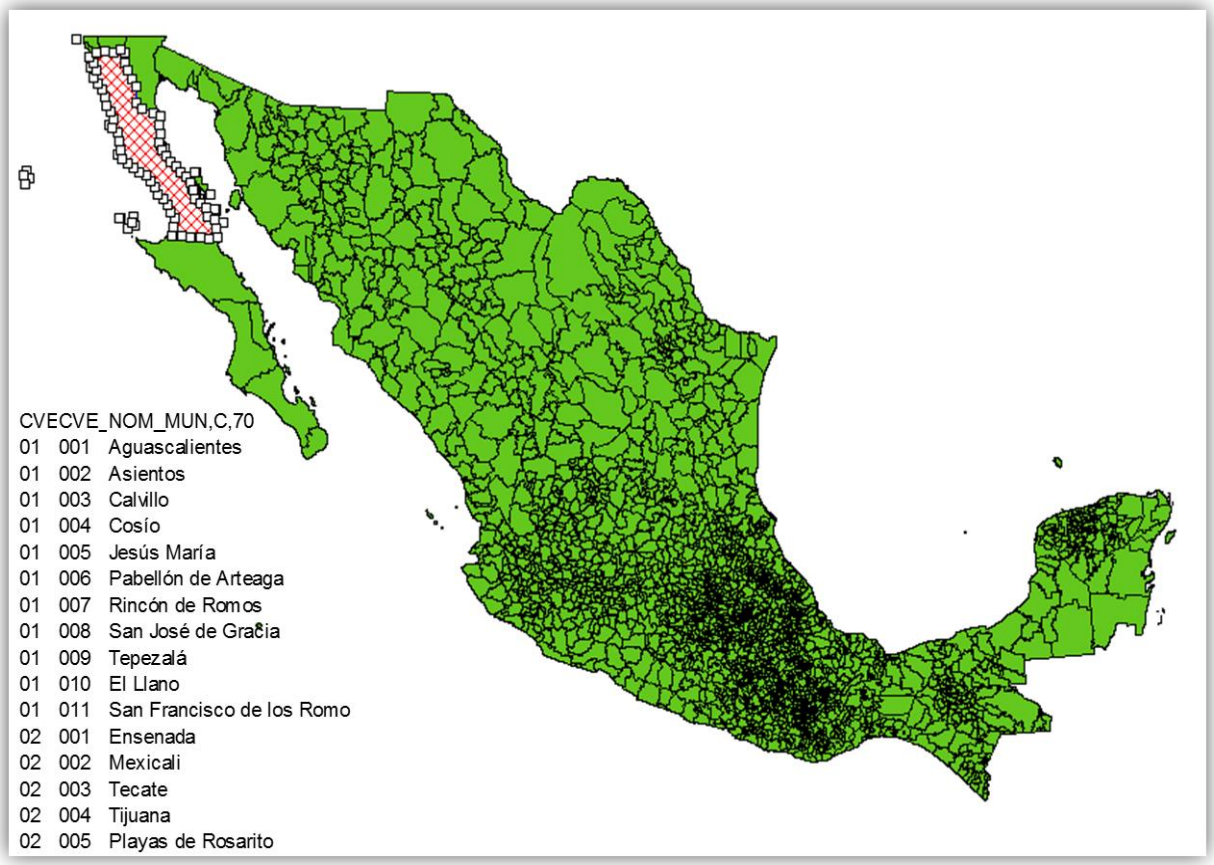

Figure 2; Example of codes and Polygons of municipal areas from the MGN

The graphical representation of the MGN is concreted in a set of maps, sketches, plans and catalogs at different geographical levels of representation which is known as the "Statistical Cartography" and it's used to support the activities of planning, execution, processing and dissemination of results of censuses and surveys carried out by INEGI. Maintenance and update of this cartography is permanent.

