UNITED NATIONS

E

Economic and Social Council

Distr.
General

ECE/CES/GE.41/2007/12
25 May 2007

Original: ENGLISH

ECONOMIC COMMISSION FOR EUROPE

CONFERENCE OF EUROPEAN STATISTICIANS

Group of Experts on Population and Housing Censuses

Tenth session
Astana, 4-6 June 2007
Item 3 (b) of the provisional agenda

CENSUS TECHNOLOGY: RECENT DEVELOPMENTS AND IMPLICATIONS ON CENSUS METHODOLOGY

Innovations in census mapping and census data geocoding

Submitted by Italy*

This meeting is organised jointly with Eurostat.

Summary

The Bureau of the Conference of European Statisticians (CES), at its meeting held in Washington, D.C. (United States) on 19-20 October 2006, approved the renewed terms of reference for the Steering Group on Population and Housing Censuses and the plan for future CES activities on population and housing censuses. The CES Bureau also agreed that the Steering Group would coordinate the work on the diverse types of meetings. The present paper was prepared on request by the Steering Group on Population and Housing Censuses, for presentation and discussion at the Joint UNECE/Eurostat Meeting on Population and Housing Censuses in Astana (Kazakhstan), 4-6 June 2007. The paper provides substantive basis for the discussion in the session of the meeting dedicated to "Census technology: recent developments and implications on census methodology".

* This paper has been prepared by National Institute of Statistics of Italy (ISTAT) at the invitation of the secretariat.

I. INTRODUCTION

1 In the last years several countries have increased their efforts to move from a traditional approach to new approaches, as register-based ones, combination of conventional census and sample survey, combination of register based census and sample survey, rolling sample, other emerging approaches[1].

2. The Board of the National Institute of Statistics of Italy, in occasion of the 2006 reorganisation of the Institute, established a new directorate, the Central Directorate for General Censuses (DCCG). The mandate of DCCG, is to determine pros and cons of new approaches, emerging census methodologies and techniques, organizational options, to allow to take the most appropriate decisions for the next census round in Italy.

3. DCCG launched several studies to identify factors, which are critical in explaining the quality of census data in recent census round. It was showed that demographic size of municipalities is one of the most important risk factor. In Italian population and housing census the fieldwork is traditionally a task of municipality offices, which, immediately before the census date, have to recruit the enumerators and additional staff. Offices of the larger-sized municipalities[2], in a relatively short period, have to face a huge increase in human resources to be trained and employed in census fieldwork. Moreover the impact of municipality-size on data quality is expected to be much stronger in the incoming population and housing census round when it will become more difficult to find people at home. To face this problem it is necessary to organise fieldwork and to include actions aiming to improve quality control especially in large-sized municipality. Two projects may be of crucial importance to plan these actions: a feasibility study to produce geocoded lists of address numbers and a project to define new area units in census mapping which may allow to employ area sampling methods in a Long-Short form approach.

4. According to the *Regolamento Anagrafico*[3], besides the coordination of census fieldwork, offices of municipalities have the task to maintain: a) the boundaries and codes of census mapping units (enumeration areas and localities) in the latest version validated by the National Institute of Statistics; b) a list of address numbers. It is worthwhile to remark that, among the municipalities, there is an high variability in the contents, format and quality of these data. These

---

[1] United Nations, (2006) *Principles and Recommendations for Population and Housing Censuses*, NewYork.
UNECE, *CES Recommendations for the 2010 censuses of population and housing*, jointly prepared by the United Nations Economic Commission for Europe and the Statistical Office of the European Communities

[2] In this study, the large-sized municipalities are considered the municipalities, which are main town of province or in the class "50.000 inhabitants or more". In the last population census large-sized municipalities were 159 (2 per cent) covering almost the 36 per cent of the italian population.

[3] The italian law on population registers. (D.P.R. n.223 del 30 maggio 1989), ISTAT, Metodi e Norme serie B – n.29 1992.

issues will be discussed in the next two paragraphs. The last paragraph is dedicated to some concluding remarks.

