



**Economic and Social
Council**

Distr.
GENERAL

ECE/CES/GE.41/2009/24
14 October 2009

Original: ENGLISH

ECONOMIC COMMISSION FOR EUROPE

CONFERENCE OF EUROPEAN STATISTICIANS

Group of Experts on Population and Housing Censuses

Twelfth Meeting
Geneva, 28-30 October 2009
Item 6 of the provisional agenda

ONLINE DATA COLLECTION

Modern census in Poland

Note by the Central Statistical Office, Poland

I. INTRODUCTION

1. Technological progress, rapid development of IT, the improvements of administrative registries, and the development of telecommunication - all of these will be used in the realization of a task as huge as the census. Our aim is to reduce the costs and apply the least burden on the citizen's obligation to answer census questions. We are aware that censuses are held every 10 years; therefore they must be carefully prepared in order to avoid mistakes. Due to the cost and the scale, there is no possibility of repeating such a task. Our main goal was to develop a methodology which will allow us to take full advantage of all the technological benefits of the twenty first century.

2. Censuses are such a great financial and organizational effort, that even wealthy societies can execute them only once every few years. A cheaper and less demanding solution on both organizational and logistics has been sought for years. As a result of this search, we have reached for the administrative registries. As it turns out, the administrative registries often contain data covering the scope of the censuses. This particularly applies to the population and housing census.

GE.09-

3. Administrative registries have been thoroughly examined, methodologies have been developed for combining them with the collected statistical survey data, and mathematical models have been created to allow estimation and the imputation of data. As a result, with the opportunity of drastic cost reduction, the censuses may be carried out yearly. In Poland, the first, practical and widespread use of administrative sources will be implemented during the forthcoming censuses – the agricultural census in 2010 and the population and housing census in 2011.

4. The upcoming censuses are also a technological challenge, where for the first time, in addition to the extensive use of administrative sources, latest technologies of data acquisition will be used. We will entirely withdraw from the use of paper questionnaires and replace them with electronic questionnaires on mobile hand-held devices. Enumerator management will also be possible, with the aid of GIS tools (Geographic Information Systems). The hand-held device equipped with GPS will indicate the current Enumerator position and address points at which they must collect census data on an orthophotomap.

5. For the first time, self-enumeration over the Internet is planned. This particularly applies to people who are temporarily out of the country but are subject to a census. All citizens of Poland will be able to self-enumerate, which will spare them a census enumerator visit. The process of obtaining census data will also be supported by the computer assisted telephone interviews (CATI method). The whole process will be managed through regional centers supported by appropriate software-based GIS technologies.

6. The usage of administrative registries and advanced data acquisition technology will reduce the number of census enumerators working in the area by 90 per cent - from almost 200 thousand in the past census to about 20 thousand in the upcoming census. This allows the reduction of overall census costs by approximately 140 million PLN, or about 32 million euros.

II. LEGAL BASE

7. The forthcoming censuses will be the first since Poland's accession to the European Union. This requires strict adherence to the principles set forth by the respective legislation of the European Union. These acts are an EU-level regulation and are directly applicable in the legal systems of the Member States, and they ensure comparability of census results in terms of time and methodology. This results from the need to create a common agricultural policy, demographic comparisons and uniform analysis of the social and economic developments taking place in Europe. The UNECE recommendations are intended to standardize terms and methodologies of censuses in order to allow comparison of their results on an international level.

8. With regards to the national population and housing census current legislation in force are:

- (a) Regulation (EC) of the European Parliament and Council No 763/2008 dated 9 July 2008 for the population and housing censuses (Journal of Laws, EU No L 218 dated 13 August 2008.);
- (b) A National Census of Population and Housing 2011 bill, which is currently discussed on a legislative level and will soon be passed on to the parliament;

- (c) Public Statistics Regulation dated June 29, 1995 which regulates the legal basis and the census issues.

III. NEW IMPLEMENTATION MODEL

9. In the forthcoming censuses Poland, like most European countries, will implement a mixed model approach consisting of a combination of data from administrative sources with data acquired from statistical surveys. We believe this method to be the most effective and secure at the present development level of administrative sources and the degree of advancement of the methodological work enabling the use of administrative sources.