Geographical coordinates must be referred to a geoid, which is a mathematical model to represent the Earth. The Statistical Cartography of the MGN uses the standard model called ITRF92 epoch 1988.0 as is datum. It is generated by the International Reference System (ITRS) and is widely used to support navigation systems. The projection used to represent the maps is the geographic projection known as Lambert Conformal Conic (CCL).

INEGI has established that the official reference datum to be used for geodesic measures is ITRF08 epoch 2010.0, so currently, Statistical Cartography is in an update process to change coordinates from ITRF92 to ITRF08.

B. Technical standard for the generation, collection and integration of cadastral and registry data for statistical and geographical purposes

Cadaster is a systematically organized public inventory of each one of the land properties of a country. It contains a set of registries showing the extent, value and ownership of the land.

In January of 2012, INEGI published the "Technical standard for the generation, collection and integration of cadastral and registry data for statistical and geographical purposes". The use of this standard is obligatory for any institution that generates, collect or integrate cadastral or registry data. The objective of this standard is to harmonize and homogenize the information of this kind that is integrated to the National Statistical and Geographical Information System (SNIEG).

Each land property has a unique identifier conformed by 31 characters codified under a hierarchical structure. The components of a cadastral code are:

- EE, State represented by two digits
- RRR, Cadastral region represented by three digits
- MMM, Municipality represented by three digits
- ZZ, Cadastral zone represented by two digits
- LLLL, Locality represented by four digits
- SSS, Cadastral sector represented by three digits
- BBB, Block represented by three digits
- PPPPP, Property represented by five digits
- DD, Building represented by two digits
- UUUU, Unit represented by four digits

The standard describes the data and metadata that must be included in any cadastral database. Accordingly with the standard, each cadastral element must contain a polygonal figure and the necessary information to locate, describe and represent it.

