**STATISTICAL COMMISSION and**
**ECONOMIC COMMISSION FOR EUROPE**

Working Paper No. 19
English only

**CONFERENCE OF EUROPEAN STATISTICIANS**

**Joint UNECE/EUROSTAT Work Session on Methodological Issues Involving the Integration of Statistics and Geography**
(Tallinn, Estonia, 25-28 September 2001)

Topic (ii): New technological solutions, including those based on online data access

# GEOGRAPHY AND NEW CENSUS IN FRANCE

Submitted by INSEE, France[1]

**Contributed paper**

## I. INTRODUCTION

1. French National Institute of Statistics (INSEE) plans to change the way the census of the population is carried out. Geography will play an important role in the new method, not only in data dissemination, but also in data collection.

2. In the first section, this paper will sum up the new method. The second section will concern data collection, and the third section, data dissemination.

3. A law, voted by Parliament, will only allow the implementation of the new census in France. At the time of writing (April 2001), the bill has not yet been presented by the Government.

## II. NEW CENSUS IN FRANCE[2]

4. Over the last decade, reductions in public spending have been observed in many Western countries. In a number of countries, it has thus become increasingly difficult for national statistical offices to justify and obtain sufficient funding for a complete enumeration of the population.[3]

5. INSEE has embarked on a complete overhaul of its general census. Aiming at the same objectives (legal population, statistical description of the population and to sample population for statistical surveys) while lowering the response burden, the new census (named RRP, from the French words: New Census of the Population) could be characterized by the data collection within a sample of the territory and the use of administrative data.

---

[1] Prepared by Michel Isnard.

[2] This paragraph is a summary of a paper presented by Jean Dumais and ali at the Population Estimates Methods Conference, Washington, D.C., 8 June 1999; available from the author.

[3] In France, the cost of '99 Census is about 190 million Euros or 170 million dollars.

### II.1 Sampling strategy

6. The pivotal unit is the commune[4] or municipality: smaller communes (less than 10,000 inhabitants) would be sampled at the average rate of 1/5 and then completely enumerated; all the larger communes would be covered every year and only a fraction (1 over 12) of the dwellings would be contacted. As the smaller communes comprise about one half of the population, the overall annual sampling fraction would be about 1/8.

7. For the domain of smaller communes, we would create 5 rotation groups of communes to be surveyed in turn. Those annual rotation groups would be created within each of 22 French regions (Nuts II French levels). Using 1999 census data, each group would be created so as to best represent each region's socio-demographic make-up, thus minimizing the year-to-year variation. Within each region and each rotation group, a complete enumeration would take place as is done in a traditional census operation. For the domain of the larger communes, we plan to use a "buildings register". The buildings register is described in Nathalie Eltchaninoff's contributed paper. The sampling plan considered for each larger commune is a stratified 2-stage sample of dwellings. Firstly, the buildings in each commune are stratified in IRIS2000[5]; in each IRIS200, five rotation groups are created as seen above for the annual survey. Secondly, in the annual rotation group, the listed buildings will be visited and a complete list of dwellings will be drafted; from this list, a random sample of about 40% of the dwellings will be drawn.

### II.2 Estimation and dissemination

8. Annual estimates for the larger communes should be straightforward expansion-type estimates. Annual estimates for the smaller communes will be possible for those subjected to the annual census.

9. The current dissemination plans are to publish by December 31, Y the national results of the census in year Y and the results for some large segments of the territory (e.g. metropolitan areas of more than 1 million inhabitants). Moreover, the detailed results of year Y-2[6] would come from a census or sample survey (smaller or larger commune respectively) or from synthetic estimation. A synthetic estimate would make use of relationships between the observed data (census or surveys) and administrative data for a given time point at a very high level of geographical detail (buildings, city blocks or communes).

## III. NEW CENSUS AND GIS

10. This paragraph will try to show the use of INSEE's Geographic Information System in the new Census. Two phases will be described: during data collection and during data dissemination.

### III.1 During data collection

11. Data collection depends on the size of the commune: for smaller communes, data collection is very similar to that of census and will not be commented upon here. In larger communes, data collection will resemble that undertaken for a statistical survey. Everything mentioned here can be applied for data collection of a survey.

---

[4] Remembering that Paris comprises 20 communes, Marseilles 16 and Lyon 9. There are 36,685 communes in France, half of them with fewer than 400 inhabitants. Communes are the French NUTS V level.

