Distr.
GENERAL

CES/SEM.40/4
8 September 1998

Original: ENGLISH

STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE

STATISTICAL OFFICE OF THE
EUROPEAN COMMUNITIES (EUROSTAT)

CONFERENCE OF EUROPEAN STATISTICIANS

Joint ECE-EUROSTAT work session
on Population and Housing Censuses [1]
(Dublin, Ireland, 9-11 November 1998)

Study topic 1

APPLICATIONS OF GEOGRAPHY FOR THE CENSUS: THE FRENCH EXPERIENCE

Invited paper submitted by INSEE, France [2]

Geography has always been playing a key role in Census taking and other data collection processes. This role is changing nowadays because of the availability of GIS.

This paper presents various uses of geography depending on Census collection methodologies and describes in particular the French experience; it also shows new perspectives opened by new geographic techniques as they have been adopted in France.

The paper will be the basis for some recommendations to discuss at the Dublin Conference organised by ECE and Eurostat.

I. Census collection methodology and geography

1 Collection methodologies currently in use in different countries are very diverse; the same is true for existing geographic databases. As a consequence, different geographic tools are needed. This paper will consider geography for data collection purposes with a wide scope, including postal address management where this is required by the collection methodology. A few countries have been selected for example, but many others could have been considered.

1 The papers which are prepared for this work session will be treated in the same manner as papers that are prepared for seminars.

2 Prepared by Michel Jacod.

2 This paper does not intend to focus on geographic issues for data dissemination. Dissemination methodologies are, in many ways, independent of collection methodologies differ a lot from one country to another. Data dissemination deals with: geo-referencing statistical results; producing thematic maps; mapping of links between territorial units; disseminating through internet; electronic supports or paper; graphic semiology.

Nevertheless, dissemination issues regarding geography should be linked to collection issues for cost and quality reasons.

3 Few examples are enough to show that diversity in Census collection methodology, as far as it concerns geography, cannot be properly approached by distinghishing the two usual categories:

- register-based Census;
- traditional Census.

4 The Danish or Fin nish experience of getting Census-like information from many connected administrative sources requires indeed a heavy use of geography by maintaining geo-referenced files of housing units. Data on place of work or place of study are also extracted from geo-referenced data files.

5 The Belgian methodology is an example of a mixed strategy: use of register and survey to check and to supplement the existing information. It makes also use of postal addresses for mailing the forms.

6 On the traditional Census si de, the «mail in» methodology, adopted for instance by the US Bureau of the Census, requires a lot of work before Census day in order to build the address list of the housing units and to geo-reference these addresses. Conceptually, that is not very different from the Belgian example, except that it is a one shot work. As soon as the US Bureau of the Census will be allowed and will commit itself to maintain this «master address file» through time, the conceptual difference on the geographic ground will almost disappear, even if the procedure will still be different.

7 The «mail out» methodology (Canada) looks substantially different. An address list should also be set up, for instance to manage response rates, but that can be done after the return of the Census forms. And there is no purpose at re-using this list next time. Though, Census boundaries should be designed and local maps produced and provided to enumerators throughout the country before Census day.

II. Several aspects of Geography in the '99French Census experience

8 After this rapid overview of different geographic issues depending on various collection methodologies, this chapter will go more deeply into detailed geographic work with the example of the current preparation of the

'99 French Census. Several levels are to be investigated, from the «communal» level to the most detailed one, which is the building location. Precise purposes differ depending on the geographic level even if the general objectives are always to guarantee the collection exhaustivity and to allow the output of as many results as possible for every geographic level.

II.1 Communal Geography : managing the few changes

9 French statisticians take advantage of a very stable and very detailed administrative zoning with its 100 « *départements*» and especially its 36,000 «*communes*» (NUTS 5 in the European terminology), many of them being very small (median is about 300 inhabitants and 50% of the population live in *communes* under 10,000).

10 At these two levels (*communes* and *départements)* the maps for dissemination purposes have been based on *IGN*[3] maps and SAS software has been used since the beginning of the 80's. Production of atlases was partly contracted out with specialised firms on several occasions. More recently, electronic dissemination of data associated with communal maps came in a significant demand and *INSEE* tried to enter in this market with a French distributor of MapInfo.

