

## **Modern approach to censuses in the case of Poland – advantages and constraints**

Grażyna Marciniak  
Central Statistical Office, Warsaw, POLAND  
e-mail: [g.marciniak@stat.gov.pl](mailto:g.marciniak@stat.gov.pl)

### **Abstract**

The Agricultural Census 2010, as well as the National Census of Population and Housing 2011 were carried out in Poland in the framework of common project aiming at implementation of a mix-mode approach. The purpose of the paper is to contribute to discussions on the future of censuses. It contains a presentation of Polish experience on adaptation of the method based on data from administrative records combined with sample survey, as well as on the use of modern technologies for data collection, monitoring, data processing and elaboration of the results. Main features of the mentioned project, highlighting its innovative character, are discussed. The essence of the adopted approach was to use registers of public administration and some other information systems existing in Poland to the highest possible extent. Those were used not only for the purposes of construction of censuses' frames, but also – as it was in population and housing census' case – as the source of principal information that was completed by a survey conducted on a random sample of 20% of dwellings in the country. Different techniques of data gathering were applied, including acquisition of administrative records transformed to statistical data sets and various computer assisted modes of interview (CAII, CATI and CAPI). Paper forms have been entirely eliminated. The census performance in the modern framework required developing an appropriate IT architecture, preparing an organizational structure adapted to the new approach, preparing staff to the completion of census tasks in new technological and organizational circumstances, developing a coordination and quality management system, as well as the system for integration of data drawn from various sources and suitable processing technology.

Key words: data integration, management system, mix-mode approach, use of administrative records.

### **1. Introduction**

The 2010 National Agricultural Census and the 2011 National Population and Housing Census in Poland were carried out in line with the applicable EU and national laws governing their subject matters, scope, form, mode and limits of statistical obligations. Secondary legislation for Polish acts was particularly useful in providing broad access to administrative and non-administrative data sources. As a result, 28 registers and information systems, maintained by 16 different administrators, were applied for the purposes of the censuses. Bringing these two tasks under one project helped develop a coherent system for organising the censuses, manage the field staff and its work, monitor the whole process, and also equip the enumerators with modern data-recording technologies. Above all, however, it resulted in building IT infrastructure which was used in both censuses.

The Population and Housing Census, carried out using data from administrative and non-administrative sources, was combined with a survey on a random sample involving over 2.7 million dwellings, an unprecedented size of sample in Poland's statistical practice. Because the Population and Housing census employed a wide variety of methods and techniques, further on I will focus mainly on this part of the project, though as it was mentioned, both censuses were performed using organisational and technological infrastructure developed for the entire project to be managed in line with PRINCE2 method.

## **2. The idea of the new approach to Polish censuses**

So far, the statistical practice adhered to by Poland in performing population and housing censuses was based on traditional methods, involving a great number of enumerators visiting all dwellings and other inhabited premises and making records of information provided by their inhabitants (or by the administrators of collective housing facilities) on paper forms. Data from the forms was then recorded (in the 2002 census - using the OCR technology) and electronically processed. Due to the substantial costs generated by the traditional census approach, with substantial labour consumption and respondent burden that it entailed, it was decided that the next population census would employ different methodology and organisation based on mixed model.

The underlying idea behind the new approach to census was to make all possible use of central and dispersed administrative systems and non-administrative information systems available in Poland to perform a full-scope survey. The thematic scope of the census was complemented with extra statistical surveys, and then the information acquired from different sources was integrated in a single data base.

Digital maps were worked out combining information from state geodetic and cartographic resources and materials held by the statistical authorities. Once created, the digital maps made it possible to develop a census frame (an electronic list of addresses and housing facilities) in which the address identifiers of the buildings recorded in the National Register of the Official Territorial Division of the Country (TERYT) were integrated with the geodetic x, y coordinates to help identify precise address points. Installed in the hand-held device application equipped with GPS, these maps served as an auxiliary tool for the enumerators and an essential device for the monitoring and managing of their field work.

