



Economic and Social Council

Distr.: General
1 July 2010

English
Original: French

Economic Commission for Europe

Conference of European Statisticians

Group of Experts on Population and Housing Censuses

Thirteenth Meeting

Geneva, 7-9 July 2010

Item 2 of the provisional agenda

Census communication and dissemination, including the use of Geographic Information Systems (GIS)

The dissemination of French census results since 2009

Note by the French National Institute of Statistics

Summary

The dissemination of census results in France is annual since 2009. This distribution takes place exclusively online, under the shape of varied products adapted to different categories of users. Documentation and specific assistance are intended to make users autonomous in their use of data. Dissemination products have been designed after consultation of users, taking into account methodological characteristics of the census and the opinion of the National Commission on Informatics and Freedoms, committee in charge of monitoring respect for freedoms and privacy. On the second edition in 2010, products have evolved to include early user feedback.

I. Introduction

1. Like any census, the French census allows on one hand to develop "official" population data at all geographical levels, and on the other to build detailed statistics describing the population and society at geographical levels. In the case of France, we must consider two features: the significantly numerous and varied administrative units (36,681 Local Administrative Units 2 (LAU 2) units where half the municipalities have fewer than 420 head, but half of the population lives in the 930 largest municipalities) and the specific census method that allows the update of all data annually but requires, in return, some precautions.

2. This text covers only the dissemination of statistical results: it does not address the official populations, released annually online at the end of each year¹.

3. It will first describe the strategy defined by the French National Institute of Statistics (Institute national de la statistique et des études économiques (INSEE)) for building its data offering, then the nature of dissemination products themselves, ending with some points of assessment and the proposed evolutions.

II. The data dissemination strategy

A. All results updated every year

4. The census method implemented in France since 2004 (see Appendix) allows the development of full census results every year. The complete results use data from the last 5 annual census surveys, all brought to the same base year. So in 2009 we released data from census surveys from 2004 to 2008, with the base year 2006 and 2010 collections of data from 2005 to 2009, corresponding to the reference year 2007, etc. Every year, we therefore produce new figures, but only partially built with new information (only one fifth of the information is really new). Data of successive dates do not constitute actual yearly series. Only results from two cycles of five years disjointed collection can be compared strictly speaking because only they are totally independent.

5. Despite this difficulty, INSEE decided to publish each year a complete set of census results: this answers to a strong demand from users to have available for use the latest information at every moment.

6. This also allows consistency between statistical results and the official population figures. The latter, according to the law governing the census, must be updated annually. So every year in late December, population figures from all geographical entities are released and authenticated. Six months later, in July, all statistics are published in full consistency with the previous ones. This alignment was a strong demand from users, including local public actors.

B. A broadcast exclusively on Internet

7. Until the 1999 census, the detailed results of population censuses were available as paper volumes, the addition of which represented several m³. This distribution was

¹ <http://www.insee.fr/fr/ppp/bases-de-donnees/recensement/populations-legales/>

completed for delivery over the Internet later. With the logical decision to update all the annual results, it was no longer possible to consider maintaining all publications in paper form: in addition to arguments of cost and sustainable development, taking in account printing and distribution delays, they were "outdated" almost as soon as they were published.

8. Since 2009, the detailed results of the census are available only online. This offer is completely free, as any supply of INSEE, that requires no "on demand" treatment. The free access was extended compared to previous censuses, to types of previously paid and / or restricted to certain audiences products, including micro-data files.

9. Note that even the official population figures for local units, which are authenticated by decree, are not printed, as the text states that authenticates that "The population figures [...] of *Communes*, [...] are appointed to the values contained in tables available on the website of the National Institute of Statistics and Economic Studies (www.insee.fr). "

C. User wishes and statistical confidentiality

10. The census users were consulted through the CNIS (National Statistical Information Board), a consultation organisation between users and producers of official statistics, where meet together government officials, social partners and associations. A special working group on the use and dissemination of census data worked in 2003 and 2004, bringing together more than 70 users from very different backgrounds. Its recommendations included the need to reflect the diversity of students (from national to local levels and from elected to the researcher), and the requirement of good information on special precautions because of the new method. INSEE has tried to take into account all of these recommendations during the definition of the products.