[5] IRIS = "îlots regroupés selon des indicateurs sociodémographiques" or city blocks regrouped along statistical indicators; each grouping is connected, rather homogenous and comprises about 2,000 people.

[6] The 2-year lag is what is currently expected to be the time needed to receive and process the administrative data.

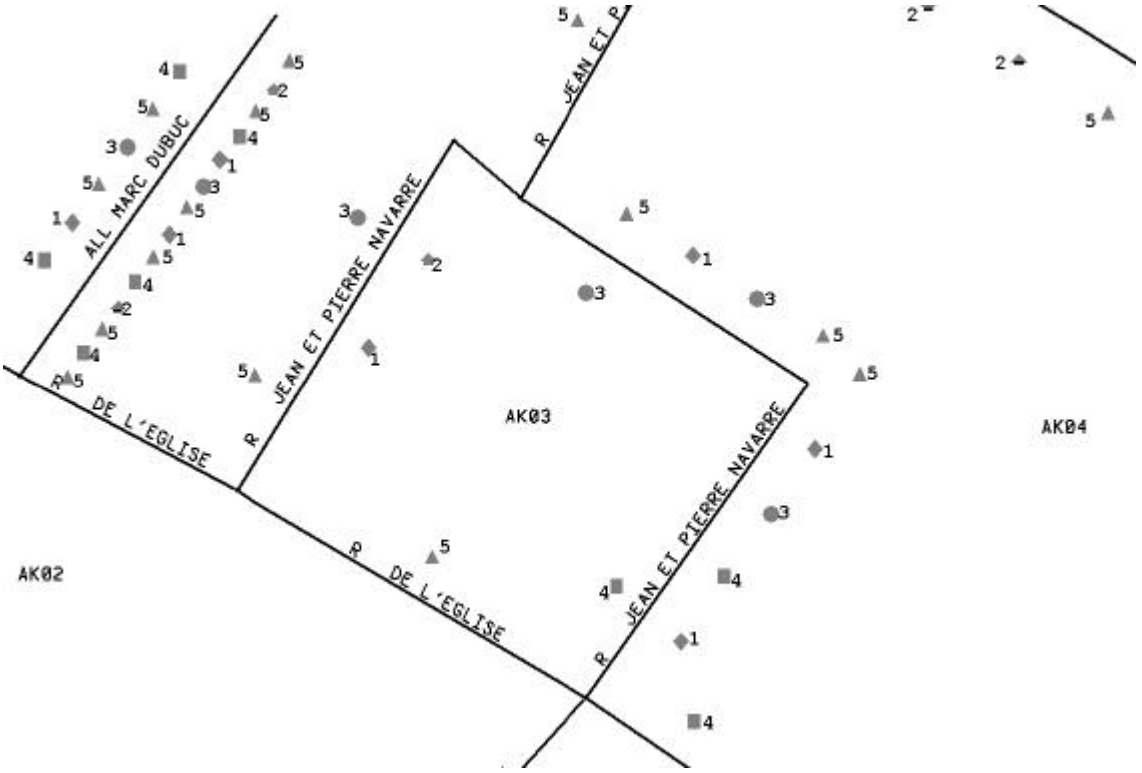


**Figure 1: Buildings Register with rotation groups**

12. Figure 1 is a description of INSEE's buildings register, each building marked with its rotation group. Buildings marked 1, for example, will be visited during years 1, 6, 11 and so on. The work of census agents will be comprised of:

- visiting every building of the present rotation group and make a list of dwellings inside it
- procede to the statistical survey with the sample of dwellings drawn from the list described above (sample rate about 40%).

13. Figure 2 is an example of a map that could be given to the enumerator. The map will, of course, be in colors and will give a more accurate description of the buildings (address for example). But the buildings of other rotation groups will probably be present.

14. Both statistical and geographic data of the GIS will help INSEE to collect data and control data collection. Geographic data from the GIS will help census enumerators to locate buildings. Localization will be based on geographic coordinates, addresses and other significant data. Statistical data from the GIS (especially number of dwellings at last census) will help INSEE to control the results of data collection: to verify that the right building has been visited and that the number of dwellings is roughly accurate. Administrative data from dwelling tax rolls, which will be localized with the help of GIS information, will also be used to control data collection.

