

**Economic and Social Council**Distr.: General
4 September 2013

Original: English

Economic Commission for Europe

Conference of European Statisticians

Group of Experts on Population and Housing Censuses**Fifteenth Meeting**

Geneva, 30 September – 3 October 2013

Item 3 of the provisional agenda

Census technology, innovation and outsourcing**New technologies used in 2010 Census Round – Polish case study¹****Note by the Central Statistical Office, Poland***Summary*

In the 2010 census round many countries used for the first time a new technologies allowing the collection of data from different sources and channels (i.e. administrative registers, Internet self-enumeration, direct interviews conducted by census enumerators using electronic questionnaires and hand-held devices, or by telephone). In Poland, simultaneous data collection from these channels was used - exclusively with the use of electronic questionnaires. Electronic questionnaires were initially completed with data from administrative registers. Thanks to such applied methods, paper questionnaires were completely eliminated, and were replaced by ICT solutions.

Furthermore, Poland has implemented GIS technology, adopted solutions ensuring the high security level for data processed, developed modern statistical data processing technologies, and established a comprehensive tele-information structure.

This paper presents the technologies used and the main organisational and methodological principles adopted at the outset of census work.

¹ Prepared by Janusz Dygaszewicz, Director of Central Census Bureau, Director of Programming and Coordination Surveys Division.

I. Introduction

1. The cyclical organisation of censuses of population and housing is an enterprise of great essence and significance, implemented in various economic and social conditions. However, regardless of the country's economic potential or the degree of social awareness, census implementation usually involves high costs and considerable citizen engagement. Therefore, the natural tendency is to attempt at reducing all expenses incurred for census operations, and at lowering the burden of respondents, while simultaneously maintaining high quality of the results.

2. It should be noted that the improved on modes of statistical production, not only national but also global, affect on improving their quality and effectiveness, and the continuously raised issue of minimising the burden of respondents is one of the organisational guidelines for the statistical survey conducted. It is influenced by rapidly developing IT and telecommunications technologies and constantly enhanced alternative sources of information, which enable us to reorganise the way censuses are carried out.

3. The organisation and course of the 2010 Census Round in Poland greatly contributed to the development of information society, which in fact was evidenced by the course of the census itself. Therefore, it has been proven that technological innovativeness in statistics is not only possible, but it also leads to an improved quality of the results, as well as allows to minimise the burden of respondents and reduce implementation costs.

II. Data collection channels in 2010 Census Round

4. The National Census of Population and Housing (NSP 2011) conducted in Poland in 2011 was designed and implemented with the application of a mixed model, i.e. employing data from administrative registers and data obtained from respondents, exclusively with the use of electronic questionnaires. As a result, paper questionnaires were eliminated altogether. It is a significant achievement on a global scale.

5. The starting point was the use of administrative and sources already existing within the State administration structures. In accordance with the National Census Act, all entities maintaining IT systems of public administration and non-administrative in the selected scope shall deliver data in the framework of census operations in the scope and time as specified therein.

6. In the framework of NSP 2011, a total of 27 registers maintained by 15 public administration institutions were used. Additionally, datasets from 3 non-administrative systems were applied.

7. The necessity to use data from administrative systems in Polish statistics resulted from:

(a) economic reasons – demand for effectiveness: minimisation of the costs of statistics production, including administrative burdens on respondents,

(b) the risk of an increased non-response in statistical surveys, including censuses,

(c) an intensive development of IT systems of public administration, based on advanced technologies.

8. Census implementation based on administrative and non-administrative systems has brought numerous benefits, including:

- (a) an effective use of administrative and non-administrative systems,
- (b) reduced census costs,
- (c) reduced social burdens connected with data transfer,
- (d) an improvement in data safety,
- (e) a guarantee of surveys harmonisation,
- (f) the availability information from future annual census based on registers,
- (g) the availability of data from administrative registers for any level of territorial disaggregation,
- (h) the possibility to identify double entry errors (overcounting),
- (I) the creation of a micro-database supporting indirect estimation – modelling at the unit level,
- (g) an improvement in estimation for small areas,
- (k) an improvement in the coherence and reliability of statistical data.

9. Data from administrative systems was used in the census:

- (a) as a direct source of census data (personalisation of questionnaires).
- (b) to create:
 - compilations of buildings, dwellings and persons,
 - an address-residence register,
 - a sampling frame.