11. Obviously, the definition of census dissemination products has been the subject of discussions with the CNIL (National Commission on Informatics and Freedoms). From these discussions emerged from the text of a 'regulation on the dissemination of census " which states the nature of the products and authorized levels of detail (spatial accuracy or classification) for each product. This text also limits the specific variables considered as "sensitive", i.e. the countries of birth, country of previous residence and date of arrival in France. All dissemination products have been designed in accordance with these requirements.

D. Make the user independent

12. Census data are traditionally, among those released by the INS, those that are most used, especially by local actors. Thanks to Internet and because of the offer growth, we could expect an increase in the number of users. In addition, the new census method (survey large municipalities, spread over five years of collection) make some precautions necessary. Finally, as with every census, some developments of questionnaire and concepts make comparisons with previous censuses difficult.

13. But INSEE cannot provide training or even a personalized help to each user. Even if it is organized to respond to questions from users, the risk of an excessive number of questions, in regard to the capacity of personalised response, was strong. So a considerable effort has been provided in the website design to make the user independent as much as possible.

14. The main statistical results are contained in a specific sub-website of the main website [insee.fr](http://www.insee.fr): <http://www.recensement.insee.fr>. This website opened on 1 July 2009, 15

months after the end of the last survey taken into account. It has been designed to facilitate the consultation through an intuitive navigation mode.

15. The site is divided into three sections devoted to statistics (marked in red) and two others to promote the autonomy of the Internet (marked in blue): Documentation section and Help section (for geographical research or use of downloadable databases).

16. The "Documentation" section is particularly well supplied: it includes first a set of recommendations for use of data, covering both the specificities of the method (accuracy, temporal effect) and the use of the data subject by subject (including information on the comparability with the previous census). Also included is a list of variables with their definitions, a presentation of the organization of geographic levels in census data and a presentation of the major types of dissemination products.

17. In addition, on each page containing results (sheet, table, or database) are also included documentation files in .pdf format, easily downloadable, adapted to the theme of the data and nature of the results involved.

18. Even if INSEE is not able to train all users of the census, a supporting presentation was designed in June 2009. This presentation recalls the collection principles, presents the distribution of products and includes key tips for using the results. Accuracy and topics concerned by changes from 1999 are particularly detailed. INSEE regional teams have presented it to 5000 Statisticians from state services, local authorities and major research organizations in the fields of urban planning and local development.

III. Products designed for different audiences

19. Even if all the products on the website are in fact accessible to all Internet users, they have been designed according to "targets" which could be described as follows:

(a) General public: whether general public, policymakers, politicians, eager users, media ... everyone that must have the finished product immediately available;

(b) Professionals: they can build independently information tailored to their own needs or those of their sponsors by exploiting online available databases.

Outside these two categories, a specific, pay offer, also exists.

A. Products for the "public" immediately usable.

20. For these users, ergonomics must be adapted: available for consultation in the screen, easy to print and being able to be directly used and valued, without reprocessing. These data were selected to be used without special precautions, especially in terms of significance. The user can choose among 40 000 areas proposed, the geographic area over which it seeks the data (a specific search engine to help them), then the theme of his research.

21. Three types of products belong to this category: sheets of figures, tables and detailed maps of indicators:

(a) Key figures gather the essential results of population census. They are presented in six thematic sheets, offering tables and simple graphics. Most data are presented in evolution, mainly by comparison with the results of the 1999 census. They are available for each *commune* (LAU 2) regardless of its size, and for habitual territories, administrative or study, consisting of a set of whole *communes*. In 2009, each sheet can be exported in .pdf format;

(b) Detailed tables include a set of more than 80 tables. They cover all the topics of population census and provide more detailed information than the cards 'key figures'. The detailed tables provide the numbers corresponding to the intersection of two, three or four statistical variables that relate only to the most recent census. They do not provide comparative data over time, but they are organized according to the same thematic structure that results from the 1999 census. They are only available for areas of more than 2,000 inhabitants. Depending on the size of the territory concerned (2-10 000, 10 000-50 000, more than 50,000 inhabitants), some tables are more or less detailed. For example, the sector of economic activity will be available in 5, 16 or 36 positions depending on the size of the area. Each table is exportable to .xls or .txt formats;