11 There are very few boundary changes between *communes*, and many of them concern only exchange of agricultural lands without houses. Some mergings or partitions occur between two Censuses, for instance 16 mergings and 25 partitions occurred from 1990 to 1998. They are not numerous, and moreover, most of them are localised in few regions of France. *INSEE* keeps records of these changes inasmuch they interfere with population and the historical reference is influenced. It publishes regularly an official geographic codebook and makes it available each year on electronic media. For Census purposes, the parts of a *commune* which were merged are censused separately as a precaution: in fact, they could be splitted again some years later because of modification of local consensus.

12 For 28,300 *communes* under 1,000 inhabitants [4] (total population: 9,3 M in 1990), there is only one enumerator per commune and the territory is not divided into several enumeration areas. Infra-communal is not very relevant in those communes for Census taking as well as for results dissemination; people knows each other better than any Census could ever do.

13 For 7,400 *communes* between 1,000 and 10,000 inhabitants (total population: 19.5 M), several enumerators should be hired and enumeration areas be designed. In order to limit the workload and to concentrate the resources on important towns rather than on small ones and villages where Census taking is objectively easier, *INSEE* let municipalities produce their infra-communal geography without a strong a priori control.

---

3 *IGN* : *Institut Géographique National*, the French mapping agency

4 This threshold and the next ones are not strictly applied.

14 *INSEE* is mainly involved in geography for the 800 *communes* over 10,000 inhabitants (total population: 27.8 M). In this case, a lower level called «*îlot*» and not the communal level is considered.

II.2 Infra-communal Geography : a critical Challenge

15 The French collection methodology, as in most countries where traditional Census is taken, raises two groups of issues:

- dividing the territory in practicable units;
- insuring that each enumerator is correctly located and each portion of the territory is actually censused once and only once.

II.2.1 Dividing the territory

16 The objective here is to design small parcels of the territory adequate for dissemination and collection purposes [5]. The French experience showed us that it is not easy to deal, in some case, with these two kinds of constraints and led us to define two geographic concepts, which happen very frequently to be identical on the field:

- «*îlots*»,
- «*districts*».

For the '90 Census, 220,000 *îlots* and 230,000 *districts* were designed in *communes* over 10,000.

17 No information will never be made available at a thinner level than the *îlots* and that is precluded by limiting geographic identification in dissemination files at that level. The question is to define useful geographic units, that means that they should be small [6] and that their limits should follow any existing administrative boundary and some visible limits like streets, rivers, railroads, ... In fact, *îlots* are generally the smallest parts of the communal territory limited by streets. They should also vary as little as possible through time from one Census to the next one. Finally, they should permit the implementation of confidentiality and quality rules. As for confidentiality, in accordance with CNIL (the French privacy authority), information will be released to public authorities (including local ones) after their formal commitment not to further disseminate.

18 The *district* is the geographic unit used for collection. Its characteristics are the size limit (under 800 inhabitants [7]) the visibility of its limits for a little skilled enumerator. Very generally, both units, d*istrict* and *îlot,* are identical but somewhere one *îlot* is split into several *districts* because of its size; that is the case when very large

5 These two purposes are clearly separated in some countries.

6 and sometimes they tend to be too small

7 Again, that is not a strict threshold.

housing buildings are present on an area with no streets in between. On the field, in general there are no difficulties in distinguishing the different *districts* but, apart from exceptions, the present *INSEE* GIS does not describe buildings but only streets.

19 *INSEE* cooperates with local municipalities in order to design *îlots* and *districts* before Census. Their interest comes from the future output they are waiting for. This process lasts a little less than a year: it began in January 1998 and should be finished in November 1998, for a Census day on March 8th, 1999.

20 Availability of a GIS is not stri ctly necessary, one could design *îlots* on local maps provided for instance by the municipality; however, a GIS helps a lot by giving the means of recording uniformously these limits as soon as they are defined for future collection or dissemination purposes.

II.2.2 One district, one enumerator

21 Each *district* is assigned to an enumerator but an enumerator can be assinged to census several *districts,* depending on their. The normal workload of an enumerator is between 500 and 600 inhabitants. It could be a little bit over 600 when it is too difficult to divide a *district* into two units because of lack of visible internal limits. In a town, it is never fixed significantly under 500 because the enumerator wage would not be motivating enough [8].

