



Economic and Social Council

Distr.: General
29 June 2010

Original: English

Economic Commission for Europe

Conference of European Statisticians

Group of Experts on Population and Housing Censuses

Thirteenth Meeting

Geneva, 7-9 July 2010

Item 3 of the provisional agenda

Plans for census quality evaluation

Using administrative sources in the National Census 2011 Experiences after pilot census

Note by the Central Statistical Office, Poland

Summary

In Poland the 2011 Population and Housing Census will be based on a mixed mode which will include a wide use of information obtained from registers and administrative systems, as well as data collected in full-scale and sample surveys. This paper describes the various stages through which administrative data from approximately 28 registers and administrative systems will be used in the 2011 census, as well as the sample survey that will be carried out on about 20-25% of households, to collect information on specific topics. Information is also provided on the enumeration of institutional households as well as the homeless persons.

1. In the practice of Polish public-statistics, data from administrative systems have been used to a limited extent so far. Administrative data constitute approx. 20% of source data for statistical surveys. However, a strategy adopted in the National Population and Housing Census involved the widest possible use of public administration IT systems as sources of data for the Census. In Poland the Population and Housing Census in 2011 was based on a mixed implementation model of censuses through wide use of information obtained from registers and administrative systems, as well as data collected in full-scale and sample surveys in the following stages:

- (a) Collecting data from IT systems;
- (b) Internet self-enumeration, consisting of acceptance or correction of data obtained from IT systems within a certain time period;
- (c) Computer-assisted telephone interview (CATI) carried out by a statistical interviewer;
- (d) Interview, registered on a portable electronic device (Hand Held), carried out by an enumerator (CAPI).

2. About 300 available administrative registers have been examined, whereas approx. 28 registers and administrative systems are intended for use in the National Census in 2011.

3. Information obtained from registers is subject to the appropriate mechanisms and procedures used in obtaining data from registers.

4. The first stage of work included:

- (a) Work related to transforming administrative data sets into statistical data sets,
- (b) The comparison of values included in administrative data sets with the corresponding values from statistical surveys;
- (c) Developing the methodology of the quality assessment of public administration IT systems (criteria of assessment, indices, measures);
- (d) Work on data sets from the public administration IT systems selected from the whole country; transforming administrative data sets into statistical data sets;
- (e) Quality assessment of public administration IT systems with regard to using small area statistics in the National Census 2011;
- (f) Mathematical-statistical modelling of data for the census – statistics based on administrative data and surveys related to one another and forming one system.

5. The next stage concerned work carried out on unit data sets and included normalisation procedures as well as checking and correcting data sets from administrative systems:

- (a) Integrity check;
- (b) Formal and accounting check as well as logical check of sets;
- (c) Recurring/duplicated records check;
- (d) Missing data in records check;
- (e) Logical coherence of data check;
- (f) Development of assumptions for the substantive check of sets;
- (e) Quality assessment of data (development of quality indices for data sets).

6. The analyses show that almost all data may be available literally “right away”. Some registers, the so-called central registers, may be used for statistical purposes after a few weeks from the end of the reporting period; e.g. population register, after performing appropriate cleaning actions and transforming administrative register into statistical register. The longer availability period of variables involves so-called dispersed systems, which require the collection of partial systems from particular administrators and integrating them into one set. Nevertheless, some registers, for instance the tax register, may be used by a statistician only after a lapse of a year, due to the nature of this set, as well as the time synchronisation of the data. Some information obtained from registers may be used straightforwardly as census data.

7. Procedures for census variables calculation from administrative systems, which indicate the method of obtaining these values, have been developed:

(a) Direct transfer of values from one or several systems (selection of “the best” value from the list of available variables);

(b) Combining the variable value from the values of several variables from several administrative systems.

8. Nevertheless, in order to obtain the maximum statistical data, appropriate procedures and logical rules must be developed and applied, in a way which will make it possible to calculate the most probable values of statistical variables.