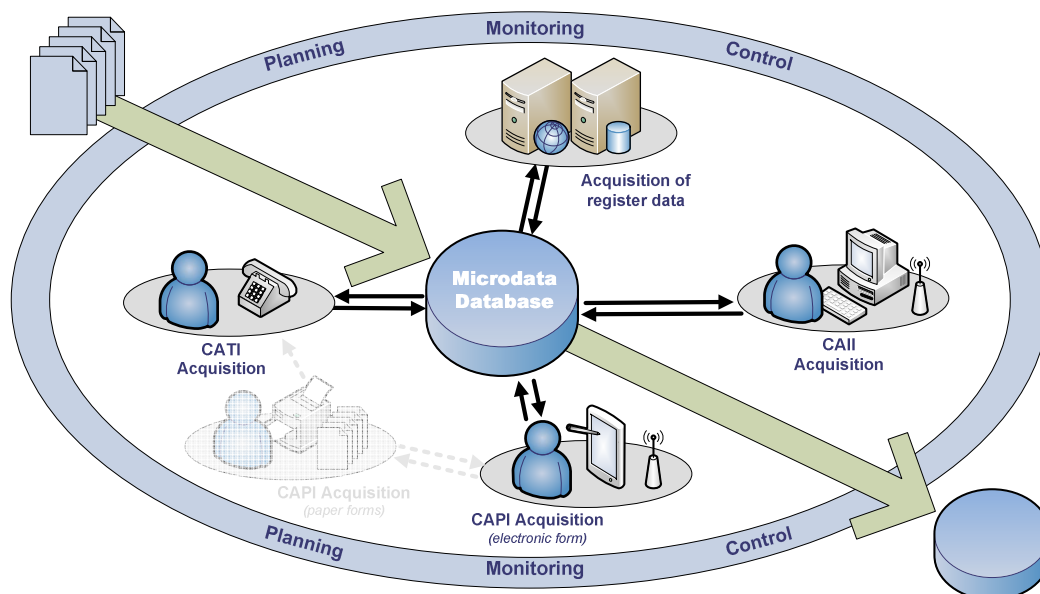
10. Poland, however - as one of the first countries in the world – will implement a fully innovative method involving the use of multiple modern census data collection techniques, at the same time. This especially applies to replacement of paper questionnaires with electronic questionnaires, which allow more efficient data collection. These include Internet technologies enabling self-enumeration over the Internet. The electronic questionnaire runs both online and off-line. A technology will be implemented in order to allow telephone interviews (CATI method). Also, census enumerators equipped with portable hand-held devices will use electronic questionnaires. Hand-held devices will enable census data collection, on-line transmission of data and the use of digital maps, which will eliminate the need for paper maps and situation sketches. The combination of digital maps and aerial photos with built-in GPS receivers is a revolutionary change in the development and census management before and during the census, it also enables multi-dimensional spatial analysis of the census results.

11. All of the above technologies will be used in the upcoming census, at the same time. This means, that the mixed method will use four channels for data acquisition:

- (a) administrative registries;
- (b) Internet – self-enumeration (CAII);
- (c) telephone interview (CATI);
- (d) census enumerator visit with the use of hand-held devices (CAPI).