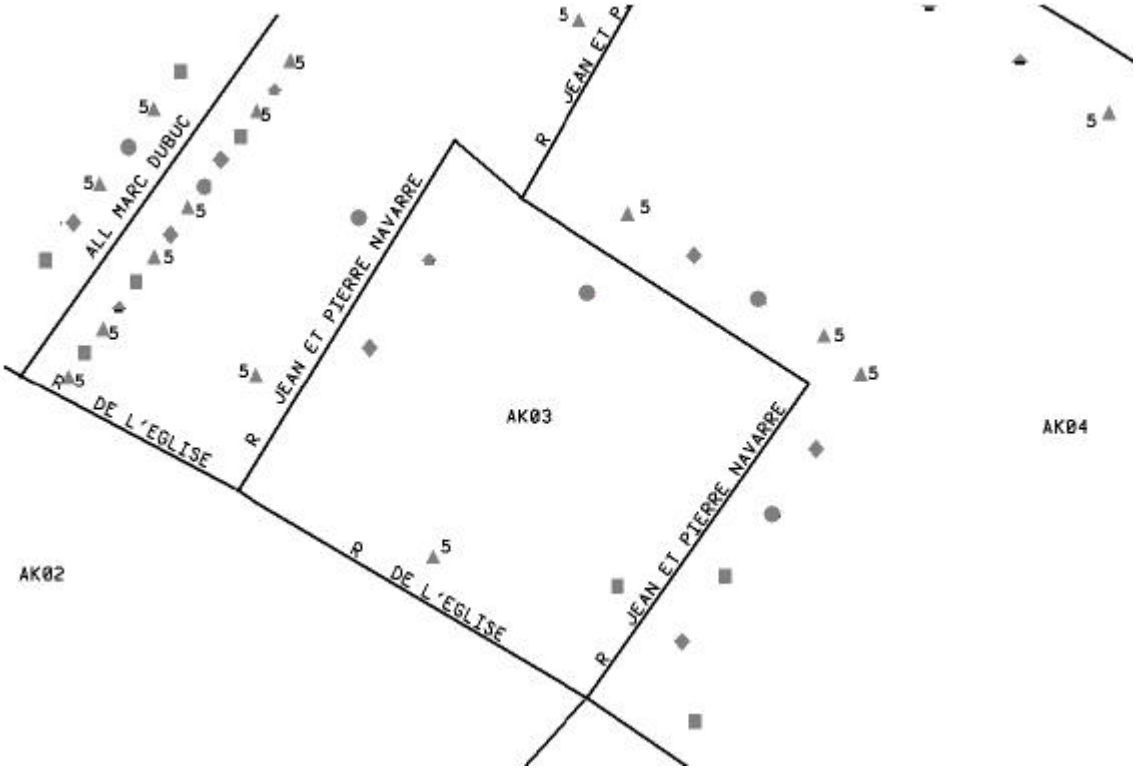


**Figure 2: Example of map given to the dwellings enumerator**

15. As described earlier, the results of that visit are a very accurate description of every dwelling inside the buildings of a given rotation group.

| Address | Dwelling # | Description | To be censused |
|---|---|---|---|
| 1 allée Marc DUBUC | 1 | Ground Floor | NO |
| | 2 | First Floor | YES |
| 5 allée Marc DUBUC | 1 | | YES |
| 11 allée M. DUBUC | 1 | First Floor, Left door | YES |
| | 2 | First Floor, Right Door | NO |
| | 3 | Second Floor | NO |
| | 4 | Third Floor, Left door | YES |
| | 5 | Third Floor, Right Door | NO |

**Table 1: Results of the enumeration of the buildings**

16. After the first visit, a sample of dwellings will be drawn up, as shown in Table 1 above, and data will be collected inside the dwellings on that list. Data collection control will also use data from the geographic information system.

17. The quality control will also use the expected number of buildings and dwellings, drawn from GIS and previous collections. This data will allow INSEE to follow the response rate of data collection and therefore take measures to improve it.

### III.2 Data dissemination

18. In France, Census data dissemination falls under a law named "Informatique et Libertés", which deals with data treatment and dissemination of data. An independent body (named "Commission Nationale pour l'Informatique et les Libertés", CNIL) authorizes every action undertaken by INSEE. For the '99 census data dissemination, the rules are the following:

- the end-user could only obtain tables at the IRIS level (see above). Individual files could be at a geographic level containing at least 50 000 persons. Specific variables (such as nationalities or birth countries) could only be obtained at a 5 000 inhabitants level.
- very specific users (mainly the communes) can obtain, for specific purposes, tables at an enumeration district level (roughly 600 inhabitants).