22 Enumerators are provided with a localization map for each *district* and a list of the street segments they have to census [9]. These two pieces of information do not come from the same tool because of the current *INSEE* organisation. We maintain concurrently two geographic databases:

- «*REPLIC*[10]» is a list of the streets segments related to each *îlot* with postal addresses at each end of the segments . This alphanumeric database was created after the 1982 Census for data dissemination purposes; a second version was created before 1990 for collection purposes and the third one is scheduled for November 1998.
- «*Base-Îlot*» is a GIS built in partnership with IGN since 1993. The software is ArcInfo. It includes the description of the street axis and the postal addresses at the ends of the segments limiting the *îlot.* Limits other than streets are also described. In few

8 Enumerators are directly renumerated according to the number of completed forms that they return to *INSEE*.

9 See an example of this material in annex.

10 For « *REPertoire de Localisation Infra-Communale» (infra-communal localization register).*

words, it could be said that *IGN* provides *INSEE* with the street geometry, *INSEE* adds the limits of îlots on the GIS, and both bring information about postal addresses.

23. For the 99'Census, these two tools have been extended with the description of the *districts* when they are different from the *îlots*.

24. The data contents of the two bases are identical about the *îlot* or *district* identification numbers and their composition on the field; there are a few differences in street names, for instance a name has been updated in a base and not yet in the other one. Postal addresses are here the main concern. That is because there is no physical link between the two bases; neither *IGN* nor *INSEE* have yet succeeded in solving this technical difficulty. But basic information is gathered only once from local authorities and other sources and this information is normally used for maintaining the two bases but with different processes.

25. Another relevant issue is from the conceptual difficulty in defining postal addresses in some cases: for instance where there is no building on the extreme lot of a street segment or there is a store and no housing units, the number of the last building unit to be enumerated in a Census as well as the number of the last lot can be different depending on scopes of the source and not according to a statistical rule.

26. There are therefore some inconsistencies in the material provided to enumerators. Tests showed that they will succeed in managing those differences when canvassing the area.

27. New fonctionnalities will soon be added to the *Base-Îlot* software; in addition, in 2000 *REPLIC* will be derived from *Base-Îlot* after that postal addresses in the latter one have been improved by an automated comparison with *REPLIC*.

28 Another issue was raised in *INSEE* when the decision of replacing the usual cadastral maps with digitized maps was made in 1993. The existing cadastral maps showed a lot of details about lots and buildings but were not frequently updated. Despite the frequent delay for updating, people in *INSEE* were used to employ those maps and most of them regarded such a detailed information about lots and buildings as necessary for enumerators. Without this information, an enumerator could be lost and canvass a neighbour *district* instead of his assignment or he could omit secondary buildings which are not located by the street but, for instance, in a second row.

29 A comprehensive test was conducted at the end of 1993 on 108,000 housing units in 1,075 *districts.* Each enumerator canvassed two groups of *districts*, one with cadastral maps and the other with numerised maps; and a especially trained supervisor went as well on the field with all the existing material to constitute a reference. To minimize the enumerator effect, a Wilcoxon test was made on the comparison of the units numbers by *district.* It showed no significant differences between the two methods except where the building

organisation is particularly complex; in this case, it was decided to digitize some supplementary objects in order to help the enumerator for a good localisation. [FANOUILLET]

II.2.3 Enumerators hiring and assignment

30 Formally in France, enumerators are hired by municipalities which get the necessary money from INSEE. INSEE only has to check the ability of candidates who are presented by the local authorities train them and assign them a reasonable workload. *INSEE* is in charge of all the paper work about supervisors, and of course their training and assignment.

31 For the '99 Census, this task began in June 1998 and we did not use any GIS facilities except in some local offices where supervisor assignments were designed with our GIS. In that field there is clearly place for future improvement.

II.2.4 Collection management

32 Collection period lasts four weeks in France. It is followed by about six months during which local counts are checked concurrently by local authorities and *INSEE* before the output of provisional figures (July 1999) and definite counts (end of 1999). All this process is decentralized: a regional and local hierarchy is constituted for the collection period and the count processing period is undertaken through the 22 permanent *INSEE* regional offices.

33 Collection geography, from the thinnest level to the whole country, has been defined before Census day except some local updates at the beginning of the fieldwork that cannot be avoided. An alphanumeric file has been derived from the GIS used to produce enumerator maps *Base-Îlots*; it comprises the comprehensive list of all geographic units. This file constitutes the backbone of the management software to be used during the processing period.

34 We choose not to use GIS facilities to assist this management. GIS advantages did not appear obviously at that time because all units must be processed in the same way and there is no use of topographic links between these units making of the GIS a unique tool.