The census was performed as a full-scope survey covering basic data, mostly acquired from administrative systems. Those data integrated in personalised questionnaires were made available to the population via a special, on-line application in a protected environment for verification and supplementation. The full thematic scope of the census was covered by the sample survey, with its electronic form also being made available on-line to inhabitants of the sampled dwellings. The Population and Housing Census was carried out using several channels of data gathering – in addition to data sets acquired from the different information systems, a full-scope survey and a sample survey were performed on-line (CAII), followed by a sample survey carried out by enumerators, who used hand-held devices for recording the information obtained (CAPI); furthermore, special call centres were launched for the census purposes with statistical interviewers gathering data via phone (CATI), to supplement the available data. In the collective households and places where homeless persons stayed, the census was performed by way of a complete survey involving enumerators. The process of managing and coordinating census data collection and monitoring census performance was essentially facilitated by a special application for management of census completeness.

Underlying these organisational and methodological approaches was the need to rationalise the costs and reduce the respondent burden. At the same time, the new approach to censuses' performance created the opportunity to modernise the interviewing methods, the monitoring of the interviewers'/enumerators' work, and the streamlining of the survey process management.

## **3. The main tasks in the framework of modern census**

In order to perform the censuses using the modern approach, it was essential to appropriately plan all the actions and the required financial and fixed resources, and also to develop and enact legislation imposing the statutory census obligations on the population and institutions and providing the official statistics with an access to administrative and non-administrative sources of information. With both censuses

performed under one project, it was easier to properly prepare and efficiently complete all the stages of census performance.

Set up for the project's purposes, the organisational framework played an important role in central, but also regional and local, management. The Central Census Office was established, with the Census Management Centre, and also 16 Voivodship Census Offices, with Voivodship Management Centres and Voivodship Call Centres. Additionally, working groups and task forces were formed to address specific methodological and organisational problems.

During the census preparation work, metadata was collected from nearly 300 administrative registers and other non-administrative information sources. These registers were evaluated in terms of the potential they had for providing data on the population, buildings and dwellings. The evaluation was performed in line with the definitions and classifications recommended by the United Nations Statistics Division and Eurostat. The analysis resulted in choosing 28 information systems, to be used later in the census as a direct source of survey data or a resource based on which a list of entities being subject to the census could be compiled. The registers and other information systems chosen according to their usefulness for statistical purposes were also meant to provide data for imputation, data estimation, comparisons and verifications, and also for results' quality evaluation.

In order for the census to be carried out in line with the new organisational, methodological and technological approach, special IT infrastructure had to be developed. To ensure the smooth management of the modernised census tasks and to monitor its performance, an electronic platform, with systems, sub-systems and applications was set up to support the processes of collecting, processing and disseminating census data. A proper infrastructure was ensured for the census purposes, including but not limited to, the Operational Microdata Base and the Analytical Microdata Base, together with a sub-system for metadata, a support system for various communication channels, and also application designed for management of census completeness.

Derived from 28 systems, the sets of input data were assembled in the Operational Microdata Base, and then integrated and converted into statistical data by using procedures of cleaning-up, standardisation, validation, deduplication and selection of appropriate source for given attribute. Data from administrative systems which conformed to the quality requirements was used to prepare and update the address and dwelling list, and subsequently to develop the sampling frame of addresses and dwellings for the purposes of the sample survey. It also served as direct census data.

Based on the combination of inputs, the *master record*<sup>1</sup> was defined, and a census subject list was compiled, to be used in a complete survey; it was also used to generate prompts in electronic census forms to help respondents fill them in on their own and facilitate the process of self-enumerating.

The census completeness management application was the main tool supporting operational managing and coordinating the process of census data collection and monitoring the course of the census. Functionalities of the application ensured fully automated data exchange with the systems supporting individual CAxI channels, and also with the Operational Data Base. They also supported choosing census method for a given census entity and ensured the verification and confirmation of the questionnaires filled in through on-line self-enumeration, and also helped monitoring the course of the census (including the monitoring of the enumerators' route).

In order to perform a modernised census, it was essential to ensure the proper training and motivation of the staff employed for this purpose. Before the actual census, training was provided to all the persons involved in the process, and covered the census methodology, the implementation of ICT technologies, and also the way the

---

<sup>1</sup> It contained personalised data on the census entity, including its statistical feature values based on the information available from registers and administrative systems

census-support applications and systems worked. The training was delivered by statistics professionals and covered the census methodology and organisation, as well as by the representatives of the outsourced companies which had delivered individual components for the IT system. The latter training covered the use of the implemented IT solutions.