II GEOCODED LISTS OF ADDRESS NUMBERS

5. A first approach to produce a good list of geocoded address numbers is possible when a National Address Point System is available[4]. Some countries have developed a National Address Point System in which each building address is represented by a coordinate in a proper geographic reference system[5]. A National Address Point System could be create either collecting a coordinate for each building in the country using field techniques, either by digitizing from available small-scale topographic and city maps. Availability of a National Address Point System is the most favorable context and the resulting georeferenced frame would in this case represent a reference system of strong detail for censuses. The points that represent households for which census data are available can then be aggregated to any desirable zones, using simple GIS operations and a National Address Point System may be employed in census field operation to list addresses for any EA or other territorial units of the hierarchical census geography. The completeness of lists obtained by National Address Point System strongly depends on the National Address Point System quality and update. A National Address Point System project requires to employ a considerable large amount of financial and human resources usually not available in the census budgets. It is therefore usually only employed in countries where other authorities, such as the postal office, also have an interest in creating such a database.

6 A "second best" approach is possible when a National Street Network is available. Often, it is the private sector that is creating a street network for commercial and marketing applications. A National Street Network consists of streets and road segments, a stretch of a street between two intersections, where an intersection is defined as the place where street segments meet or where the street changes its name. For each line segment in the street database, the range of address numbers is recorded for both sides of the street, usually the numbers on one side of the street are even-numbered, and those on the other side are odd-numbered. Employing Gis and National Street Network, each street address from a list may be evaluated and matched to a location on the corresponding street segment. The location determined by interpolation of the address number in proportion to the range of addresses on the street segment. The result is not exact, but usually the approximation is quite good to be used for census purposes.

7. It must be remarked that a National Street Network itself, does not allow to produce a good list of addresses for each EA. Employing National Street Network it is possible to produce a "pseudo-list" including all the addresses in the range of the address number of each road segment, but this "pseudo-list" is usually of poor quality. Number in the range which are lacking

---

4 United Nations (2000) *Handbook on geographic information systems and digital mapping*. United Nations Publication, ST/ESA/STAT/SER.f/79. Printed in United Nations, New York, 2000.

5 This is the case for example of the Ordnance Survey in the United Kingdom. Ordnance Survey (1993). *Address-Point User Guide*. Southampton, United Kingdom: Ordnance Survey.

in the field are counted, while number repeated with different index (as 40a, 40b, etc.) are counted only once. When a good external list of addresses is available, a National Street Network may be employed to approximate location of addresses in this list and to assign them to EA. The quality of the resulting lists for EA strongly depend on the update and completeness of the external list, on the update, completeness and geographical quality of National Street Network, on the geographical quality of EAs boundaries.

8. A National Address Point System it is not available in Italy. National Street Network are available only from commercial sources. In the last two round of censuses it was required municipalities to produce a form in electronic or paper format describing the "enumeration path" (ISTAT Mod CP5) for each EA. The form contains the list of road segments, portions of streets in the EA, and the ranges of address numbers in each segment. A copy of this form was sent to ISTAT and was used to check coherence of EAs codes in census questionnaires and ranges of the addresses numbers in the ISTAT Mod CP5.

9. Even if National Address Point System and official National Street Network are not available, most of large-sized municipalities developed Local Address Point Systems and Local Street Networks. Due to latest technological developments in Geographical Information Systems (GIS), World Wide Web services, Global Positioning Systems (GPS), cartography, mapping tools, the availability of geographic data in digital format at local level highly increased in recent years. Data may be more easily combined with other data to create relevant information for decision making and spatial analysis and geographic information, in its digital form, is exchanged more rapidly. Some municipalities may have already a geocoded lists of addresses. In other case a Local Address Point Systems could be employed to produce geocoded list. When it is available neither an already geocoded list neither a Local Address Point System, it would be possible to use external lists of addresses[6] and geocode them with the help of a commercial National Street Network.

10. This suggest a possible innovation for the next census round to be introduced at least in the larger-sized municipalities: using local Address Point Systems or local information on addresses the objective is to produce, in advance of enumerators fieldwork, a comprehensive list of address numbers geocoded to EAs. The quality and update of the list should be verified in advance as well. For this reason ISTAT launched a feasibility study for an archive of addresses geocoded to EAs. In some municipality[7] local and commercial sources were employed to build experimental integrated archives. To the purpose of planning a process of local data integration, it is necessary to know precisely contents, formats, and characteristics of geographical data, including lists of georeferenced addresses, street network, etc, available at municipality level. To investigate on these issues, a web survey on large sized municipalities was recently launched.