19. These specific rules were built in order to:

- avoid individual identification by way of public census tables (is the only foreign doctor in an IRIS married?)
- avoid social characterization for small zones, especially for trade companies or banks. Municipalities with specific purposes could obtain data at enumeration district level.

20. The problem encountered by municipalities is that enumeration districts are fixed once permanently and can not be used for all purposes. On-purpose zones could lead to identification, because census is (should be) exhaustive.

21. New census could change dissemination rules with the help of GIS. First, in larger communes, new census looks very much like a survey and therefore is not exhaustive. The treatment of the collected data (building of synthetic results) combines data from the results of 5-year data collection and administrative data. Therefore, individual identification is quite impossible, because it is not even known if an individual was included in the census.

22. The same features could open up a new kind of data dissemination for on purpose zones.[7]

[7] CNIL has not been consulted on this topic and its decision will be necessary to put into effect that kind of data dissemination.

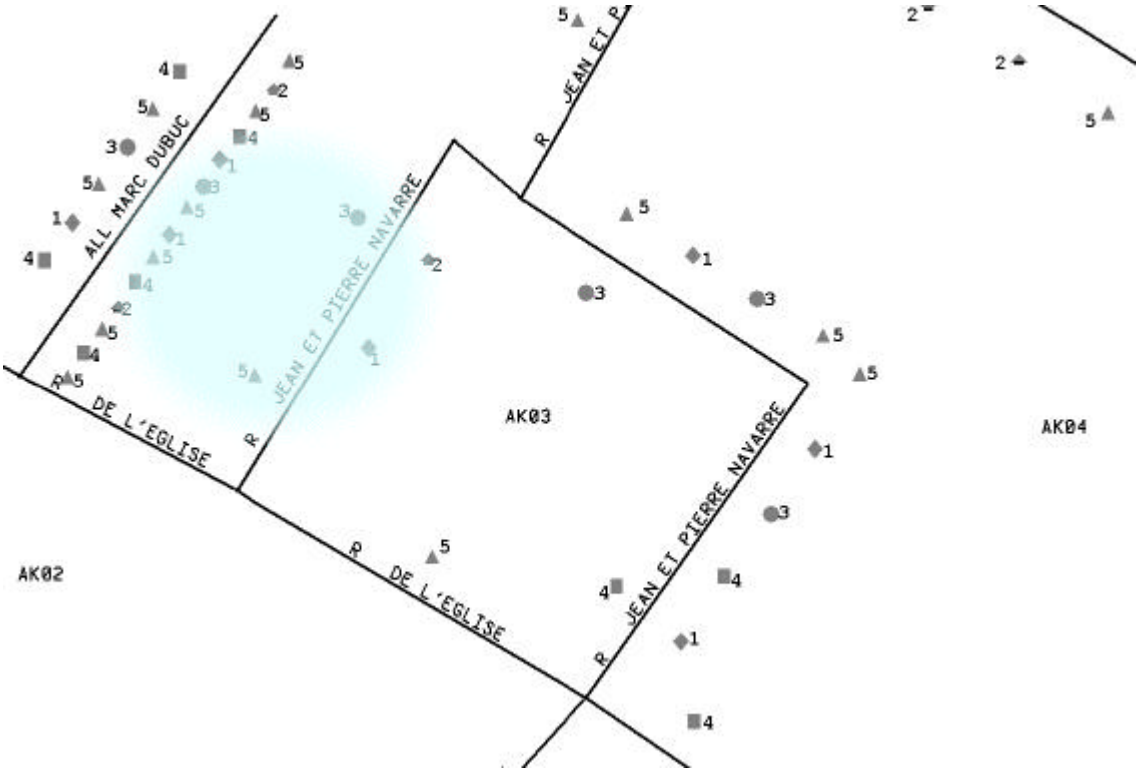


**Figure 3: Data dissemination on specific zones**

23. The individual file used for census dissemination (even if it won't be disseminated in that form) will be closely linked to the GIS. It will therefore be quite easy to compute tables for that specific zone and therefore disseminate data for it.

24. Individual identification could not occur, because of the sampling that occurred during data collection. Of course, sampling errors will be very important for small zones.