10. To enable the administrators to transfer data from dispersed systems via tele-transmission, the Central Statistical Office of Poland (the CSO) constructed an electronic platform for data collection and processing, together with a net-based application for a direct data transfer via electronic means in a secure connection. These solutions were also applied when collecting from over 2500 local governments.

11. Apart from the use of IT systems of public administration, various data collection methods were applied, based on functioning of three channels simultaneously (known under the common name of CAxI):

(a) CAII/CAWI (Computer Assisted Internet Interview/Computer-Assisted Web Interview) – an online self-administered questionnaire, which entails checking by the respondent data obtained from administrative sources, within a specified time frame, and, if need be, correcting the same and providing missing information.

(b) CATI (Computer Assisted Telephone Interview) – a computer assisted telephone interview, conducted by a statistical interviewer.

(c) CAPI (Computer Assisted Personal Interview) – an interview conducted by a census enumerator, registered on a hand-held device.

12. All three channels were based exclusively on an adaptive electronic questionnaire, ensuring high quality of data at the collection stage. The electronic questionnaire was adjusted and implemented in accordance with the technology assisting particular modes of obtaining data based on CAxI. An appropriate questionnaire application (available at a

mobile terminal or Internet browser) verified if the questionnaire had been filled in accurately, among other things, through logical and accounting control.

III. Census architecture - The IT Census System

13. To enable the optimal application of advanced IT and telecommunications technologies in censuses, an appropriate census architecture had to be constructed. For the purposes of census design and conduction, the Central Statistical Office of Poland implemented the IT Census System (ISS). The system, consisting of more than 10 components implemented by different contractors, provided IT assistance for all operations within census. The ISS integrated various technologies (from applications installed on mobile terminals, through applications managing and assisting in telephone interviews, to specialist bases, data warehouses and analytical and reporting tools).

14. The ISS employed various solutions ensuring a high level of security for processed census data. Certain organisational means were implemented, obliging census participants to observe statistical confidentiality and to guarantee the protection of personal data.

15. As part of the work involving the processing of census data, including data from administrative registers, numerous rules and theoretical solutions were established in the Central Statistical Office, which were later implemented in practice.

16. The theoretical solutions and rules included:

(a) the preparation of normalisation, control, and correction rules for datasets from administrative systems, including data imputation (in administrative datasets),

(b) the preparation of rules for the synchronisation of data from administrative systems – the harmonisation of base periods – tables of transition from the state in which data from administrative systems were acquired to the desired state,

(c) the preparation of rules for supplementing missing census data – imputation and calibration,

(d) the preparation of rules for linking data from various administrative systems – the methods of precise/clear linking,

(e) the determination of the values of variables included in censuses (data-source rules),

(f) the preparation of the rules for calculating the values of census variables,

(g) the preparation of rules for creating derived objects – creating new objects (households, families),

(h) the preparation of a data-estimation model/method using data from administrative systems and statistical surveys,

(I) the preparation of data anonymisation rules.

17. A mobile application implemented on a mobile terminal (CAPI method) supported census enumerators in the field operation. The central dispatcher (responsible for the monitoring and coordination of NSP 2011 at the central level) and a voivodship dispatcher (a voivodship coordinator of the data collection process in the census, directly responsible for the census monitoring and coordination via the CAPI method) received assistance in the form of a dispatching application.

18. A module of a dispatching application was used by gmina leaders, who constantly support works of census enumerators working in their gmina. They were equipped with a

mobile terminal facilitating communication with the census enumerator; they functioned as the first-level support for the enumerator. The proper functioning of the applications was overseen by the administrators of the census systems, who worked in every statistical office as well as the Central Statistical Office.

19. The Stat Call Center system was used to support the CATI-method census. It consisted of functional components supporting the tasks that were given to interviewers:

(a) an application supporting the use and management of the incoming traffic (the Interaction Attendant application), i.e. a helpline that was used by respondents,

(b) an application supporting the use and management of the outgoing traffic (the Interaction Scripter.NET application), i.e. establishing connections as part of a CATI-based census,

(c) an application enabling the constant monitoring of the interviewers' work, providing access to reports and previews of the interviewers' work,

(d) an electronic form used to conduct interviews with respondents,

(e) an application allowing interviewers to arrange enumerators' meetings.

20. The CATI census was conducted by interviewers using the so-called SIP (Session Initiation Protocol) Stations equipped with headphones and microphones.