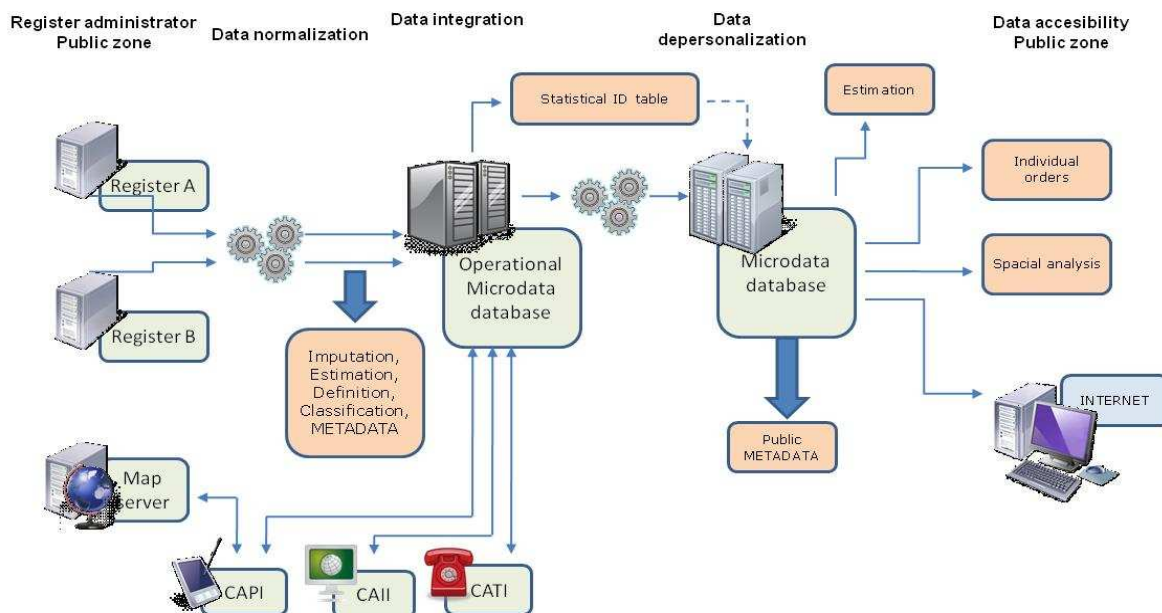
12. The last three channels will be supported by on-line electronic questionnaires. In exceptional cases, in emergency mode, the use of paper questionnaires may be applied. Data collected with these questionnaires will be imputed into the database through CATI or CAPI.

Picture 1. Primary data acquisition sources



IV. DATA PROCESSING

13. Data acquired from the four preliminary sources (described above) will be integrated in an Operational Microdata Database. This is one of the most important stages of census data processing, as it will decide upon the completeness of census data, its quality and the possibilities of data imputation and estimation. Data integration in the operational microdata database with regards to data from administrative registries is preceded by a very important stage of cleaning, standardization, de-duplication and normalization of data. Each information source is represented in the operating database, as a separate layer of data. The integration will rely on combining those layers basing on defined identifiers or key attributes. The Operational Microdata Database will allow advanced cleaning and repair of variables in the context of data available from all information sources. Ultimately, it will be possible to calculate, according to defined procedures and rules, the most probable values for individual census variables and transfer them to the target microdata database while depersonalizing individual data. The microdata database will provide the basis for creating products such as census output tables, spatial analysis and basic result aggregates. It will also be a source database for the provision of individual data for scientific research, while maintaining the principles of statistical confidentiality. Microdata database, together with its accompanying metadata system will contribute to the public information system of the official statistics.



Picture 2. Data processing chart

V. ON-LINE COLLECTED DATA SECURITY

14. In Poland, the issue of collected and stored data protection by public statistics is governed by the provisions of the Act of June 29, 1995 on public statistics

Article 10 Collected and stored in the statistical surveys of the official statistics individual data and personal information are confidential and are subject to special protection, this data can only be used for studies, summaries, statistical analysis and to create the services of official statistics frame for statistical surveys conducted by the departments to share or use personal data and personal information for purposes other than those indicated is prohibited (statistical confidentiality).

15. Official statistics collects and processes data from administrative registries in accordance with applicable Polish law, in particular:

- Personal Data Protection Act dated August 29, 1997,
- Regulation of the Ministry of Internal Affairs and Administration dated April 29, on personal data processing documentation and technical and organizational conditions which should correspond to the devices and systems used for personal data processing,
- the Act dated June 29, 1995 on public statistics.

