**UNITED
NATIONS**

E

Economic and Social Council

Distr.
GENERAL

ECE/CES/GE.41/2007/SP/3
29 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**

**Coverage and quality issues in the preparation of the Population and Housing Census in
Kosovo: The role of mapping, IT and logistics**

Submitted by Statistical Office of Kosovo-UNMIK and TA project "Support to Statistical
System and Preparation on Census, Kosovo"

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 for 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".

GE.07-

I. SUMMARY

1. The Statistical Office of Kosovo (SOK) is carrying out the preparatory activities for the realization of the next population and housing census that will provide a new demographic and socioeconomic portrait of the region, after recent changes in the amount, distribution and characteristics of its population. It will be based on a traditional approach following the enumerator method.

2. In the absence of reliable administrative records, comparable previous census data and trustworthy household surveys, a set of technical preparatory census activities were considered particularly important for improving the coverage rate of the foreseen Kosovo census. The adoption of accurate quality control procedures in census preparatory activities is a crucial element in the absence of "traditional" tools to plan a successful census. These factors are also expected to improve data accuracy at small area level and to ensure public and institutional confidence in census results.

3. This paper aims at describing the activities carried out on mapping, IT and logistics, and their evaluation trough the pilot census undertaken in November 2006, underlining measures already adopted and further activities to maximize the completeness of coverage.

II. INTRODUCTION

4. The results of a population and housing census are always affected by a degree of uncertainty on the coverage of dwelling units, households and individuals. Coverage errors refer to the non inclusion or double-inclusion of units that may originate under-counting or over-counting, respectively. They may originate in the enumeration phase or during data processing procedures, from operational errors.

5. The occurrence of these errors may be reduced in the census preparatory work mainly by the realization of updated and well organised digital census maps in a GIS environment, and by well planned and organised data processing techniques, training activities, and fieldwork operations. Other critical factors may also affect census coverage: the way of designing questionnaires, the definition of usual residence, and the effectiveness of the communication campaign; these aspects, however, are not described in this paper.

6. Data processing procedures are described also with respect to the detection and correction of inconsistency errors that may originate by respondents, enumerators, supervisors and data entry operators, in order to provide an overview on data quality control as well.

III. THE ROLE OF MAPPING

7. Since the beginning of 2005, the Cartography and Geographic Information System (GIS) Unit of the Statistical Office of Kosovo (SOK) has developed a geo-referenced cartographic database for the preparation of census maps as a component of a comprehensive digital spatial infrastructure, in a GIS environment. The main objective was to create a geo-referenced system for the support of planning and monitoring activities of census operations and for the dissemination of spatially aggregated census results, including data for small areas.

8.   On the basis of ortho-rectified colour aerial photos taken in 2004, the administrative boundaries (municipalities and cadastral zones) and limits of geographic entities (settlements) were digitalized, named and coded. About 5,000 census enumeration areas (EAs) were demarcated and coded, and around 520,000 buildings identified mainly through visual interpretation in aerial images, their perimeters digitalized and coded with unique ID numbers.

9.   A first classification between residential and non-residential buildings was implemented trough visual interpretation from aerial images, as well.

10.   The EAs were aggregated to form areas pertaining to one controller, and, as a further aggregation of the latter, to one supervisor, in order to plan fieldwork activities. The main roads were also represented with their corresponding name, and some points of interest were also entered into the system to facilitate the orientation of enumerators in the field.

11.   The EAs were demarcated on the basis of the amount of dwelling units estimated in each "potential" residential building and taking into account household estimates by populated areas derived by information from local authorities and local communities, in connection with the average household size in urban and rural areas, as obtained from a previous household survey. A list of multi-storey buildings was also provided by local municipalities.

12.   As a result, SOK estimated that the number of households per EA would be between 40 to 60 households in rural areas and from 80 to 120 in urban areas.