(c) The maps of indicators: Designed for a wide audience, these thematic maps are published on the site "Local Statistics" from the INSEE website (direct link from the site of the census). They are also easy for professional to use with the mapping tool. Indeed, the maps are dynamic in France by region, department, employment area, township or municipality. The indicators mapped are mostly found in the records key figures. Wherever possible, a history of several years or more periods is proposed for these indicators. All data are exportable.

B. For professionals, downloadable databases

22. In addition to the products immediately exploitable, professionals need data that can be reprocessed. More sensitive to questions of statistical significance, comparability between vintages, from changes in concepts, they should consult the documentation and use their own tools for processing and editing.

23. INSEE provides them with data download of two types: the first type, called "bases" requires only a good knowledge of the functionality of a spreadsheet. They are most often available in two formats, XLS and TXT. The "micro data files" which are the second type, must be read and processed with software processing database or statistical software. Full documentation and advice for using the products are presented as .pdf files associated with different files.

24. The first set of bases includes all the indicators of "key figures" or "Detailed Tables" in large tables (hypercubes) with one line for each *commune* and one column by indicator. These bases allow the easy construction of data on any supra-communal area. They therefore include detailed data for even small communities, in order not to constrain the use by experts.

25. Municipalities with more than 5,000 people are cut into quarters of approximately 2,000 inhabitants: the Iris. This division was made in 1999 to allow the dissemination of sub-municipal public statistics respecting the confidentiality rules. Five basic themes provide essential data at the neighborhood level "Iris", in order to make the clusters of sub-municipal areas consist of a set of Iris. These bases are accompanied by statistical robustness indicators (for each zone) and coefficients of variation (for all terms distributed variables), which allows the expert user to adapt its work to the accuracy of data.

26. Bases on mobility flows provide figures for residential migration, commuting to workplace and commuting place of study, total flows of mobility between municipalities.

27. The bases give users access to all communal aggregations of indicators already calculated. To build their own tables in order to study particular sub-populations, micro data files are offered to experienced professionals (public organizations, local authorities, large companies, consultants and consultancy, research...). These files provide information at individuals or units level and allow for exploratory data analysis, modeling behavior, or just make a tab on an under-specified population. Because of the method, each statistical

unit has a weight, as in the surveys. But the file sizes are not similar. These large files, available in .txt format, require databases managers or statistical packages. They are of three types:

(a) « housing », data located in the LAU 2 level (*commune*), describing the characteristics of dwellings and households that occupy them;

(b) 2 files of persons:

(i) One with data localized to the Nomenclature of Territorial Units for Statistics (NUTS 2) level (region), with the characteristics of each individual identified, those of his household, and those of his principal residence. Variables are available with detailed breakdowns (including nationality, country of birth, age of arrival in France);

(ii) The other with data located at LAU 1 level, with grouped breakdowns for reasons of confidentiality;

(c) 5 files “twice located” which allow people to study, commune to commune, travel between home and workplace, between home and place of study and residential migration. The number of variables descriptive of individuals is limited.

C. Few specific paying services

28. For specific requests that do not find satisfaction by means of previous products, INSEE offers “customized products” that are subject to tariffs. In addition, available only to some organizations with a mission of public service, and with an additional fee, a service for figures on customer-built *infra-communal* areas” will be offered in 2011: municipalities can place orders at INSEE to get key figures on areas of at least 1,000 homes, according to their political needs (noise zones, urban restructuring, impact of equipment...).

IV. Initial assessment, early prospects

A. The cost of the device

29. The establishment of this entire distribution device had a significant cost. It was necessary to define precisely the content of each product (form, table or database), to build variables or underlying modalities, to conceive the organization of the website, develop it, and test everything.

30. The labor force that mobilized statisticians, product designers and computer specialists, between 2007 and 2009 can be estimated as worth 11 years of work of senior executives. The development of the site itself, subcontracted to a specialist company, has cost INSEE 280,000 euros. In total, the implementation of the dissemination of statistical results of the French census has cost 1.4 million euros.