II.2.5 Supplementary surveys

35 In France, every person is due to answer the same form , there are no short and long forms (as in the US, for instance). Therefore, there is no need to make a distinction between households. But, there is traditionally a supplementary survey about family organisation, fertility, female work versus child breeding. The sample for this survey is drawn from the list of the *districts;* more precisely it is a sample of enumerator areas. Apart from that, the survey is taken with the same process and in the same time as the normal census. Because of the collection process, the forms are voluntarily short (four pages) and should be self explanatory; however, the sample size

can be very large for a very low cost (340,000 respondents in 1990).

36 In 1999, there will be three surveys of this kind:

- «study» on family history (to be filled in by 250,000 women)
- «study» on family history (to be filled in by 150,000 men, 1999 innovation)
- survey «daily life and health» (a screening survey on 200,000 households which will permit to take a comprehensive survey about handicap at the end of 1999 on 20 000 persons, a second innovation for 1999).

37 The design of these samples has two objectives : regional representativity and minimization of interferences with the work of the enumerators and of their supervisors. These constraints led us to a sort of cluster technique at both levels: *districts* of course and also groups of enumerators supervised by the same person. The limits of this sample design are tempered by the opportunity of raking on many socio-demographic variables offered by the simultaneity of the Census.

38 For 1999, no GIS software has been used for sampling though it appears clearly to be a geographic work. Here again, we could have improved our productivity and maybe the reliability of the results by this mean.

II.3 Control of buildings existence and localization

39 The experience of previous Censuses shows that about half of under-enumeration derives from housing unit's omissions. For the first time in 1990, external files were used to check housing units' exhaustivity. Occupancy tax files are produced by the tax administration with some delay but even provisional data are of interest for this objective. In 1990, these files were used in towns over 20,000 after having been sorted by street names. This sort was not satisfying because streets are frequently very long and concern several *districts*; the effect on collection exhaustivity was therefore impressive: the post enumeration survey showed that the omission rate falls in the towns where the process was successfully implemented: from 0.6 % to 0.2 %. [COEFFIC]

40 For 1999, the decision was made to go further with an extension to all towns over 10,000 and with a preparation of these files to make them easily usable by enumerators and supervisors: housing units [11] and buildings will be geo-referenced in January 1999 at the *district* level using a facility developed by *INSEE* under the name of « *îlotage*». We use *REPLIC* (see before) as reference file for localization. The automatic coding is very efficient with coding rate over 95% in almost all towns. These rates should not surprise

11 These files describe housing units but also other units (garages, independent cellars, apartments transformed into professional units , ...). They should be cleaned as possible before use.

because there are close links between local governments and tax offices for maintaining these files [12].

41 Instead of the alphanumeric file *REPLIC*, we could have used our *GIS Base-Îlot* with GIS software facilities for geocoding. Thus, tests are under way using MapInfo software but their objectives go further than the present one for the '99 Census (see below). As the objective is only to provide enumerators and supervisors with local lists, GIS facilities are in fact not strictly necessary.

II.4 Geography and data processing

II.4.1 An industrial process

42 Census data processing comprises data capture, coding, imputation, sampling if the whole information is not processed in the same detail or at the same pace, and in some countries estimation using external data at a regional or provincial level or even at a local level. The last step of this list is not currently in use in France, it will not be considered here but it will constitute one of the challenges we shall have to face in the next future with a continuous Census (see below).

43 Data processing at the scale of a Census is an industrial issue and it should be organised as such. One key issue here is the fractionating of this bulk of data in subsets. As our processing is divided in three important parts, the criterion used for the constitution of subsets will not be the same during the whole process.

44 For the '99 Census, the three important parts are the following:

- Legal counts: the work is organised by *commune*, roughly independently of its size; it is due in October 1999.

- First statistical processing: all forms are scanned and the images are processed (OCR and data capture from images for open questions) except not for questions about work (producing occupation and industry codes); the whole processing is organised by group of *communes* as soon as the forms are available for scanning; a group comprises fulfilled forms of 250,000 persons. This process should end and results made available one *département* after the other during the second and third terms of 2000.

- Second statistical processing: a sample of about 1/4 of the forms is processed for questions about work. Again processing is organised by group of *communes* but not the same groups as previously because their size will be more important for two reasons : less work for each form, quality control samples designed on larger units for cost limitation. Output is expected during the third and fourth terms of 2000.