13.   The system was tested in 17 EAs of three Kosovo Municipalities, Pejë/Pec, Kamenicë/Kamenica, and Štrpce/Shtërpcë during the 2006 Kosovo pilot census carried out from 30 October to 14^t November 2006. The test was carried out by SOK in collaboration with experts from the EU funded technical assistance project "Support to the statistical system and preparation for the census, Kosovo", implemented by a consortium composed by the Italian National Institute of Statistics (ISTAT), a research centre (CIRPS) of the University of Rome "La Sapienza" and the Albanian National Institute of Statistics (INSTAT).

14.   The objective of the test, for what concerns census mapping, aimed at analysing mainly three aspects in relation to the census coverage issue:

    a)   The appropriate demarcation of EAs to verify that the amount of households was approximately equal in rural and urban areas;
    b)   The amount and typology of physical changes occurred at building level, given that aerial photos dated back to 2004;
    c)   The adequacy of paper census maps, prepared in Albanian and Serbian languages, for the organization and implementation of fieldwork operations

15.   The enumeration was anticipated by a two days fieldwork activities in the selected EAs carried out by enumerators, in order to identify and classify all buildings and to update the corresponding census maps. This activity was particularly important to analyse building changes occurred since 2004, and differences between the data on the identification of residential buildings made through aerial photos and the one obtained with fieldwork investigation. All changes were recorded in the *building list form* (BL1).

16.   The data by EA used for the analysis were:

a)   The recorded number of households, individuals and dwelling units;
b)   The estimated number of households and buildings prior to enumeration;
c)   The recorded number of buildings with one or more private households;
d)   The number of buildings classified as residential and non residential prior to enumeration;
e)   The number of buildings classified during field operations in the conditions of in state of ruin;
f)   The number of buildings not existing anymore at the census date, but reported in 2004 orthophotos;
g)   The number of buildings built since 2004;
h)   The number of buildings under construction.

17.   The results of the analysis on the data and activities undertaken showed that:

a)   All buildings were covered by visits of enumerators and that in few cases households were found residing also inside non-residential buildings;
b)   Approximately 10% of physical changes (buildings not existing anymore or in state of ruin, new buildings or buildings under construction) occurred at building level in two years period in the considered EAs, and that those changes were mainly concentrated in urban areas with a significant presence in selected rural areas;
c)   Although the criteria used to demarcate the EAs gave on the whole good results (on average, the overestimation was equal to only 8 households per EA), many differences were recorded in some urban areas due to the large amount of not inhabited dwelling units, and in few rural areas were the number of households was much greater than expected.

IV.   THE ROLE OF IT

18.   The Data Intake System implemented at the Statistical Office of Kosovo and tested with the entry and checking of the Kosovo pilot data, is a client/server application developed in CSProX environment. This application allows the simultaneous recording of questionnaires from different clients to a MySQL database. Taking advantage of the client/server technology, this system increases data security and provides for greater control over the processing of questionnaires.

19.   The most important characteristics of the Data Intake System are:

a)   The capacity to record data from the Building Form (BL1) and the Record and Summary Form (RS1), allowing the linkage - for quality checking purposes only - between these data and the data contained in the questionnaires;
b)   The automatic storage of errors (inconsistencies, missing values, overridden skips) detected during the census form's data entry, in a specific database section;
c)   The run of a batch editing procedure, using as input the stored data, and the production of an organized report of all the inconsistencies, overridden skips or missing values ignored by the operator;

d) The recording of any change performed on the original data, change by change, in a specific section of the database in order to ensure the transparency of the whole data intake operation;

e) The keeping of a copy of the data entered during the data entry operation and a copy of the same data after batch editing and data change;

f) The availability of utilities to export data in SPSS, SAS, and STATA format.

20. Taking into consideration that at present no reliable data is available in Kosovo on dwellings, households and individuals, the only coverage analysis that could be carried out entailed the comparison of the data collected through the questionnaires, with the data recorded in the BL1 and RS1 forms.