#### **4. Activities needed after census completion**

Gathered from different sources and using various technologies, data was assembled in the Operational Microdata Base and verified, checked in terms of completeness and deduplicated. Information obtained through the full-scope survey and the sample survey was integrated and incorporated in the so-called *golden record*<sup>2</sup>. The golden record is the result of choosing the questionnaire (for given census entity) which contained the most valuable information coming from given channel (CAPI, CATI or CAII). The selection was based on the completeness indicators calculated for each of 13 thematic modules covered the census scope. Because the questionnaires were completed not for all respondents, the sample survey results were generalised using the corrected weights and then also calibrated weights. The population composition by sex and age, as determined in the complete survey, was used as the reference population. The golden record, containing final and depersonalised census results, was transferred to the Analytical Microdata Base (AMB), where further processing, data analysis and dissemination took place. The Analytical Microdata Base was used for calculating derivative variables requiring more complex algorithms. The tools which the Analytical Microdata Base is equipped with enable the creation of microaggregates and aggregates, and the development of predefined tables to help the users employ the census results and conduct statistical analyses.

Moreover, a geostatistical web portal was launched to expand the electronic platform created for the census. The infrastructure in place is equipped with functions, enabling both ready-made and custom-made spatial analyses, with the possibility of disseminating the results and present them in graphical form. Data users will be able to compile their own maps based on a selected phenomenon (its attribute) and covering even the smallest areas, according to the needs (in compliance with statistical confidentiality). The geostatistical web portal allows for updating the address points and makes it possible to conduct analyses irrespective of the changes to the administrative division of the country.

#### **5. Advantages and constraints in modern census completion**

With the adoption of new approach, it was possible to facilitate the censuses' organisation, streamline the management of enumerators' work, and to monitor the run of the census processes off-site. The identification of the premises to be covered by the census was substantially facilitated by providing enumerators with electronic devices for recording respondent information with embedded digital maps and marked address points. With the implementation of new methodological and technological solutions, a much smaller number of enumerators were employed compared to the traditional approach. The use of electronic forms helped cut the costs which the previous censuses generated due to the printing of paper forms, and to reduce the labour consumption related to the recording of the forms. It is worth noting that the experience gained through the completion of mixed-mode census, together with the IT infrastructure that was set up for such organisation and methodology of a survey, pave the way for further modernisation in carrying out statistical surveys and the dissemination of their results.

It should be mentioned however that during the process of performing the census following adopted organizational and methodological assumptions a range of

---

<sup>2</sup> It contained personalised data about census units (entities) and a set of feature values gathered directly from the full-scope survey and the sample survey or calculated based on the gathered information

difficulties occurred. As far as the factors that made it hard to perform the census by new approach are concerned, it is important to recall that the Polish registers and information systems were ultimately not designed for statistical purposes. Each of the systems was designed to serve very specific needs related to administrative and/or operational tasks, and was only to a limited extent useful for statistics. It took a great deal of work to integrate data from different systems (including dispersed systems) and transform them into a census-usable form. Due to quality of information and its accessibility in accordance with the census scope, the collection of data received from registers for direct use in the census was rather limited.

What disappointed us was that a relatively small proportion of people participated in the on-line census, which, in addition to the administrative sources, was meant to be the second-most important channel of gathering information on population. Approximately 10% of respondents took part in the full-scope survey via Internet. Because of this a much greater importance had to be attached to processing the sample survey results which, at the same time, became the primary source of information for the majority of recommended census topics.

In the sample survey carried out using CAPI and CAII methods (the latter was used solely by 2% of respondents) the information on over 8 million people residing in 2774 thousand dwellings was collected. Register sourced data was used particularly to determine the size and structure of the population by sex and age, and for the purposes of imputation and estimation.