9. Information from administrative registers will be integrated into the operating micro database as a separate stratum of data. Data integration will consist of combining strata according to identifiers or key attributes. The final data will constitute the basis for preparing statistical information.

10. Regular updating of the micro database with data from registers and administrative systems will allow its use for current statistical surveys, as well as for preparing specialist statistical frames in order to, for instance, conduct social surveys.

11. The following fields of using data from registers and administrative systems are planned:

(a) Direct source of data for surveys;

(b) Data source for preparing lists of objects of surveys, and frames;

(c) Data source for completion of missing data – imputation;

(d) Data source for data estimation;

(e) Data source for making comparisons with data obtained in statistical surveys carried out using statistical forms.

12. As information technology is rapidly developing in Poland, one may expect that the next censuses in Poland will be based on these administrative systems.

13. Moreover, in the National Census 2011, a sample survey will be carried out on a sample of about 20-25% of households, which will concern information related to subjects, the obtaining of which in other ways is impossible or highly biased.

14. The form for a sample surveys will include questions concerning first of all:

(a) In-depth demographic characteristics;

(b) The socio-economic characteristics of the population;

(c) The situation of persons on the labour market (an in-depth survey of current and permanent activity);

- (d) Disability;
- (e) The fertility rate of women;
- (f) In-depth characteristics of houses and buildings (including, for instance, improvement in the standard of houses by extensive renovation, modifying buildings to the needs of disabled persons, etc.);
- (g) Surveys concerning the emigration of Poles, including emigration for work.

15. These subjects concern more difficult issues and require great knowledge of the surveyed issue; that is why surveys are conducted by the technique of personal survey between enumerator and particular persons from sampled dwellings.

Other subjects examined in the census

A. Institutional households

16. The group of persons which has to be taken into consideration in a census survey are residents of institutional households. These households may be divided into two categories:

(a) Permanently-inhabited dwellings (for instance, social welfare homes for chronically-ill persons, homes for pensioners and elderly persons, and orders and monasteries);

(b) Temporarily-inhabited dwellings (employee hostels, student dormitories, sanatoria, tourist establishments, etc.).

17. In the 2010 trial census a specially-prepared Internet application was used for the survey of institutional households. Administrators (owners) of buildings had to verify the list of persons living/staying in the building as at the date of census, taken from administrative registers, and add persons not included on the registers – it concerned, first of all, persons staying temporarily. The application included basic questions identifying a given person (PESEL number [*National Identification Number*], date of birth, sex, stay status).

18. The method applied for the first time was used, first for all, for checking the survey tool. The gathered data have been transferred – as a separate stratum – to the census database, where they will be further developed. By Identity Number (PESEL) it is possible to collect from registers the remaining data characterising a given person. After the analysis of the trial census, we will make additions to the prepared application and it may be used in full in the 2011 census.

B. Homeless persons

19. In the trial census there was an attempt to enumerate homeless persons, both those staying in institutional households, such as homeless shelters or common lodging-houses, as well as persons with no roof over their heads – staying on the street, in parks, on stations, etc.

20. Persons staying in organised dwellings were enumerated by persons administering these buildings, according to the general rules concerning institutional households (see above).

21. Persons “with no roof over their heads” were enumerated by interviewers at a specific moment (Friday evening and early morning on Saturday) in places previously

identified by employees of the Gmina Census Offices and institutions taking care of homeless persons. Groups formed in such a way have proved successful and it was possible to access places where the homeless were sleeping, and in many cases it was possible to communicate with them and obtain answers to some basic questions concerning sex and age (it was necessary to determine sex in order to enumerate homeless persons).

22. In the trial census, the survey of homeless persons was conducted in 4 gminas. In the 2011 census we intend to expand the application for surveying homeless persons and add new questions.

23. The preparation work being carried out at the moment, as well as the results obtained in the trial census, indicate that Polish statistics will conduct the Population and Housing Census in 2011 at the time and within the scope consistent with EU requirements and according to national needs.
