



Economic and Social Council

Distr.: General
15 September 2014

Original: English

Economic Commission for Europe

Conference of European Statisticians

**Group of Experts on Population and Housing Censuses
organized jointly with Eurostat**

Sixteenth Meeting

Geneva, 23-26 September 2014

Item 5 of the provisional agenda

Operational aspects of censuses

Innovative products for disseminating geographic information

Note by the National Statistics Institute, Madrid, Spain

Summary

One of the most important issues of 2011 Spanish population and housing Census is that it has collected, for the first time, information about the geographical coordinates of the buildings.

This information has been used for developing new products that connect geographical data with census topics (demographic, economic, educational...). Furthermore, users will have the option of obtaining data associated to a 1km² cell and independent from the administrative boundaries.

Possibilities of new ways of disseminating census information are great, and it will be more visual and easier to transmit it to all kind of users. Conclusions, difficult to reach without the geographical data, can be obtained at a glance by looking at the suitable product.

I. Introduction

1. Every day it is more common to see information summarized in maps or graphics in the different types of media like televisions, newspapers or Internet.

2. Visual products have several advantages:
 - (a) First of all, users can understand in an easy way the information transmitted, with very few or no explanations.
 - (b) Furthermore huge amounts of data can be condensed into a simple picture. Sometimes it would be impossible to sum up all this information in a textual format.
 - (c) And the impact that they have in our memory is normally greater compared to written information. Most of us are able to retain in our brain some graphics, especially if the contents are impressive.
3. In addition, it seems clear that the development of technology has helped to raise the amount of information that appears in the media with a visual format. Nowadays, it is possible to process much greater amounts of data than several years ago, and the software programmes of graphic design have also increased their possibilities.
4. All these general comments have even more importance in the context of statistical information, where the numeric data appear everywhere.
5. INE-Spain was aware of this new reality and taking advantage of the execution of the Census in 2011 to collect, for its first time in history, the geographic coordinates of all the buildings in the territory, with which a new range of dissemination products were possible.
6. Among the different tools developed, users can discover modern features, for example:
 - (a) Using the geographic coordinates, it is possible to make analysis by cells, independent from traditional administrative divisions.
 - (b) Selecting a geographic area directly from the map, and then a guided menu will intuitively ask about the details of the query.
7. This paper tries to give an overview of the details of these products.

II. Products developed

8. The amount of information that an operation like the Census produces is much higher than a conventional statistical survey. Moreover, the number of users of any sort (professors, researchers, employees of public companies, consultants...) that are interested in the Census is also very important.
9. If we consider these premises, it seems quite reasonable to develop different types of Census products depending on the user's ability, and the details (geographical and number/type of variables) of their requests. INE-Spain made several products available to users:

A. Predefined tables

10. Predefined tables, with fixed variables and breakdowns, are the most classical way to disseminate statistical information. In this occasion only a few sets of tables (it is possible to obtain all the information through other products) among all the possibilities were included in our webpage.

11. Users can obtain tables for different types of administrative boundaries like national (NUTS1), autonomous communities (NUTS2), provinces (NUTS3) or municipal¹ (LAU2) from each one of the following concepts:

- (a) People
- (b) Family nuclei
- (c) Households
- (d) Dwellings
- (e) Buildings
- (f) People living in collective living quarters

12. Additionally, in some cases, several municipal rankings (i.e. Municipalities with the highest/lowest amount/percentage/rate...) were also included. This innovative product was appreciated very much by the media.

13. Predefined tables are suitable for those users that want a “panoramic view” of the Census information without bothering to analyse the details.

B. Indicators

14. Together with these tables, INE-Spain has created a set of 145 main indicators. These indicators contain data about people, dwellings or households. Some examples of the information that can be consulted are:

- (a) Foreign people aged less than 16.
- (b) Dwellings, with surface between 61 and 75 m².
- (c) 3 people households

15. Indicators are disseminated from two points of view:

(a) Administrative criterion: Spain is divided into almost 36,000 census enumeration areas.

(b) Geographic criterion: Spain has a surface area exceeding 500,000 km². Indicators are calculated for those 80,000 cells of 1km² with at least one dwelling² inside.

16. Users can download indicators files both in excel or csv format. In addition, cartography files (shape files from both enumeration areas and 1km² cells) can also be freely and directly downloaded from INE³ web page.

17. Indicators are recommended for those users that want some quick facts at a very detailed level.

C. Microdata files

18. INE-Spain has elaborated two 2011 Census microdata files:

¹ Depending on the table, information is provided for all the municipalities or only for those bigger than a certain threshold.

² Only 63,500 of the total amount of 1km² cells have population (at least one person) inside.