III CENSUS MAPPING

[6] An external list which could be employed to this purposes is the address number list of PosteItaliane, the Italian Postal Service Company.

[7] Including Firenze, Bologna. In 2007 the experiment will be extended to several municipalities, as Cuneo, Milano, Palermo, Pesaro, Trapani, etc.

11. Defining the census mapping is a crucial pre-enumeration work having big consequences on census management and enumerators fieldwork[8]. Main objective of this work are: to define lists of territorial units included in the census geographic hierarchy, to delineate their borders, to define the coding schemes, to code each unit. Usually the census geographic hierarchy includes three level: 1) Administrative borders (Country, Regions, Provinces, Municipalities); Localities (Morphological Areas); 3) Enumeration Areas (EAs)[9]. The EAs are designed to allow to easily identify boundaries of areas assigned to each enumerator, to facilitate counting of statistical units and to reduce the possibility of double counting. Natural boundaries for EAs are streets, railway lines, hydrological features such as rivers and lakes.

12. Innovations in census mapping of the last two decades aimed: 1) to build comprehensive census mapping databases in digital format; 2) to improve census mapping design using as possible new available images and cartographic sources and Gis applications to check data, and to better integrate data sources. In 1991 Italian Census, for the first time, produced a digital database covering all country at least in scale 1:25.000. The mapping process was mixed, the design of localities boundaries by interpretation of remote sensing images and the design of administrative boundaries from maps of IGMI (The Italian Military Mapping Agency) were centralized, the delineation of EAs by small-medium scale topographic and city maps was decentralized. The checking processes was done jointly by municipalities and ISTAT. The 2001 census mapping project was mostly centralized, started from the 1991 digital database, and combined several geographical sources, digital aerial photos at scale 1:10.000, local topographic maps, streets networks, railways networks, hydrological features as rivers and lakes, to update boundaries and increase positional quality. Main objectives were: i) to redesign and reduce in size census sections in extra-urban areas to get the integration of all censuses mapping, including for the first time agriculture census.; ii) to improve geographical quality and integration with other territorial databases of public interest. At the first stage localities were mapped [10]. In the second stage, using different techniques for each type of locality, EAs were mapped. A preliminary proposal was sent, in a first phase, to local authorities both in CD-Rom and printed format asking them to check and modify boundaries when necessary. In final further centralized long and costly phase maps were checked again and changes digitized.

13. For the next round, given the good quality of the digital database built in 2001, the plan is to maintain as possible the 2001 boundaries and codes of EAs and localities with the objective to concentrate resources to produce lists of address numbers and to delineate Census Areas (see next paragraph). New built areas and new buildings in existing built area requires in any case to split some EAs and change some boundaries of localities. Some EAs boundaries will need updates and corrections during enumeration. The challenge is to make all changes which are strictly necessary, taking advantage of the rapidly and constantly evolving technological framework in computer networking, in sharing web resources and Gis applications. Tools for web cooperation could be employed to digitize, edit and print maps preserving topological

---

[8] United Nations (2000) *Handbook on geographic information systems and digital mapping*. United Nations Publication, ST/ESA/STAT/SER.f/79. Printed in United Nations, New York, 2000.

[9] EAs (*Sezioni di censimento*)

[10] Types of localites: 1) Centri abitati, 2) Nuclei abitati, 3) Case sparse (scattered buildings).

properties and, whenever possible, use automatic checks. Changes in the boundary of a EA should be done only if at least a pre-determined number of new address numbers have been created inside the EA after last censuses.