21. Pursuant to the National Census Act of 2011, the Operational Microdata Base (OMB) was developed, prepared, and implemented at the Central Statistical Office. The created System included hardware-system-tool infrastructure (computer hardware, system software, tool software) and application software (computer programs that are the result of programming work),

22. This base enabled the inclusion of data transmitted in electronic form through four informational channels by entities obliged to do so by the Act, i.e. entities maintaining administrative registers, persons covered by the census (via the Internet, a telephone interview run by a statistical interviewer or a direct interview run by a census enumerator) and to conduct further data processing. In the OMB there took place processes connected with the control, correction, and linking of data, up to their complete cleansing. Next, depersonalised data were transferred to the Analytical Microdata Base (AMB).

23. The Metainformation Subsystem gathered indispensable metainformation describing data and census processes, including the processes indispensable to drawing up quality reports. The task of the Metainformation Subsystem was to ensure the coherent definition of statistical objects for the OMB and AMB. The Metainformation Subsystem was also used to store depersonalised operational metadata of the OMB and AMB systems. This Subsystem constitutes the Central Metadata Repository (CMR).

24. The role of the Analytical Microdata Base is to store depersonalised census data in their final form. In this dataset every type of statistical analyses is carried out to acquire results for publication, i.e. the census products. The AMB allows all the recipients of statistical information to quickly acquire data in the form of aggregates. The AMB system constitutes an analytical and reporting platform that currently enables the statistical preparation of the outcome data from the National Population and Housing Census 2011. The results of analyses in the form of documents, reports and breakdowns are shared with internal and external users.

25. The AMB also allows the calculation of aggregates available in the Geostatistics Portal as maps (cartograms and cartodiagrams).

IV. GIS Technology

26. In 2010 Census Round a combination of data coming from administrative sources and register containing spatial data was used for the first time. The application of digital maps and the GPS technologies brought a revolutionary change when it comes to the possibilities of planning and managing census operations, both prior to and during the census.

27. For that purpose, the data obtained from the State geodetic and cartographic resources, as well as ortophotos (processed aerial photographs), were employed. With the use of the materials obtained, both from geodetic and statistical resources, it was possible to develop sampling frames for censuses, comprising statistical address points and their spatial reference.

28. The digital maps based on the GIS technology were used during:

- (a) gmina update,
- (b) pre-census field check,
- (c) census survey.

29. Digital maps were an indispensable tool facilitating the work of census enumerators (when it comes to moving around the area, verifying the sampling frame, etc.), gmina leaders, and voivodship and central dispatchers who could verify on a map the progress of the census and, for example, the route or location of an enumerator, using a dispatching or GIS application, facilitating the work of a gmina leader.

30. The introduction of x, y coordinates and address points in statistical data enabled changing of the previous system of spatial identification and shifting from area assignment (census districts) to point assignment. It had a fundamental significance for the applications of geoinformatics in official statistics. The change of the assignment mode allowed for more flexible grouping of data collected in public statistics for even the smallest areas. It also facilitated the creation of a spatially-oriented micro database, enabling the conduction of geo-statistical analyses.

31. In accordance with the organisational principles adopted, pre-census round was made prior to census commencement. In pre-census, a census enumerator verified the existence of buildings and supplemented the register with missing address points. An additional aspect of the round involved examining the area where the census was to take place, and possibly resolving any ambiguities which had arisen during data revision on the gmina level.

32. Mobile terminals were equipped with the GIS application, which enabled revisions and showed on the map, among other things, the current location of the census enumerator (GPS) and address points assigned to him. Using the mobile application, the enumerator could change the location of an address point, delete an address point, or add an address point not included in the register, with the help of GPS device. During the pre-census stage, the enumerator was also responsible for controlling the entire areas of the census districts assigned to him. It was particularly important when the revision in the gmina district was performed only on the basis of the registers kept, and the census enumerator was the first and in many cases the only person directly involved in field work in the framework of the census operations. The GIS application was also actively applied during the census – to manage its course. It enabled the monitoring and control of the enumerator's work, as well as the tracing of his movement in the area (among other things, to ensure his safety).

V. The Geostatistics Portal

33. The Geostatistics Portal is a tool for interactive cartographic presentation and the publication of data acquired in censuses. It serves the following functions:

- (a) storing,
- (b) presenting,
- (c) sharing information for a broad group of recipients.