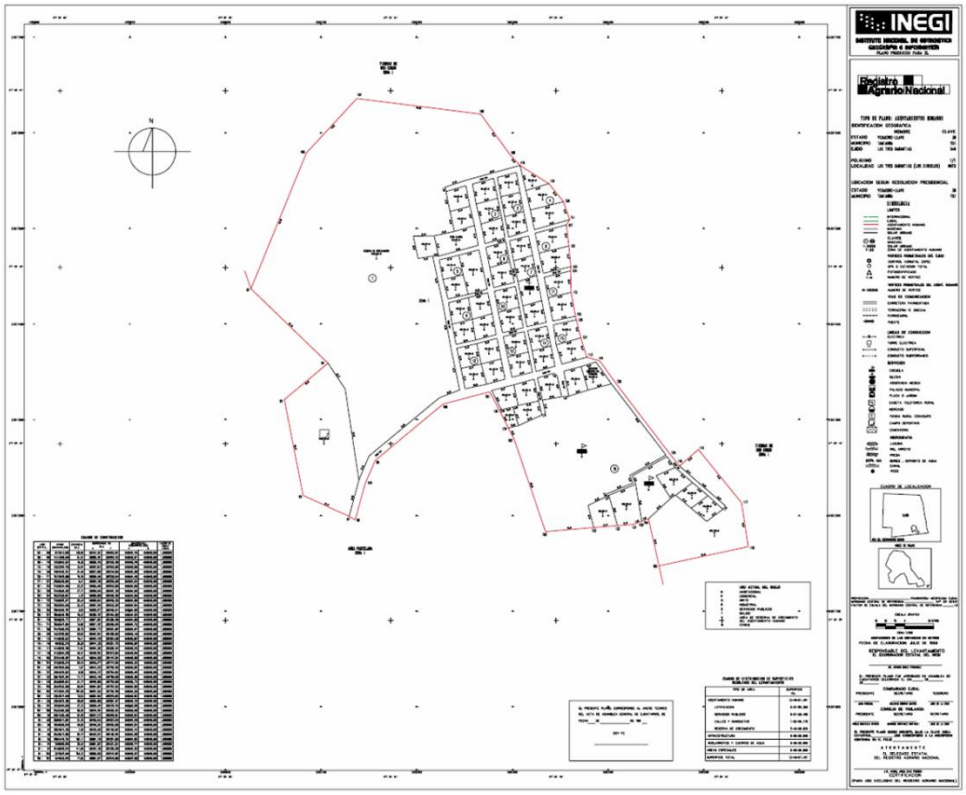

Figure 3; Representation of a communal agrarian property

Derived from the cadastral information INEGI generates several statistics like the indicators of the modernization of cadasters on each Estate, the state of the social properties, etc.

C. Technical standard on geographic addresses

A geographical address is the space within a town or referred to a communication channel that occupies a building (construction or land) where can be set one or more persons or economic units. The main objective of the Technical Standard on Geographic Addresses is to establish the component specifications and characteristics of the information that conforms a geographical address which will serve to identify a property.

Administrative records will integrate these structured, standardized and consistent addresses that are compliant with this standards; the aim is to improve quality, comparability and interoperability of them.

The standard defines that a geographical address must contain the following components:

- Spatial: street, highway, road
- Reference: external number, internal number, human settlement (neighborhood, quarter, etc.), zip code, location description (additional information to locate the address)
- Geo-statistical: AGEE, AGEM, Locality

Figure 4; Components of a geographic address

This technical standard is mandatory for INEGI, as well as the state units (a state unit is an office that has official attributions to develop statistical or geographical activities or have administrative records to produce information of interest for the nation) whose are responsible for capturing, updating and integrating administrative records containing geographical addresses, by themselves or through third parties when such activity is entrusted to them.

To make this standard operable, INEGI delivers in his site the code lists of the components, the information model to implement a conformant database of addresses, the SQL script to build it and a manual for the application of the standard (http://www.inegi.org.mx/geo/contenidos/normasTecnicas/dom_geo.aspx).

D. Other geographical standards issued by INEGI

Other standards that INEGI has published attending to its coordination responsibilities in the geographical information field are:

- Technical Standard for the National Geodesic System: This is the basis of the geodesic reference framework used in Mexico. It defines how the latitude, longitude and high must be measured accordingly to the datum adopted by INEGI (ITRF08 epoch 2010.0) which is associated to the reference ellipsoid GRS80. For cartographic purposes this standard is compatible with WGS84 and SIRGAS and the formulas to make the necessary translations between 3D and 2D systems are described. The national geodesic system is also defined by its passive and active networks which are described in the standard.

- Technical Standard for the Elaboration of Geographical Metadata: This standard defines 149 metadata concepts, 25 of them are mandatory and 124 are optional. This standard was created to make easier the coordination of how to document the metadata of geographical data sets. This standard is based in other standards which have been internationally adopted, like:

    - FGDC-STD-001-1998: Federal Geographic Data Committee -Standard-Version 001 - revised June 1998.
    - ISO 19115:2003: International Organization for Standardization 19115.

    The information that is considered in this standard is categorized in the following groups:

    - Metadata information: Data about content, quality, condition and other characteristics of the geographical data.
    - Identification of the set of spatial data or product: Basic information to find and classify in a unique way the data set or product.
    - Dates related with the set of spatial data or product: Dates related to the production of the information.
    - Information about the source responsible of the set of spatial data or product: Contact information about the organization and the specific people that has produced the data set or product.
    - Geographic localization of the set of spatial data or product: Mechanism used to represent spatial information and establish the geographical position of the set of spatial data or product.
    - Reference system: Description of the horizontal and vertical reference system used to determine the coordinates. It includes the domain references or values needed to make the codification.
    - Information quality: General valuation of the set of spatial data or product. This section includes some criteria about the data and the production process.
    - Entities and attributes: Summary and list of all the data elements included.
    - Dissemination: Information about considerations and restrictions that must be taken into account to distribute the set of spatial data or product.
    - Contact information for questions about the metadata: Reports the version of the standard used to document the metadata and the contact information about the person who was responsible of doing it.