14. The delineation of Census Areas[11] in the larger sized municipalities is a further innovation proposed for the next census round[12]. The scope is to allow in the main locality of these municipalities the use of Long and Short Forms sampling techniques. This delineation will be part of the census mapping project and Census Areas will represent a new unit in the mapping hierarchy between the "Centro abitato principale", the main locality of the municipality, and the EAs. Census Areas will be delineated in partnership with municipalities offices and will be compatible as possible with areas defined at municipality level (Municipality Areas), as wards, functional zones, etc[13]. The target population size is 10,000 inhabitants, the target expected range in population sizes is 7,000-15,000 inhabitants.[14]

15. The process to delineate Census Areas is divided in two steps: in the first step the Basic Areas are defined. Basic Areas are obtained inside the main locality by overlay of first order natural features[15] and municipality areas boundaries[16]. Basic areas may be then divided in three classes: i) Over the maximum threshold (15,000 inhabitants); ii) Between the minimum and maximum threshold; iii) under the minimum threshold (7,000 inhabitants)[17]. In the second step is employed a set of constraints to accept a Basic Area as Census Area, an area in which is possible to employ sampling techniques. In the i) case basic areas may be divided using when available second order natural boundaries or second level municipality areas. When constraints are not satisfied it is possible to employ segregation tecniques[18] and spatial autocorrelation[19], to verify if

---

[11] Bianchi G., Di Pede F. et al. (2007) *Processi per la definizione di aree di censimento sub comunali da utilizzare per il Censimento della popolazione e delle abitazioni del 2011*, Paper submitted to the XXVIII Conferenza Italiana di Scienze Regionali, Bozen, 2007.

[12] The extension to municipalities over 10,000 inhabitants will be evaluated.

[13] At municipality level local-administrative zones and/or functional zones may be defined. Among them municipi, circoscrizioni, quartieri, zone di decentramento, zone statistiche, aree elementari, unità urbanistiche, zone toponomastiche, etc..... The criteria of delineation and minimum-maximum thresholds are not harmonized.

[14] These thresholds are provisional. An ongoing study on sampling estimators performances using 2001 census results as benchmark will define the final thresholds.

[15] Natural features are streets, railway lines, hydrological features such as rivers and lakes. Natural features are employed to delineate EAs. It is possible to classify natural features selecting the most important of them by size and relevance, these are here defined first order natural features.

[16] These ensure that Basic Areas are a sum of contiguous EAs.

[17] Fig.1 shows the results of an experimental study for Florence Census Areas delineation.

[18] Wong, Davis W. S. (2003) *Spatial Decomposition of Segregation Indices: A Framework Toward Measuring Segregation at Multiple Levels.* Geographical Analysis - Volume 35, Number 3, July 2003, pp. 179-194.

[19] Goodchild M. F. (1986). *Spatial Autocorrelation.* Catmog 47, Geo Books, Norwich.

it is possible to merge contiguous Basic Areas to form a Census Area, which satisfy constraints. At the end of this process not all EAs in the municipality will be part of a Census Areas. In particular all the EAs outside the main localities and the EAs of Basic Areas for which it is not possible to find a merging which satisfy constraints will be not included in a Census Area [20].

16. The Census Areas, so delineated, may allow the adoption of sample strategies as Long-Short form. In fieldwork organization, Census Areas could be employed as supervisory area, a set of contiguous EAs to be assigned to one supervisor. Census Areas may be employed as output areas or reporting zones in data dissemination, and, in a medium term perspective they may be used as target zones on which, starting from census outputs, it will be possible to increase local data availability integrating existing data from local sources.

IV CONCLUSION

17. The innovations described in this paper may represent powerful tools to highly improve the quality control during fieldwork. Main phases of census design and fieldwork would be fruitfully supported by the availability of a good quality and list of geocoded address numbers and good quality definition of the census mapping. A list of geocoded address could be employed:

a) to find EA in which at least a pre-determined number of new address numbers have been created after last census and consequently need to be redesigned in the census mapping update;
b) to keep under control the expected number of buildings in each EA;
c) to preprint lists of address numbers for enumerators;
d) to allow enumerators to find, to enumerate and to survey statistical units; e) to allow fieldwork and centralized check of coverage.

18. Besides, a list of geocoded address numbers may be employed in many phases of data presentation and analysis:

e) geo-coding of statistical units on maps;
f) data presentation on maps,
g) to match data of several archives;
h) location-data are used to geo-reference units of archives in the GIS geographic information systems.