34. The Portal functions on two levels: for internal (official statistics) and external users, and the scope of presented data is defined through the appropriate roles and authorisations. Internal users have access to both unit and aggregated data, whereas external users only to aggregated data, published taking into account statistical confidentiality.

35. The interface of the Geostatistics Portal allows its users quick and easy access to resulting statistical information. Data are presented using such cartographical presentation methods as cartograms (choropleth map) and various cartodiagrams. It is also possible to set one's own parameters for the visualisation of a thematic area for a given cartogram. These include measure, aggregation level (territorial division unit), the number of intervals, etc. Aside from the possibility of using ready-made spatial analyses, in the Geostatistics Portal, internal users can draw up custom thematic maps based on a selected feature of the data model, using dynamic spatial analyses, i.e. linear or distance analyses, or object buffering.

VI. Census organisation support systems

36. To facilitate efficient communication with all the members of the census organisation, two systems for the support of information and material exchange were implemented – the Notification System and the Knowledge Base.

37. In the Notification System, several thematic sections, so-called projects, were created. Each project encompassed a selected thematic area. Each notification was assigned to a suitable project and person responsible for resolving the reported issue.

38. The up-to-date versions of documents, including instructions, training materials, and the operational schedule were gathered in the Knowledge Base.

39. To ensure effective management of the remote enumerator instruction process, the m-learning application was employed. It was installed on mobile terminals and enabled conducting of tests, quizzes, and supplementary training using the available training materials. The application had been designed in such a way as to match its interface with the mobile application used in enumerators' work.

VII. Census staff organisation

40. The organisation of the census frame had a hierarchical structure, corresponding to the territorial division of the country and the systematics of data collected in census. It consisted of the central level (the Central Census Office), voivodship level (Voivodship Census Offices) and gmina level (Gmina Census Offices). Some part of the census structure was permanent and some of the census frame (voivodship and gmina levels) was established only for the time of preparation and implementation of NSP 2011 in the field.

VIII. Summary

41. The census in Poland turned out to be innovative project not only countrywide but also worldwide on grounds of the following facts and figures:

(a) simultaneous data collection, without paper, from four different channels (i.e. administrative registers, Internet self-enumeration (CAII), direct interviews conducted by census enumerators, using electronic questionnaires (CAPI), and telephone interviews conducted by statistical interviewers (CATI)) was used and implemented on such a large scale for the first time in Europe,

(b) data from 27 administrative registers and 3 non-administrative systems were effectively integrated,

(c) paper questionnaires were completely eliminated, and were replaced by ICT solutions,

(d) the use of GIS technology helped to conduct the census preparatory work and an ongoing census process monitoring and give possibility to compile and present census results based on multi-dimensional spatial analyses,

(e) IT Census System comprised a number of solutions ensuring the high level of security of the processed data,

the modern statistical data processing technologies have been developed – they will have a considerable influence on the methodology of future statistical surveys,

(f) A comprehensive tele-information structure was established, considerably increasing the automation of statistical data processing.

42. A comprehensive analysis of census conduction, accounting for all its participants, thus also the members of the field census frame, allows one to draw certain conclusions and to assess the possibility for their further implementation. The new technology applied in the census has proven that it can also be implemented in questionnaire-based surveys. It is cheaper, employs up-to-date control mechanisms, enhancing the quality of the material collected, and, in consequence, reduces the burden of respondents.

43. The time frame proved right, since the deadlines set for the data collection stage (specified in the Act on NSP) did not have to be extended, and neither did the census budget. The detailed schedule for the implementation of NSP 2011 was regularly updated. The framework schedule comprised over 250 items, and the detailed schedules for tasks included therein (for example the preparation and procedure of a control census, support of the census systems) were kept in separate files. The schedule comprised a total of several thousand tasks.

44. It should be noted that the effectiveness of census implementation was owed both to the methodological as well as organisational and logistic preparations.

45. There is another round of censuses ahead. Thanks to current experiences, in the next round of censuses Poland is going to use newer technology making it more effective and more inventive, which is still delivered and modified by the IT world.

46. Until then, census implementation methods should be developed and implemented, treating the experience gathered in 2011 as the starting point. Considerable efforts need to be expended with a view to developing a new census strategy, so as to guarantee progressive solutions. Attempts should be made at:

(a) reducing census costs,

(b) using administrative sources in an effective way,

- (c) reducing social burdens connected with data transfer,
 - (d) improving the safety of transferred data,
 - (e) improving the coherence and reliability of statistical data.
-