16. Official statistics, by protecting personal information obtained from administrative sources, ensures that the data is subject to high-level security protection arisen from the above laws and regulations. Statistics protect personal data through:

- (a) application of technical and organizational measures to safeguard such data from being disclosed to any unauthorized persons;

- (b) providing access to the data only for people with appropriate authorization;
- (c) ensuring which personal data, when and by whom were introduced to the set and to whom they are transferred (e.g. computer system identifies the actions of its users);
- (d) keeping records of persons entitled to process the data (Article 39 of the Act);
- (e) application of safeguards with particular emphasis on the cryptographic protection for data sent over the public network.

17. In case of personal data being sent over the Internet, official statistics applies security measures at a high level through the use of cryptographic protection for the data as well as authentication data. This means that the software used to transfer to official statistics offices through ICT data channel - meets the regulatory requirements for personal data safety. In addition, official statistics keeps a record of personal data processing methods and documentation specifying the measures to protect such data. This documentation consists of the Security Policy and an on-line data collection computer system management instruction.

VI. THE USE OF ADMINISTRATIVE SOURCES

18. The following questionnaires of IT systems usage is being planned during the census:

- (a) direct source of research data,
- (b) source of information to create a list of entities covered by the census (address-housing survey, farm survey),
- (c) in addition, a source of information for:
 - (i) imputation,
 - (ii) data estimation,
 - (iii) comparison and determining the quality of the data.

19. The subject of data usage from administrative registries requires a thorough identification of information resources, which these registers carry. Currently, all sources and potentially useful variables are being analyzed. Necessary metadata from about 300 administrative registries has been gathered, which helped to identify the 30 most valuable registries. A distinct specification was created for each of these registries and all variables from those sources were submitted for evaluation. All variables were assessed for definition and classification compliance with the existing dictionaries in Polish and EU statistics. It resulted in priceless knowledge which the statistical service now possesses, as to the relevance and feasibility of integrating different public administration registries. The knowledge of the quality and usefulness of variables from different registries is reflected in the rules for combining data, its estimation and imputation in the created Operational Microdata Database.

20. In the forthcoming census in Poland, a use of approximately 28 sources is expected; from the government, local government and from external public administration data administrators such as property management, housing cooperatives, electric utility providers and telecom operators. All database administrators have displayed an understanding for the needs of statistics in relation to the censuses, and have provided all of the necessary information resources for the preparatory work. Currently, the statistical service is in possession or in the process of collecting

necessary data from administrative registries listed in the program of statistical surveys of official statistics for 2009.

VII. ADDRESS AS A UNIVERSAL ADMINISTRATIVE REGISTER DATA CONNECTOR

21. One of the main objectives of the official national registry of territorial division of the country - TERYT led by the President of the Central Statistical Office, is to provide unambiguous identification of objects with different levels of territorial detail, such as: voivodship (province), county, municipality, town, village, statistical district, census region, street, building and apartment. TERYT allows the collection of data for these spatial objects and provides the conditions for comparison and analysis, which is a very important factor in the implementation of the INSPIRE 2007/2/EC Directive of the European Parliament and the Council dated March 14, 2007 which establishes an Infrastructure for Spatial Information in the European Union (Journal of Laws, EU L 108, 25.04.2007, p. 1-13).

22. In order to ensure the technical conditions of modern information technology and standards required by the directive, the Central Statistical Office began work to ensure the identification of spatial objects on digital maps. The goal of this work is to obtain numerical boundaries for municipalities, districts and provinces. In addition, the user address ID, which determines individual buildings in the TERYT registry will be augmented with the x, y coordinates of the building.

23. The introduction of address points will allow a change of the existing system of spatial identification and will enable a transition from area allocation (census areas) to point allocation. This has crucial, revolutionary importance for geomatic usage in statistics. Changing the allocation will allow a more flexible aggregation of data collected in the national census, for any area.

24. It will also allow creation of microdata database with a spatial reference, which will enable geostatistical analysis of various effects such as:

- (a) demographics (e.g. the average distance between children and parents within the country, voivodship, county, municipality, village, township, or street blocks or any other described area, the average distance from work, school, hospital, etc.)
- (b) town and country planning (e.g. to assist in determining the borders of metropolitan regions, cities, development of spatial management plans)
- (c) agriculture and environment (e.g. crop structure survey, environment contamination),
- (d) the economy (e.g. study the effects of arduous road and industrial investments).