³ See the section named “Advanced query systems” in the following link:

http://www.ine.es/en/censos2011_datos/cen11_datos_resultados_en.htm

(a) Population file: It has 4.1M records. Each record has information (138 variables) about one person and its family nucleus, household, dwelling and building. Apart from that, some variables about the father, mother and partner of the person are also included.

(b) Dwelling file: Contains 2.3M records (26 different variables in each record) with information about all the dwellings and the buildings they belong to.

19. Depending on the size of the municipality, territorial detail identification⁴ is masked or not. Municipalities with less than 20,000 inhabitants are grouped in several strata into their corresponding NUTS3 region.

20. Furthermore, other considerations related to the microdata files are:

(a) Information about army members is not included. Workers that belong to this category are aggregated into another one.

(b) Detailed occupation (3 digits) and industry (3 digits) variables are not included. In both cases, 2 digit variables appear in the microdata file.

21. Microdata files are suggested for those advanced users (often researchers or university professors) with computer skills⁵ that want to deeply analyse Census variables.

D. The Census Dissemination System in Spain (CDS)

22. Dissemination of the 2001 Census was a success, basically because the main product (with almost 3 million accesses) was designed in a very user-friendly way, and it was easy for users to find the information they needed.

23. Experience from 2001 was very useful and a similar system, with some important new developments, was designed for the 2011 Census. The main features of the CDS are:

(a) Online web product. Only a browser is needed to consult all of the data.

(b) It is possible to consult information related both with administrative boundaries (nation, autonomous regions, provinces, municipalities, districts or enumeration areas) and also with cells, which is one of the biggest novelties of this Census.

(c) Guided menus to obtain all the information

24. Several general considerations about how the CDS deals with confidentiality⁶:

(a) Information is rounded to the closest integer multiple of 5.

(b) All the cells that contain less than 10 sampling units are marked with an asterisk⁷.

(c) Results are disseminated only when the quotient $m/d \geq 10$ is satisfied⁸

⁴ This affects to the variable “place of residence” but also to other geographical variable like “location of place of work”, “location of place of study”, “place of usual residence 10 years prior to the census”...

⁵ Information is disseminated in compressed .txt files. In order to carry out analysis of the information, statistical software like SAS, R or SPSS is highly recommended.

⁶ Some issues are common with other types of products.

⁷ INE-Spain is considering relaxing the rule and reduce the threshold from 10 to 5.

⁸ “m” is the amount of sampling units. “d” is the amount of breakdowns of the table corrected by a sensibility factor (usually variables have a sensibility factor of 1, but in some occasions, depending on the variable, it can be 2 or even 3).

25. The CDS system is a suitable tool for all types of users. Because of its simplicity and all the different possibilities and tools that contains, it can be used by users, ranging from those with no census experience, to experts with detailed knowledge. Users can interact with the CDS from two points of view (thematic and geographic navigation) that are explained above.

D.(i) CDS thematic navigation

26. The CDS thematic navigation system allows users to create personalized queries. Its main features are:

(a) Dynamic creation of tables. First of all, users select the geographical area from a predefined list. Secondly, the scope of the request: People, dwellings, households or family nuclei. Finally the variables of interest, using an intuitive tree list plug-in.

(b) Variables can be allocated both in rows or columns. The maximum number of variables that can be crossed in the same table is 6.

(c) Different types of units of measure (some examples: people, average age, percentage of foreign people, percentage of unemployed people or percentage over total).

(d) Heavy queries are derived to a batch processing system in order not to overload the system. Once the query is done, an email is sent to the user with the data download link.

(e) Advanced tools (most of them only available once registered): creating and saving filters (users can select only some breakdowns of the variables), table rotation, sorting data, saving the query, possibility of creating personalized breakdowns...

Figure 1: Selection of variables with the CDS thematic navigation

The screenshot displays the CDS thematic navigation interface. At the top, there are four steps: Step 1 (What do you wish to do?), Step 2 (Select the geographical scope), Step 3 (Select collective), and Step 4 (Create the table and establish filters). Below these steps are buttons for 'Create tables', 'Help', 'Glosary', 'National', 'Autonomous Community', 'Province', 'Municipality', 'Inframunicipal', 'Persons', 'Dwellings', 'Households', 'Couples and other family nuclei', 'Design table', and 'Filters'.

The main interface shows the following configuration:

- Geographical Scope:** National
- Colective:** Resident in main dwellings
- Select variables:** A tree view showing 'Resident in main dwellings' expanded to 'Basic demographic data', which includes 'Sex', 'Year of birth', 'Marital Status', and 'Age'. 'Age' is further expanded to 'Age in large groups', 'Age (fortnightly groups)', and 'Age year to year'. 'Age in large groups' is selected.
- Rows:** Sex
- Columns:** Age in large groups
- Select unit of measure:** People, Average age, Percentage of people under 16 years, Percentage of people between the ages of 16 and 64.