- Technical Standard for Positional Exactitude: This standard sets the technical specifications to make geodesic measurements and how to represent and document them in order to be certain of its accuracy level.

III. Some Examples of Systems Developed by INEGI Related With Georeferencing Statistical Data

A. Mexico in Numbers

Contains information from national to municipal level with information of different topics (economy, environment, demography, society, government) integrated from various censuses and surveys). That information can be visualized in charts or maps. Maps can be thematic, cartographic or from satellite.

(http://www3.inegi.org.mx/sistemas/mexicocifras/default.aspx).

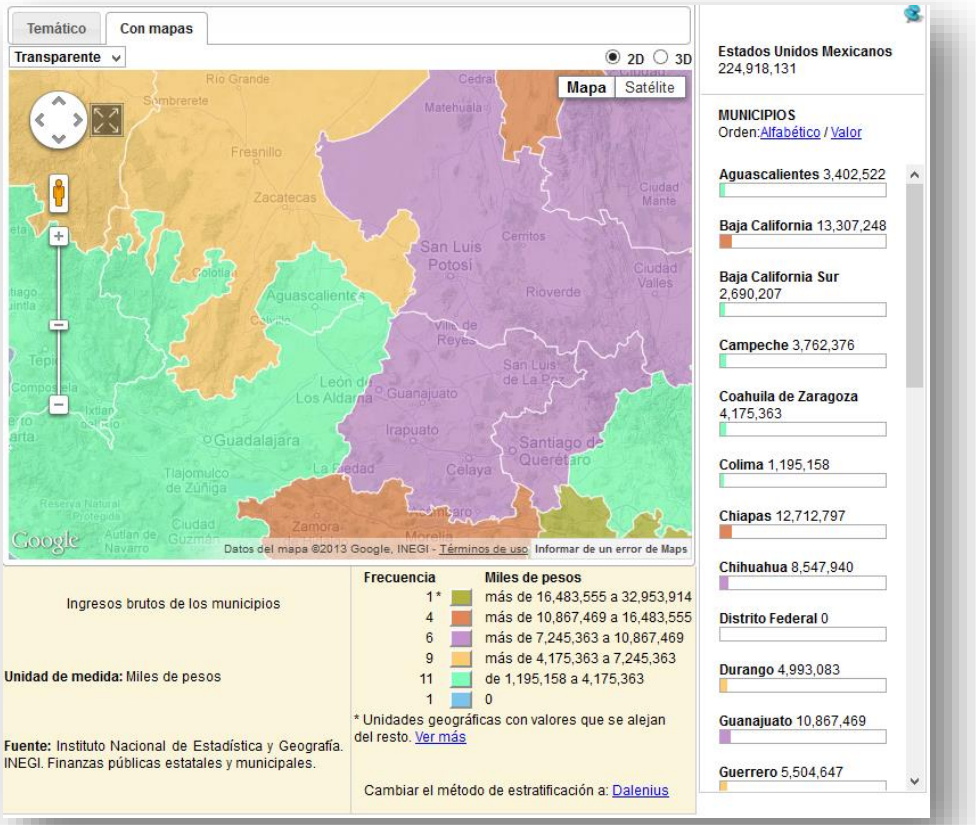

Figure 5; Mexico in numbers

B. National Statistical Directory of Business Units (DENUE)

This system delivers information about the identification and localization of the 4'400,943 business units that are active in Mexico accordingly to the data obtained in 2012. Is necessary to say that updating of the contained information is permanent and that is annually verified.

The observation units of the system are at two levels establishment and enterprise. All information has been geo-referenced using the National Geostatistical Framework.

The system allows user to select establishments by economic activity, size (strata based on employed persons) and geographical area. The system let the user to see the maps of the selected subset of data. By georeferencing business units at block level and street frontage, offers the possibility of dimension the distribution of economic activity in the geographic space chosen by the user.