25. Allocating points with x, y coordinates will also allow independence from disruptive changes in the territorial division of the country, typically resulting in changes in census areas and the resulting arduous recounts. This will facilitate comparative analysis of time series irrespectively to changes in the division. An additional advantage is the aggregation of data in

both the administrative division of the NUTS system as well as the GRID divisions developed within the EU GEOSTAT project.



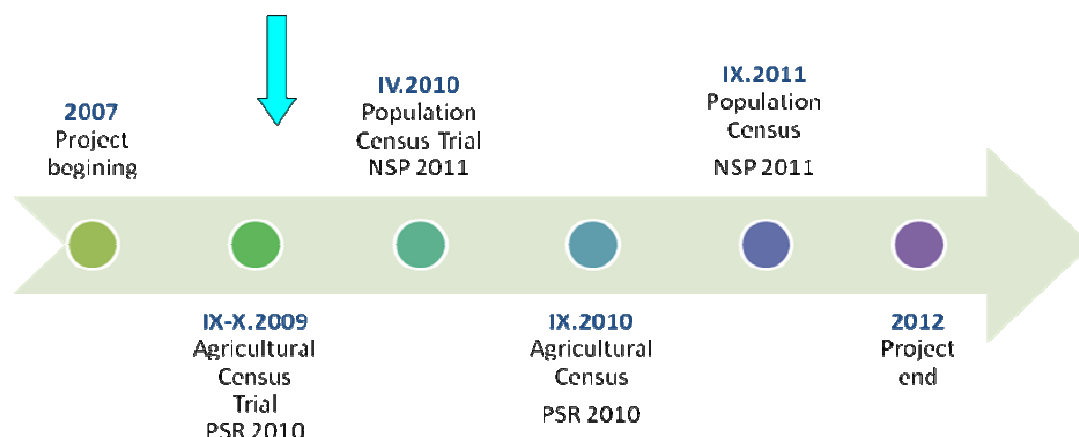
Picture 3. An example of a traditional map (paper) with the applied census area



Picture 4. An example of a modern ortophotomap with the applied address points and the boundary of a census area (the same area as above).

VIII. THE CURRENT STATE OF CENSUS PREPARATION

26. The census is preceded by a trial census. Its task is to comprehensively evaluate the technical, organizational and methodological solutions. Based on the experience from the trial census, we will be able to verify assumptions and modify those aspects of activities which may be necessary for censuses to be held in a smooth manner and with maximum efficiency. We are currently past the trial census which precedes the agricultural census. In the second quarter of 2010 we are planning a trial census preceding the population and housing census. Based on the current experience it can be concluded that correct assumptions were adopted and the identified problems are more of the psychological and mental matter than a technological or organizational matter.



IX. INSTEAD OF A CONCLUSION...

27. Methodological work related to the use of modern technology and the use of administrative registries for the censuses is very difficult and laborious. The experience of statisticians from Nordic countries says that simple relationship between the administrative sources and the variables, which one wants to obtain during the census, rarely exist. Most variables are to be deduced by combining data from multiple sources and processing them. In addition, the researchers know that during a traditional census we ask questions which we want to ask, but do we ever get a true answer (?). Whereas, the registries usually contain accurate data, but do they correspond to the questions that we would like to ask? This dilemma will always accompany statistical tests and is now particularly troubling Polish statistics. The answer may be provided solely by intensive methodological work aimed at combining administrative registries with statistical studies.

28. Usage of new technologies, a new approach to censuses, the abandonment of paper, data collection through on-line electronic channels and the wide use of data from administrative registries requires enormous organizational, intellectual and financial efforts, which are applied into preparatory work. The support of the government administration, local government, as well as the society is also crucial. All involved in the organization of the census are aware of the importance of this task. Also, all along we believe in absolute success of censuses carried out using the latest technology.