In consequence, despite shortening the time both of the census' performance in the field and of data registration, the full range of final results of the census was disseminated later than expected. Elaboration of the results required the use of multiple procedures, i.e. analysis, control and choosing the best survey questionnaire, as well as imputation in case of position non-responses. The involvement of academics was necessary. The academic staff developed a method and performed generalization of the survey results using calibrated weights. The precision indicators for each variable were determined and it was also assessed what degree of details was acceptable for the presentation of particular variables.

Because of its design the census could not provide data that enabled the direct publication of results at the lower level of territorial division of the country (below LAU1), and so did not entirely meet the expectations of the Polish public that the information be ensured even at the local level (LAU2). In this case too, knowledge and experience of academics were sought. Small area estimation will help bridge this gap, with the scope of data presented being limited to the information meeting the precision criteria.

## **6. Building on the census experience**

Developed for the census purposes, the IT infrastructure was later used to modernise the approach to conducting random-sample based social surveys and agricultural surveys. Major progress was made in the area of the methods employed to gather information, with the gradual removing from the use of paper forms. Thanks to employing electronic devices and applications designed for data recording and for the call centres in the census, computer-assisted interviews and phone interviews are being consistently implemented. Moreover, the address and dwellings register prepared for the purposes of the Population and Housing Census, and updated to include selected census results, became the basis of creating the sampling frame for social surveys, enabling more effective stratified sampling and sample allocation. This frame contains personal data and will be updated in a protected environment using data sourced from administrative and other designated information sources.

The organisation of the census, which made it necessary to integrate data from different data sources, implement new technologies, and to train the staff to work in a new technological and organisational environment, provided us with valuable experience. We made use of this experience to develop a concept of an integrated

model of realization of statistical surveys. This concept, which involves the generic-model-based<sup>3</sup> overhaul of the processes of collecting, processing, analysing and disseminating statistical surveys results, will undergo a trial in the area of agricultural surveys this year. Should this pilot programme bring positive results, the generic-model-based surveys realization system will be consistently introduced in the statistical practice.

We believe that with implementation of these modernisation solutions, the standardisation and integration of the processes related to conducting statistical surveys, their organisation and coordination and also the dissemination of their results, will be achievable, and so result in more effective production of statistical output and better satisfaction of the ever-growing requirements of the statistical data users..

## 7. Conclusions

The decision to carry out the censuses using the new approach, with the underlying objective to make the highest possible use of the nationally available information resources, resulted from the rational need to manage the budget in the most efficient way and to diminish the burden put on respondents. As well as this, the mixed-mode performance of the censuses using administrative sources and different communication channels, involving the setting up of an electronic platform for census data integration, processing and dissemination, proved a major step towards modernising the methodology of statistical surveys. What it also did was set the direction for future development in this area of statistics.

The census performed according to traditional methods is certainly a history. There is no doubt that efforts should be made to use the information resources of the administrative registers existing in a country, as well as of other data sources on population and dwellings. Therefore, taking in consideration the preparation works for the next census round, its important part will be to ensure the regular transfer of data from administrative and non-administrative records, as well as to evaluate the collected information and to undertake actions for improvement of its quality. Before making a decision as regards the application of the new approach, the mentioned information sources must be identified in detail and properly prepared, so that the sourced information meets the criteria of statistical data.

The electronic platform cooperating with administrative systems and ensuring the possibility of data transfer from various communication channels, as well as control and processing of data, should be developed early enough before the census in order to have all its components thoroughly checked. The system performance and its functionalities, including the process of data integration from different sources, should be tested in the pilot census phase. Also, while preparing the next census round, much attention should be attached to promotion and education activities, i.e. explaining to the public the advantages of the census performed according to the new approach, as well as implementation of information activities aiming at ensuring broader participation of the population in the census via Internet.

## References:

„Lessons learned from the Population and Housing Census in Poland. Notes by the Central Statistical Office of Poland”, Economic Commission for Europe, Conference of European Statisticians, Paris, 6-8 June 2012;

„Ludność. Stan i struktura demograficzno-społeczna (*Population. Size and demographic-social composition*), The Population and Housing Census 2011”, (preface, contents and general remarks in English), Central Statistical Office, Warsaw 2013.

[www.stat.gov.pl/e-census/index.html](http://www.stat.gov.pl/e-census/index.html)

---

<sup>3</sup> Generic Statistical Business Process Model