31. Its annual update, which includes the complete renewal of the data, error correction and some developments, involves approximately 4 years / people and outsourcing resources, costs about 400,000 euros.

B. Successes and new applications

32. Since its opening, the website devoted to census has been visited by nearly 60,000 visitors a month. A survey, conducted in summer 2009, and the feedback received since

then demonstrates that the site is very appreciated. Highly anticipated in terms of content, it is affordable, easy to use thanks to a straightforward presentation, with introductory texts considered explicit enough and has an easy feel access to key data. The new interface is classic, familiar and appropriate, its appropriation immediate. The site globally seduced all public, including that of the professionals who appreciate its statistical information wealth.

33. But, as always, this dissemination has generated additional requests, which are expressed mainly through several channels:

(a) Internet users directly questioned INSEE, by mail or telephone, and expressed their wishes or suggestions;

(b) A group of National Council of Statistic Information (CNIS), bringing together diverse audiences (local authorities, enterprises, researchers, state services) has supported proposals to improve product distribution of the census and its accompaniment;

(c) When studies or expertise works are conducted by the regional directorates of INSEE with public local authorities, they share their comments and requests;

(d) When the site opened in July of 2010, an online survey was organized to complement the information.

34. User requirements are sometimes on the form of dissemination products, sometimes on the data themselves, sometimes on the site documentation. INSEE analyses the most recurrent requests and tries, whenever and wherever possible, to answer them in subsequent versions of the site.

C. Developments in 2010

35. With the second edition of the site, several improvements have been made. Here are some examples:

(a) It is now possible to edit and export in a single .pdf file for all records "key figures" of a territory, which were divided into six thematic sheets in 1999. The user can thus obtain a single-click the portrait of its territory in the census;

(b) Data sheets "Key figures" of a territory can be exported in .xls, instead of only .pdf format;

(c) Geographic levels are added in the micro data files. In the file housing, there are particular code district IRIS. In this area, INSEE went to the limits of what has been authorized by the CNIL;

(d) New downloadable databases have been built to allow direct comparisons between the 26 regions (NUTS 2) and 100 counties (NUTS 3) on all indicators. In the 2009 version, the user himself had to rebuild these areas from communal basis.

V. Conclusion

36. Dissemination of detailed census statistics can be regarded as a success: a lot of data, reflecting the expectations of various users, have been made available on schedule. They were well received and are already widely used.

37. For the future, INSEE must consider the manner of providing such information less costly in human costs, in order to both meet the need for annual updates and respond to new demands. These new process will also help integrate the census results in the overall

organization of the INSEE website: it must shift from a source organization to a topic organization. The timing of this consolidation is not defined.

Appendix : The new census method

1. The new census method replaces the traditional organized counting every eight or nine years with an annual survey technique, that allows municipalities to have fresh data, updated annually. The method of survey varies according to a population threshold fixed by law to 10,000 inhabitants.
 2. **The municipalities of less than 10,000 inhabitants are recorded once every five years by rotation.** They were divided into five groups according to specific rules that provide an equivalent number of inhabitants in each group. Each year, the census survey covers the entire population and housing of Commons of the group. After five years, the entire population of municipalities with fewer than 10,000 inhabitants will have been surveyed.
 3. **In municipalities with 10,000 inhabitants or more, a sample of the population is surveyed each year.** The annual collection focuses on a sample of addresses representing about 8% of the population. After 5 years, the entire territory of each municipality has been taken into account and 40% of the inhabitants of these municipalities have been surveyed.
 4. The collection frequency is thus quinquennial for municipalities of less than 10,000 inhabitants and annual for those of 10,000 inhabitants or more. The census survey is exhaustive in the first case, a sample survey in the second case. Specifically, about 9 million people are reported each year, i.e. 14% of the population living in France.
 5. **The data released are from the last 5 surveys:** these apply to 45 million people or 70% of the population. The data, collected during five years, are returned to the same relevant year (the median year) by methods of interpolation, extrapolation and moving averages.
-