Buttons for 'See Table' and 'Apply and go to Filters' are visible at the bottom. The text 'Personal queries' is located at the bottom right of the interface.

(f) Other application functions are the presence of a glossary, with explanation of the Census concepts or a help document.

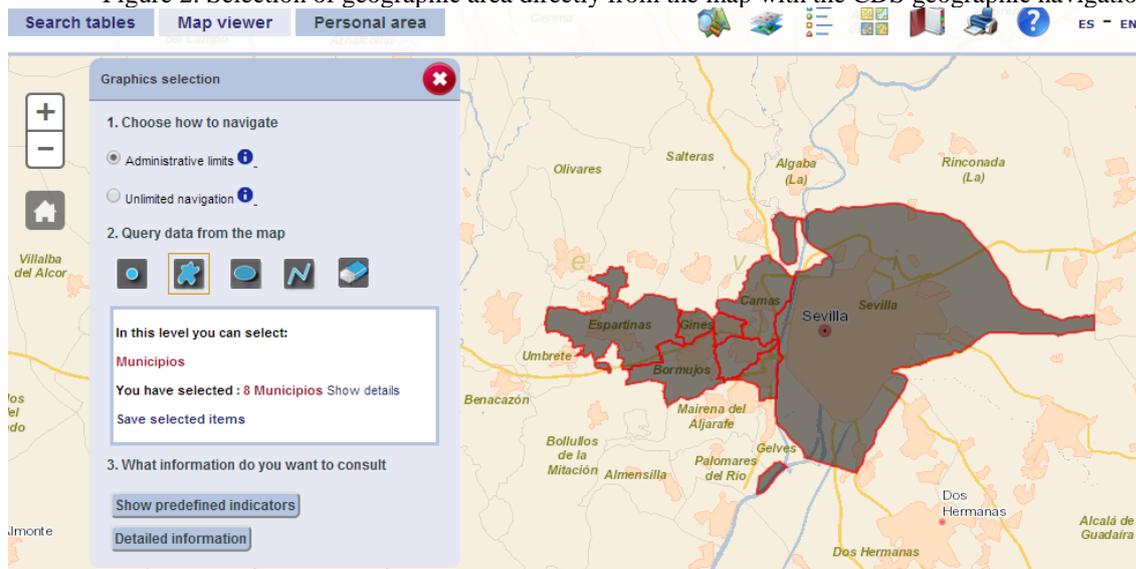
(g) Information is shown directly in the screen, and can be downloaded in different formats like Excel, xml or pc-axis.

D.(ii) CDS geographic navigation

26. The most important novelty of the 2011 Spanish Census is the CDS geographic navigation system. It allows users to analyse information from a visual point of view⁹ and its main characteristics are:

- (a) Users can consult information directly from the map through two different procedures:
 - (i) Selecting different administrative units continuous or not (from the whole nation to enumeration areas) and creating a query.

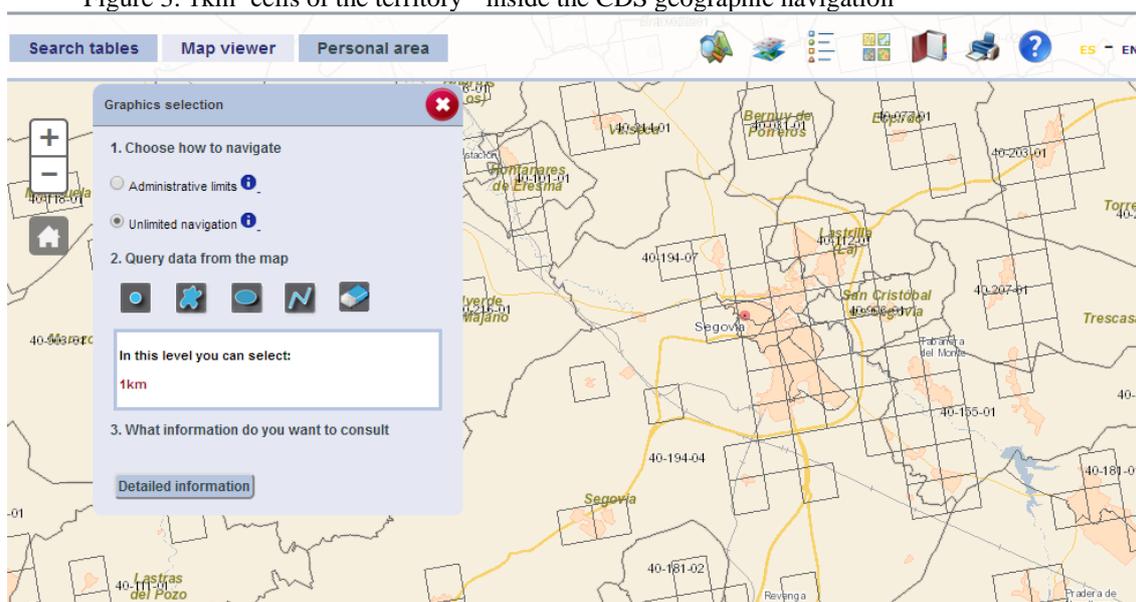
Figure 2: Selection of geographic area directly from the map with the CDS geographic navigation