---

12 Occupancy tax is an important part of municipality income.

45 This choice could be discussed, even in the French context. For instance, we might have spared time by a better co-ordination between the two statistical processing described before. Now as the whole process was reengineered with the recent decision of using optical reading techniques [13], this organisation could appear a bit conservative. In any case, geography would widely interfere with data processing.

II.4.2 Sample for processing

46 The sample (1/4) mentioned for this second stage of processing is a sample of households drawn one *district* after the other. The technique might be improved for 1999 using data captured during the first statistical processing like household size and head of household socio-economic status indicators. Geography interference is very smooth; its only reason is to keep the local representativity of results at the highest possible level.

47 Early during this process, another sample is drawn to test in almost real size the whole process and eventually to produce some first demographic and socio-economic reports at the level of regions and country. The sample rate is one in twenty.

48 Traditional techniques, before optical reading, let us with no alternative for drawing this sample: for cost reason, it had to be a sample of *districts* where one household in four had to be coded to be used in the ¼ sample processing afterwards. For the '90 Census, methodological research was made and a balanced sample of *districts* was drawn to optimize its quality [DEVILLE]. For 1999, the scanning of all the forms will be completed before December 1999; that gives the possibility of drawing directly a sample of households with no significant costs.

49 Thus geography is no longer involved in that work, and statistical quality improved by the way.

II.4.3 Imputation and geography

50 Geography is on the contrary in the core of imputation techniques employed in France with the hotdeck technique which is widely used as soon as non-responses or inconsistent responses cannot be replaced by deterministic imputation from the rest of the questionnaire. Hotdeck implies proximity on socio-demographic characteristics but as well geography.

51 The geography effect has not been studied in depth, but locally one should consider that it probably exists.

---

13 The decision was made in June 1998, but the preparation of this alternative to traditional data capture began two years before that date.

III. Perspectives in French Census collection and Geography

52 In September 1994, the next Census initially prepared for 1997 was postponed to 1999 for budgetary reasons. A wide range of reasons pushed INSEE to think of new methodology about Census [JC DEVILLE, M JACOD]. The main one is related to the French intercensal period (from 7 to 9 years during the last decades) which the users regard as too long and uncertain (just consider the postponement to 1999 of the Census); another one is the potential breech of confidentiality implied by co-operation with local governments; a third one is about the budgetary issue with internal and external aspects.

53 The decision has not yet been made but we have already get over several obstacles: internal organisation to cope with such a project in short term (the starting date for collection could be the end of 2001), discussion with experts of the government about the implications on the regulations they are in charge of, check with the « *Conseil d'Etat* [14]» of the constitutionality of such a kind of Census.

54 Next steps are:

- to open a large consultation with users in April 1999, after the '99 Census collection period;
- to prepare a parliamentary procedure which could take place at the end of 1999;
- to build several scenarios depending on what could be the results of these two procedures in order to start the project in time to reach the 2001 deadline.

III.1 Why to imply Geography

55 The intention in this paper is only to give an idea of some of the uses of geography that we envisage because that would make of it a keystone of the new organisation [15]. A second reason for introducing this issue in this paper is that these new objectives are going to be implemented very soon; for instance, the development of a new product, the « *répertoire d'immeubles localisés*[16]», designated by its acronym *RIL*, is already under way because it appears to be worthwhile even if the new Census methodology is not to be implemented.

14 The *Conseil d'Etat* was solicited by the government for advice on this subject in April 1998 and gave an affirmative answer back in July.

15 A more comprehensive description could be found in [ISNARD].

16 Building locations register

56 Without going into the details of the procedure, it is necessary to understand that the universe of the 36,000 *communes* would be divided into two subsets :

- Traditional censuses would be taken in small *communes* (the threshold could be 10,000 inhabitants) but with a rotating sample on a five years cycle.

- Surveys on a 1/20 sample would be taken each year in each town over this threshold.

57 Yearly estimates would be delivered by about every *commune* based on this partial data collection combined with external data. These external data could be nominative or not, this point is not clear to day and their uses could be slightly different depending on this decision; they are likely to be occupancy tax data that we already use for the '99 Census (see before II.3) and social security data which we have also access to for specific operations involving *INSEE* and related to electoral control.

58 To sample small *communes*, we could use GIS facilities in order to take into account some of their topological characteristics. But the main interest of GIS would seems to be for managing data collection in towns, especially sampling and collection management.