(http://www3.inegi.org.mx/sistemas/mapa/denue/default.aspx)

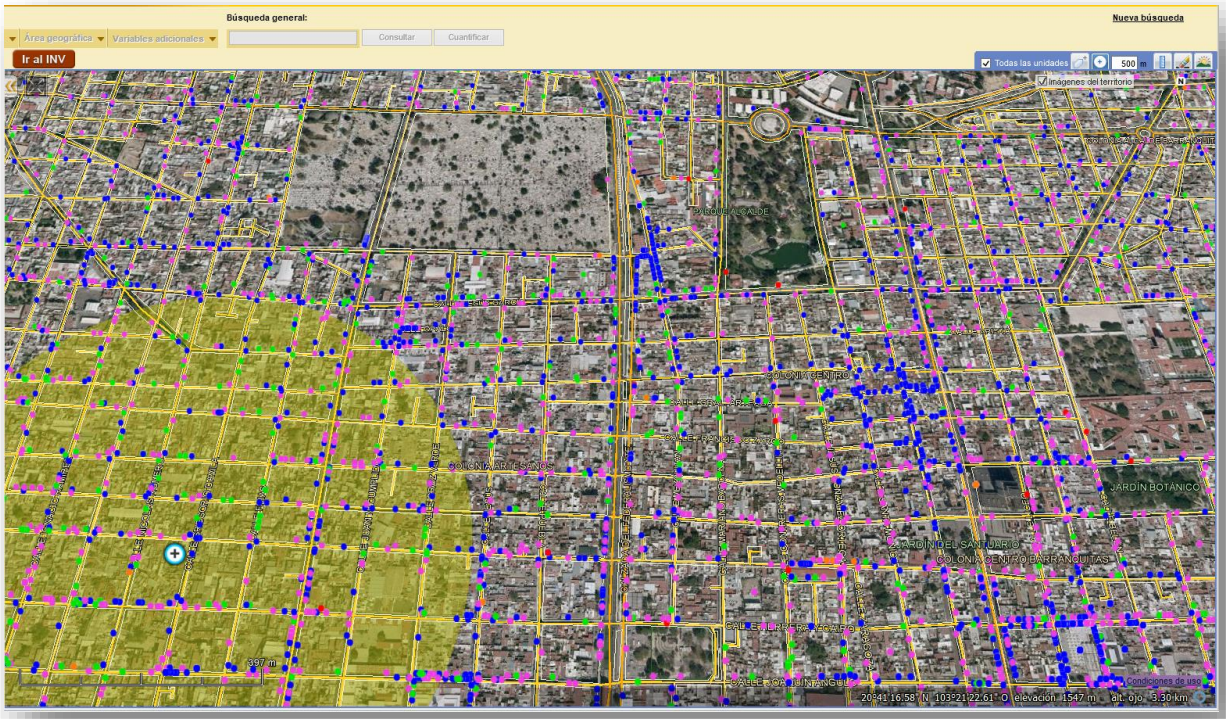

Figure 6; National Statistical Directory of Business Units (DENUE)

C. National Inventory of Dwellings (INV)

Contains data related to dwellings, population and urban environment referred to the National Population and Dwelling Census of 2010 and the grown between 2010 and 2012. Data of this system is on permanent updating.

The system has the same capabilities of DENUE but also incorporates several geographical layers to complement information from the Census giving more tools for the decision taking. From this system DENUE can be accessed (and vice versa), in this way the study of economic condition of certain area is facilitated.

INV has incorporated 12 dwelling indicators and 7 of population. Data is presented as absolute numbers an as percentages. To be complaint with the confidentiality principle it doesn´t display information for geographical unities with less than three blocks, on those localities the system only show the quantity of dwellings, the quantity of occupied dwellings and the total population.