- (ii) At grid level. The 2011 Census collected geographic coordinates for the first time in history. It was possible to link this information both with people, dwellings, households...and to have all these items georeferenced and assigned to a concrete cell.
 - (b) Four different types of grids: 100km², 10km², 1km² (this layer is the same as the European one) or 50m² were created and users can choose cells, continuous or not, from them and then create a query.
 - (i) 1km² cells were divided into several 50m² cells, according to the population they have. The more population the 1km² cell had, the greater the number of partitions that are made.
 - (c) Predefined thematic maps.
 - (d) It is possible to save the map of interest (using browser cookies) and to use it later.
 - (e) Both topographic and satellite maps are available.

⁹ Some reasons that enlighten importance of visual information:

- Visual content is social-media-ready and social-media-friendly
- Posts with visuals receive 94% more page visits and engagement than those without
- Visuals show products without telling people about them
- Visuals express ideas quickly - in a snapshot.

Figure 3: 1km² cells of the territory¹⁰ inside the CDS geographic navigation

- (f) Once the geographic area has been selected, 7 indicators (amount of people, average age, percentage of foreign people, number of dwellings, percentage of seasonal/empty dwellings, number of households and average size of the households) at different administrative levels can be visualized.
- (g) Predefined thematic maps.
- (h) It is possible to save the map of interest (using browser cookies) and to use it later.
- (i) Both topographic and satellite maps are available.

E. Predefined 1km² Maps

27. One of the most innovative products of this Census was the collection of 1km² maps that INE-Spain has carried out.

28. 15 maps were elaborated with information about the most interesting Census variables:

(a) People: Total population, percentage of foreigners, average age, average level of studies, average number of children, average time taken to place of work or to place of study.

(i) Households: Average size.

(ii) Dwellings: Percentage of secondary or empty dwellings, percentage of dwellings with Internet connection, average useful floor space, percentage of rented dwellings, percentage of dwellings with a mortgage and average period of construction.

29. Depending on the map, they can provide information of each one of the almost 80,000 cells with at least one dwelling inside.

¹⁰ Only 63,500 of the total amount of 1km² cells have population inside.

30. 1km² maps reflect information from a different and innovative point of view (all the cells have the same size) and independent from the administrative one, that people are used to. In most occasions, and especially in big cities, 1km² maps contain more detailed information and show better the reality.

31. These maps allow users to see detailed data at a glance and are recommended for all types of people, not necessarily familiarized with Census information, that want to see information from a visual perspective and to compare data among different parts of Spain.

F. Other products

F.(i) Specific reports “Cifras INE”

32. In addition to preparing products of a technical nature, INE-Spain has also elaborated some informative brochures with visual and attractive contents for the public. These brochures belong to the collection “Cifras INE” (INE figures) edited by INE.

33. In the context of the 2011 Census, 5 different “Cifras INE” reports were elaborated:

(a) The first one contained information about the Spanish municipalities with a higher amount of workers/students, and with a larger amount of non-residents who spend more than 14 nights in them.

(b) The second one was a publication elaborated for each one of the 4 most populated Spanish municipalities (Madrid, Barcelona, Valencia and Seville), with detailed information about workers and students in those municipalities.

34. Cifras INE are disseminated as a pdf file in the web page¹¹ and are recommended for those users who want a relaxed and informative reading about Census contents.

F.(ii) Tailored queries

35. Although a big effort has been made in elaborating different types of products, sometimes users (most of the times because they demand detailed information at enumeration area or municipality level) may not find the information they need for their purposes.

36. In these occasions, users have the possibility of contacting INE-Spain¹² and explaining the details of the information they want.

III. Conclusions

37. INE-Spain has made a big effort in elaborating different types of products to examine 2011 Census information, in accordance with the diversity and number of users that access to it.

¹¹http://www.ine.es/ss/Satellite?L=es_ES&c=Page&cid=1254735116567&p=1254735116567&pagename=ProductosYServicios%2FPYSLayout

¹² As of 14th July 2014, almost 300 tailored queries have been received by INE-Spain.

38. Modern dissemination products try to emphasize the importance of transmitting information on a visual way, in order that conclusions or display of results can be obtained more quickly.

39. Collection of GPS coordinates from buildings (and indirectly from dwellings, households or people) opens the door to a new set of products, very much appreciated by the users, with lots of possibilities both in the context of analysing and consulting Census information.