III.1.1 Sampling in towns

59 The yearly 1/20 [17] sampling should be representative of the town. To ensure local representativity at a local level to be defined, the sample should be spread over the whole territory of the town. And the five samples over the 5-year cycle should be independent one from the other, because collected data will be cumulated on a cycle to improve their representativity. These conditions are naturally leading us toward an increased use of GIS.

60 No exhaustive list of housing units exists in France (the occupancy tax file mentioned before can be used for improving exhaustivity but not for sampling). We think it is too difficult to obtain such a list and especially to maintain it through time. That could also raise confidentiality issues difficult to deal with. Thus, we cannot build a sample frame of housing units after adding necessary geographic properties to such a list.

61 The intention is therefore to go through area sampling, in some way like we are used to do in the Labor Force Survey or like Germans do with their Mikrozensus. But unlike LFS, the two processes (delineating areas, sampling into areas list) should be industrial, that is to say automated and with quality checks, because of their size and their consequences on the results.

62 Unlike a list of housing units, to set up and to maintain a list of

building units appears to be technically and budgetary feasible. There is no confidentiality issue, and this population is very stable. This list will be materialized as a supplementary layer of our current GIS *Base-Îlot*.

63 Several questions should receive a response (see below) :

- How to create areas on the basis of this layer ?
- How to give the right information to enumerators ?
- Could we manage collection differently ?

64 and obviously :

- What are the technical characteristics for this supplementary layer?
- Which are the investment and maintenance costs and procedures?

III.1.2. How to create areas ?

65 Let's suppose there are only small buildings with a small number of housing units. Area design would be easy by grouping together contiguous buildings up to a certain number of housing units (10 for instance). This process was tested in 1997 in one of the French overseas *département* (*La Réunion*) where the settlement has effectively this characteristic unless exceptions. This test showed that the area delineating process and the associated sampling process can be automated using topographic information included into the GIS.

66 As for larger buildings, the intention is to get enough information to be able to divide them in smaller subsets of apartments, probably by floor. Moreover, this category of buildings would be treated separately, and specific areas built comprising only parts of important buildings would constitute one or several strata.

III.1.3. What information for the enumerators ?

67 These enumerator maps for the '99 Census were described before, as well as the internal debate we had about how detailed should be the information to allow them to make a good job. The new perspective should fuel again this debate, areas visibility on the field is obviously less obvious than *îlots* visibility. A provisional conclusion is that we need maps with building locations, and that comments [18] or even sketches elaborated after a first survey might be useful in some cases.

17 This rate would be a mean rate. It could rather be 1/5 in little towns and 1/25 in the largest ones.

18 For instance: «to access this building, go through the entrance of building N° xx and go to the right end of the courtyard beyond the edge».

III.1.4 Collection management in towns: controlled quality zones

68 Time availability would characterize the new process compared to the traditional one [19]. At the same time, demand for quality assessment at the national but as well at any local level is strong, and would be probably stronger especially because local authorities are getting more and more political responsibilities and appetite for data and less possibility of interfering with data collection in this new kind of Census.

69 There is a large range of criteria to evaluate the quality of local results: exhaustivity or, in this case, sample expansion; raking techniques; underlying quality of external sources; socio-demographic heterogeneity of the population; collection difficulties in the zone measured by non-response rates; ... Taking enough time for that and having at our disposal some tools like sample size adjustment, enumerator training, collection preparation on the field making use of specific relations with local medias, imputation procedures, ... , *INSEE* could react to quality flaws. That is another of the significant advantages of the continuous Census.

70 One hypothesis is now to design «controlled quality zones» where INSEE should be committed to quality level. If this hypothesis were chosen, GIS would again be a basic tool for analysis and display of results.

III.1.5 Technical characteristics of this supplementary GIS layer

71 *Base-Îlot* will include a new layer, the *RIL*. This layer will comprise points representing building locations. We are not going to put lot limits or surfaces corresponding to the building shape as they are in cadastral maps, but only points. These points will be located approximately at least for the coming years. We cannot afford a precise (let say 2 m) positioning, and that is not essential for survey or Census taking (see the 1993 test before). What we consider important are exhaustivity, order of succession along a street segment, and localisation on the actual side of the segment. Later, we think of improving this positioning where that appears useful to avoid collection difficulties, for instance by using cadastral maps or orthophotos as a background for digitization or GPS.

72 Very few data will be tied to the points on the *RIL*: existence of housing units and an estimation of their number coming from the last survey or Census, probably existence of other purposes units (store , ...), comments likely to help enumerators to find the unit, and identification numbers for the linkage with statistical files used for Census taking.