21. As mentioned above, the BL1 form is a pre-coded list of buildings at EA level derived from the GIS database containing 2004 data. This list was used by enumerators to re-classify all the buildings (residential, partially residential, not residential, not existing anymore, in state of ruin) and to list new buildings. This list was updated in the field, during the 'Preliminary Visit', three days prior to the beginning of the enumeration.

22. The RS1 form is a kind of tracking form where enumerators write the codes of each dwelling and household visited during the fieldwork. For each household the number of members by sex is also reported. The form is basically to be used for keeping track of the dwellings visited and those to be re-visited, and to keep the daily summary of persons interviewed. This list was compiled by enumerators at the moment of the identification of the dwelling and before filling the questionnaires.

23. RS1 and BL1 form were crucial to check the completeness of the subsequent data entry of questionnaires. The comparison between BL1 data and the data form questionnaires at EA level allow to verify that all the buildings classified as "residential" or "partially residential" have at least one household inside.

24. Furthermore, the availability of the ID codes of each dwelling and household in the RS1 form makes it possible to check the records coming from the census forms with the ones coming from the RS1 form. This procedure allows a one by one check for the dwelling and the household records and, as a consequence, a check of their total figures at EA level. Moreover, it is also possible to check if the total amount of persons recorded for a specific household in the RS1 form is the same of the one in the census forms.

25. When differences were detected, it was easy to identify if they occurred because an operator forgot to enter a questionnaire, or the questionnaire never arrived to the data entry room, or if a mistake was made by the enumerator while compiling the RS1 form.

26. Even if both checks are not exactly a "measure of coverage", which indeed pertains to the post-enumeration survey, they were the only one that could be performed in a situation like Kosovo, which where no previous information was available on the distribution of the households and dwellings in the territory.

27.   In any case, the comparison between BL1, RS1 and questionnaires data was essential to guarantee the completeness of data intake operations.

28.   In relation to the improvement of the data quality, it is important to underline that the system was designed to perform several on-line consistency checks during the entry of questionnaire data, alerting the data entry operator about any kind of errors occurred during entry operations. However, the application leaves the decision to the operator on whether the errors have to be fixed or not. Moreover, the system automatically executes conditional and unconditional skips, leaving to the operator the choice to override each single skip when deemed necessary.

29.   While the data entry of a specific EA was completed, data entry operators performed the batch edit procedure. This procedure provided a report at EA level of all inconsistencies, overridden skips or missing values that, although detected during the data entry, were ignored by the operator.

30.   The last operation consisted of checking each error reported in the batch editing report against the value written in the paper questionnaires. At this stage of the work the rule was that corrections to the data were performed only if a difference between paper and recorded values was found. It is also to be underlined that all changes performed were recorded in a specific table of the database together with the ID of the questionnaire, the field in which the correction was done, the original value and the new value.

31.   Following the steps described above, it is reasonable to believe that all remaining errors are most likely not due to data entry mistakes but, mainly, to problems occurred during the interviews. Finally, all remaining errors, as inconsistencies, overridden skips or missing values, were stored in a specific table together with the ID code of the questionnaire and the field where the error appeared.

32.   In this view, the final stage of the data processing should be the imputation phase in which any inconsistency, overridden skip or missing values are substituted by a random value selected from a set of donors.

V.   THE ROLE OF LOGISTICS

33.   A plan of action to organize and manage fieldwork operations was prepared by the SOK Census Logistics Unit. It was primarily devoted to the distribution of census materials in the field and their return to SOK's HQ for data processing, and extended to the recruitment of enumerators, controllers and supervisors, to the organization of training activities and to the overall organization of fieldwork operations. The plan was tested during the November 2006 pilot census.

34.   Before the commencement of the operations, census materials in English, Albanian and Serbian languages were printed, packed and transferred to the Census Municipal Offices (CMO) of the selected Municipalities, were questionnaires and additional forms, manuals, maps and the other relevant materials have been delivered to supervisors, controllers and enumerators. SOK staff from the regional offices and representatives of the Local Governments were taking over the security and privacy of the materials stored in Census Municipal Offices.