About the urban environment, the system is presented by block and vitality. There are 14 indicators divided in three groups: vial infrastructure; equipment and services; and access and commerce on the public roads.

(http://www3.inegi.org.mx/sistemas/mapa/inv/Default.aspx)

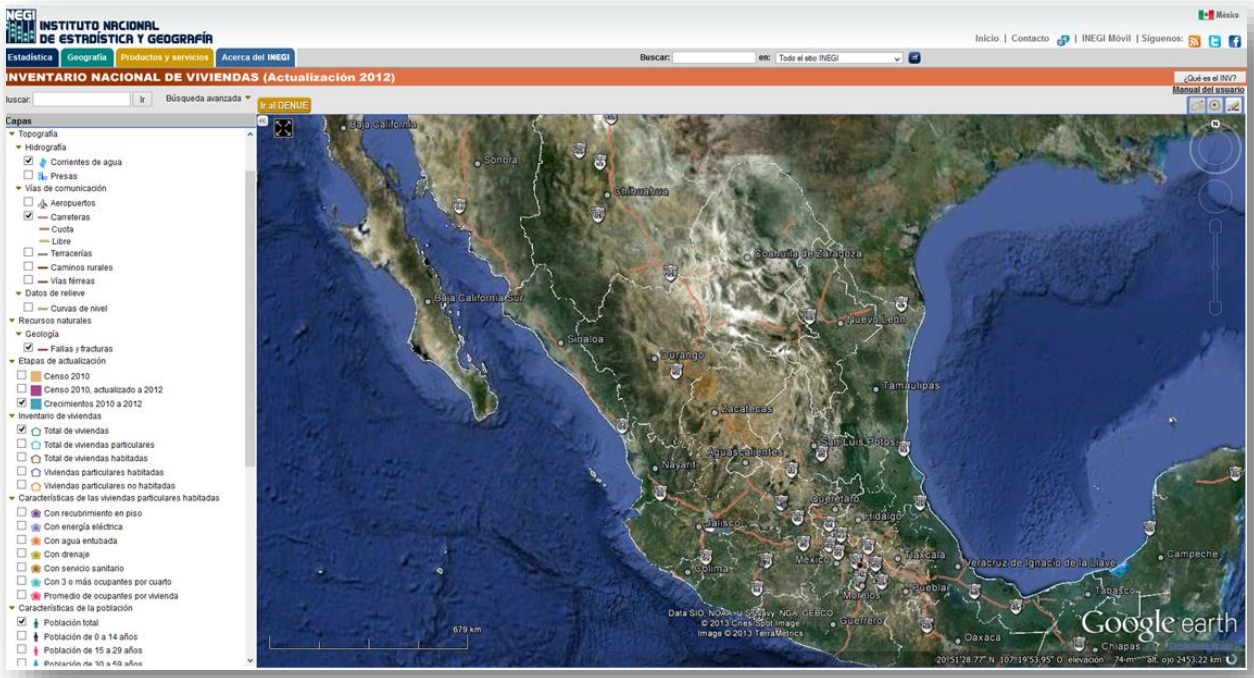

Figure 7; National Inventory of Dwellings (INV)

IV. Conclusions

Establishing a solid and congruent set of geographical standards is one of the first steps to develop a solid infrastructure of information to georeference statistical data.

The integration of statistical and geographic information can be a powerful way to present a rich set of information with meaning and context that has an added value for decision taking.

Access to big databases becomes useful to the user who can access data that is useful not only for governments or big organizations but for individual users who want to know better their own neighbourhood, and for small entrepreneurs needing to make a feasibility study to open a business.

For governments geo-referenced information is a valuable resource to make analysis and establish political policies. The richness of the information provides a context better suited to develop programs of greater impact applying resources in a more accurate way. Impact of demographic and economic phenomena on the environment can be analysed and then used to develop solutions to solve problems that are present or could be materialize in the future.