---

19 The collection period would be extended from one month in the classical Census to probably 8 or 9 months over the academic year, avoiding only holiday's period. That gives INSEE the possibility of collecting data without the help of local governments.

III.1.6 Investment and maintenance costs and procedure

73 The adaptation of the software developed for *Base-Îlot* management is currently under way. It will be operational in September 1999, in time to allow the integration of the data coming from the '99 Census (about 4.5 millions of buildings). This integration will be made in 2000 for most of the towns, using MapInfo facilities for automatic geocoding on the basis of postal addresses as they are captured during the Census. Tests are under way to evaluate the number of the non-automatically coded addresses that we must expect. The processing of these addresses is evaluated for the moment at 30,000 work hours.

74 Maintaining this geographic database implies continuous relationship with local authorities in charge of defining and implementing urbanistic regulations. *INSEE* maintain already such relations but not everywhere on a continuous basis: sometimes they are limited to the Census necessity of updating *Base-Îlot* and delineating the *îlots*. As a continuous work becomes mandatory, we need to build formal partnerships with local authorities.

75 Other sources are also considered. For instance, water supply companies (or other utilities suppliers) are well informed about any geometric changes of the street network, and are updated about building constructions and demolitions, that is important for the *RIL* maintenance.

76 Field work for the «new kind» of Census would also constitute a third source of information, but we prefer to look at the GIS (including *RIL*) as an input to the process rather than as an output because of sampling considerations.

77 Our present partnership with *IGN* is to be reset. It will probably not be extended to the *RIL* and will remain limited to the current *Base-Îlot* scope. The new basis of this partnership will probably be different: due to our continuous relation with primary sources, we have the potential to able to update *Base-Îlot* by ourselves (street geometry, street names, postal addresses and as now *îlots*); *IGN,* who is more skilled in precise cartographic work, would intervene with some delay to improve the geometric quality of the data base.

78 Yearly maintenance costs are evaluated at 80,000 work hours, half of them being required for the *RIL*. The annual number of changes in the *RIL* (building creations and suppressions) should be less than 200,000.

III.2 Extension to other purposes

79 Besides residential buildings, there are also industrial or commercial entities in specific buildings or mixed with housing units, and various public facilities (playgrounds, hospitals, schools , ...). For the new style Census canvassing, enumerators would make a better job if they know the positioning of the buildings to be surveyed relatively to the others of any

kind. Beyond this first reason for extending the scope of the RIL to other entities, there are specific reasons for every category.

80 Managing the Business Register suppose to get precise addresses which should refer to existing buildings; this control is not done thoroughly at present (only the existence of the street is checked) but the *RIL* existence will make it possible. Disseminating data from the Business Register [20] suppose to meet the demand of precise localisation of the units; that is currently done afterwards at the *îlot* level but it would be cost effective to make it straightaway when coding a new unit.

81 There is also a demand for data about localisation of public facilities. A first survey was made in 1993, but building the *RIL* gives us the possibility of renewing the methodology and producing the up-to-date data which are in demand to complete Census data on population characteristics. Again these units are not mobile and the maintenance cost should not be important.

82 For all these reasons INSEE decided to go towards *RIL* implementation, as soon as possible after the '99 Census collection phase, even if a decision about the future Census cannot yet be made.

20 An important output of *INSEE* because it is not a «statistical» Business Register, but an administrative one and the statistical law does not apply to its whole content.

COEFFIC, Nicole, L’enquête de mesure de l’exhaustivité du recensement de 1990, Recensement de la population 1990 - Innovations méthodologiques, Novembre 1995, INSEE-Méthodes n° 52-53

DEVILLE, Jean Claude, L’échantillonnage équilibré de districts et l’exploitation au 1/20 du recensement, Recensement de La population 1990 - Innovations méthodologiques, Novembre 1995, INSEE-Méthodes n° 52-53

DEVILLE, Jean Claude, JACOD, Michel, (1996) Replacing the traditional French Census by a large scale continuous population survey, Annual Research Conference, March 1996, Washington, Bureau of the Census

FANOUILLET, Jean Christophe, XXX, Recensement de la population 1990 - Innovations méthodologiques, INSEE-Méthodes n° XXX (to be published)

ISNARD, Michel, (1997), Post 2000 Census Research Project in INSEE, Working paper, available from the author at «michel.isnard@insee.fr»