35.   Training activities were organized at SOK HQ in Pristina in two different phases, one for supervisors and controllers, and one addressed to enumerators and attended also by supervisors and controllers. In both sessions, English, Albanian and Serbian languages were used. The two days training for supervisors started one week before the beginning of operations, while the second phase lasted three days and ended the day before the commencement of fieldwork activities, on 27th October.

36.   The timing of training activities was planned having as a general rule that it should be provided as close as possible to the date of commencement of field operations. Even though the planned organizational structure of the training was following a cascade scheme, whereby census experienced staff from the Statistical Institute train supervisors, and supervisors train controllers and enumerators), at the initial stage of the training it became evident that supervisors would have not provided a qualitative training to enumerators; in fact, it was perceived that supervisors would have require longer training sessions themselves and specific skills in training methods. This outcome led to the decision to extend to the supervisors the training provided to enumerators from SOK senior staff, with the support of experts from the EU funded technical assistance project.

37.   The recruitment of field staff was undertaken by Municipalities, with the support of SOK regional offices. For each EA one enumerator was designated, and it was agreed to have one controller for seven EAs and one supervisor each ten controller's areas. In EAs with ethnic mixed composition, more than one enumerator was assigned, each of them belonging to a different ethnic community and able to use the corresponding language.

38.   The pilot census exercise pointed out that the training programme was too short and with insufficient practical exercises, and that the statistical regional offices were not enough involved in the training programme and in fieldwork operations. The distribution and return of materials was generally functioning, but it was perceived that the implementation of a tool to manage and to keep track of the census materials for the field activities, would increase the overall quality of the data collected.

VI.   CONCLUSION AND FUTURE ACTIVITIES

39.   The results of the analysis of the pilot census showed that mapping activities, IT applications and logistics played an essential role in limiting underestimation and overestimation errors, and that some improvements can still be implemented in the preparatory activities.

40.   In relation to mapping, the Cartography and GIS Unit of SOK is planning map updating activities together with a revision of the estimate number of households by EA, based mainly on dwelling and household listing through fieldwork investigation and the use of high-resolution satellite images for urban areas. A list of street names provided by local municipalities is also being entered into the GIS database to better organize fieldwork operations, especially in complex urban areas.

41.   Regarding IT, it is planned to carry out the double entry of a sample of questionnaires used in the pilot census. Indeed, the differences between the recorded data and the values on paper questionnaires have demonstrated the need to perform a double entry operation. It is expected that this activity would drastically reduce the amount of time required to perform batch

editing and final checks, and could allow for a better analysis of the quality of the data intake procedure.

42.   With reference to logistics, the implementation of an IT tool to organize and manage materials and staff during fieldwork operations, is seen as an effective instrument for monitoring the overall data collection.

43.   The Statistical Office of Kosovo, in collaboration with experts from the EU funded technical assistance project, implemented by EAR, is investigating the feasibility to develop an integrated and interactive Census Management Tool that would help to monitor at the same time fieldwork operations and data intake process in each phase. This tool would be integrated to the main data processing application. This solution, even though more complex, could be a more flexible and effective approach for planning census logistics and monitoring census coverage.

44.   The system would allow to associate census materials to enumerators, controllers and supervisors, and at the same time to their assigned EAs, with the possibility to retrieve easily the information.

45.   From the data processing application side, it would be possible to monitor each phase at the EA level (starting from the reception of the questionnaires up to the editing of the data) obtaining management reports such as productivity and phase completion reports. In this way, it would be possible to request at any time a report that contains information about the status of each processing phase.

46.   As an additional benefit, this IT module would provide a mechanism to validate the geographic identification of the entered data, to avoid duplications or omissions of EAs, and to identify the EAs where the RS1 forms and amounts of filled questionnaires are not consistent with the expected numbers. It would also allow the aggregation of census data at different geographic levels to easily produce census preliminary results.

*****