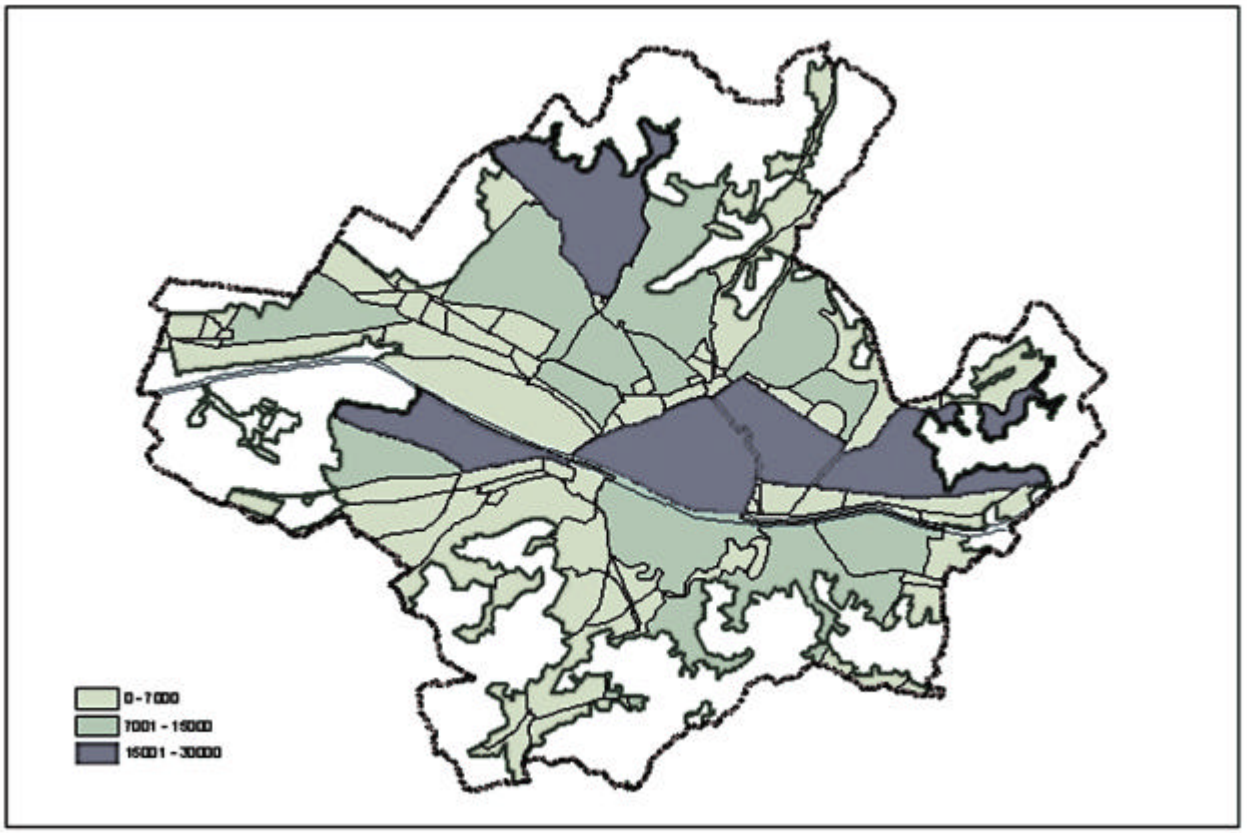
19. The delineation of Census Areas represents a further important innovation to be used in next census round:

i) in census design, Census Areas may allow the adoption of sample strategies as Long-Short form, which may reduce burden on respondents;
j) in fieldwork organization, Census Areas may be employed as supervisory area, a set of contiguous EAs to be assigned to one supervisor;

[20] This is not a problem for the adoption of area sampling. The EAs not included in Census Areas may be assumed as EAs self-representative.

k) in data dissemination, Census Areas may be employed as output areas or reporting zones, useful also for confidentiality data protection. In a post-census perspective they could be used as target zones on which, starting from census outputs, integrate all available local data.

Figure.1. Basic Areas of Florence. In white the areas outside the main locality, in dark grey the Basic Areas over the maximum threshold, in light grey the Basic Areas under the minimum threshold, in grey the Basic Areas between the minimum and maximum threshold.



*****