The Population and Housing Census 2011 – Polish case study

Introduction

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 (full scope survey covering base demographic variables) and data obtained from respondents (20% sample survey), exclusively with the use of electronic questionnaires. The Act on the 2011 national population and housing census stipulated that the information systems of public administration shall be used as widely as possible as both the sources of data for the census purposes (which as a consequence meant that information to be gathered during the census was mostly obtained from available administrative sources and then used to prepare and update an address and housing register followed by preparation of an address and housing frame for samples to be used in sampling survey) as well as direct source of census data. Data did not included in the information system of public administration or data ineligible in terms of the statistical data quality were collected from persons covered by the census. As a result, paper questionnaires were eliminated altogether. It is a significant achievement on a global scale. We claimed that this method was safer and more effective, taking into consideration the present level of development of administrative sources, their quality, and the degree of advancement of methodological work concerning the estimation and imputation of missing data in administrative sources.

The use of administrative sources

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.

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

- economic reasons – demand for effectiveness: minimisation of the costs of statistics production, including administrative burdens on respondents,
- the risk of an increased non-response in statistical surveys, including censuses,
- an intensive development of IT systems of public administration, based on advanced technologies.

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

- an effective use of administrative and non-administrative systems,
- reduced census costs,
- reduced social burdens connected with data transfer,
- an improvement in data safety,
- a guarantee of surveys harmonisation,

- the availability information from future annual census based on registers,
- the availability of data from administrative registers for any level of territorial disaggregation,
- the possibility to identify double entry errors (overcounting),
- the creation of a micro-database supporting indirect estimation – modelling at the unit level,
- an improvement in estimation for small areas,
- an improvement in the coherence and reliability of statistical data.

The issue of using data from administrative sources required an in-depth recognition of information resources which were found in these sources. An analysis of all the sources and variables potentially useful for the censuses were carried out. The necessary metadata on approximately 300 administrative registers were collected, of which the 30 most useful ones were selected. For each of these registers separate records were opened, and all variables from these sources were subjected to the utility analysis. The variables were evaluated with regard to their conformity, in terms of definitions and classification, with the dictionaries existing in Polish and EU statistics. Appropriate weights were determined both for the variables and administrative registers from which these variables came from, taking into consideration their utility and quality. The knowledge concerning the quality and utility of variables from different registers was a basis for the rules of merging data, and their estimation and imputation in the operational base of microdata created. The result of this work was invaluable knowledge concerning the utility and possibility of integrating different registers of public administration which the statistical service had at its disposal.

Finally, in the census in Poland we applied 28 sources from Government and Local-Government administration, and from administrators outside public administration such as real estate administrators, housing co-operatives, power distribution plants and telecommunication operators. All the administrators of databases approached the need for statistics related to censuses with understanding and provided access to their information resources for the purposes of the population and housing census in 2011.

Data from administrative systems was used in the census:

- as a direct source of census data (personalisation of questionnaires).

And to create:

  - compilations of buildings, dwellings and persons,
  - an address-residence register,
  - a sampling frame.

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.

The unit data obtained from registers were converted into statistical registers, simultaneously being subject to the process of cleaning, de-duplication and standardisation of data. The process was carried out in the DQS SAS environment. At the same time, metadata were collected on quality of

input data obtained from registers, the applied cleaning procedures and the final quality obtained after applying DQS procedures. Data from administrative sources converted to statistical data were used to derive the Master Record.

Master Record was the set of variables derived from the registers containing information that was introduced to the census forms in order to verify (confirm or update) by the respondents.

Quality evaluation of data from administrative sources

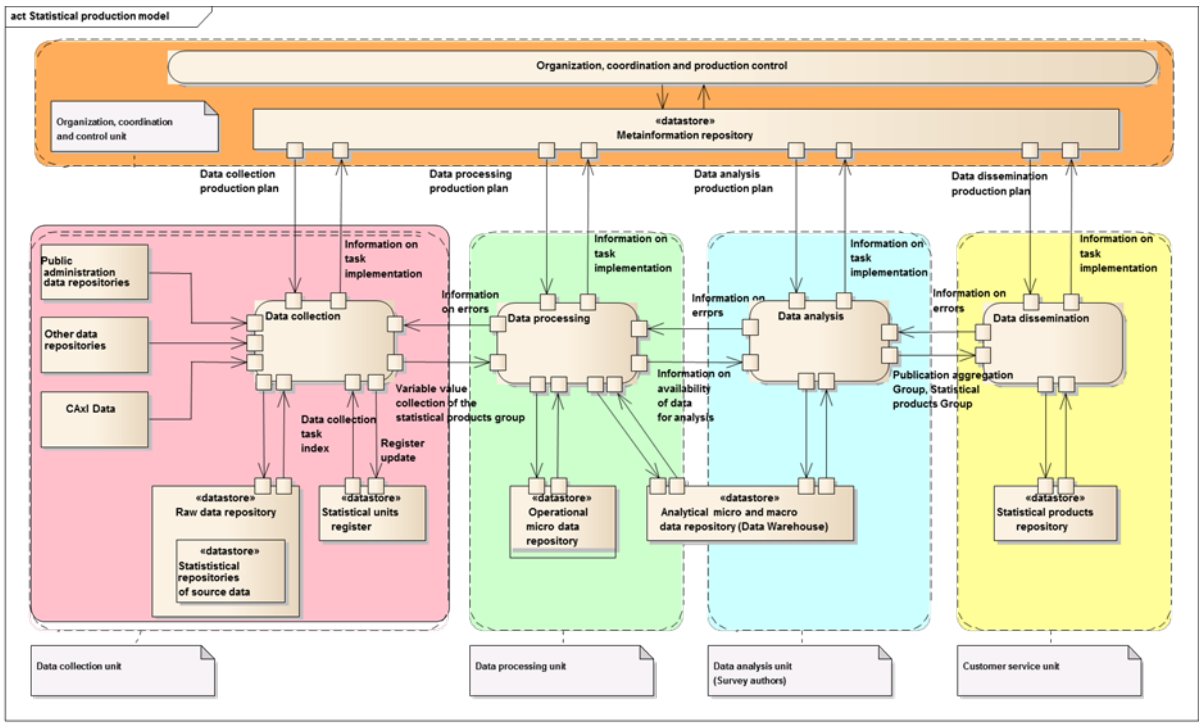
The quality of public administration systems was difficult to measure due to the complexity and multifaceted issues. Many factors make up the quality was unmeasurable. Difficulties in assessing the quality result of the inapplicability of the synthetic instrument, hence assessing the quality of systems based on multiple indicators.

Before assessment of the quality systems there are defined criteria and indicators to measure quality. For purpose of census there were three stages of quality assessment: evaluation of data sources - systems, data sets, statistical products - resulting data.

With regard to the sets from administrative data sources, quality assessment of raw data sets provided by the administrators and sets after a transformation, ie. after adjusting them for use in the census, was carrying out. Quality measurement was performed also in all stages, in all processes of development of administrative data and the data combined with data from other sources.

The scheme below shows a graphical representation of the processes in the census, including quality assessment.

"Metainformation repository" was created to collect methodological, technical and operational metainformation. It ensured process control of data processing as well as monitoring of the course of processes, including measurement and collection of metainformation concerning quality at all stages of the process of data development, i.e. on stages: data collection, data processing, data analysis, data dissemination.



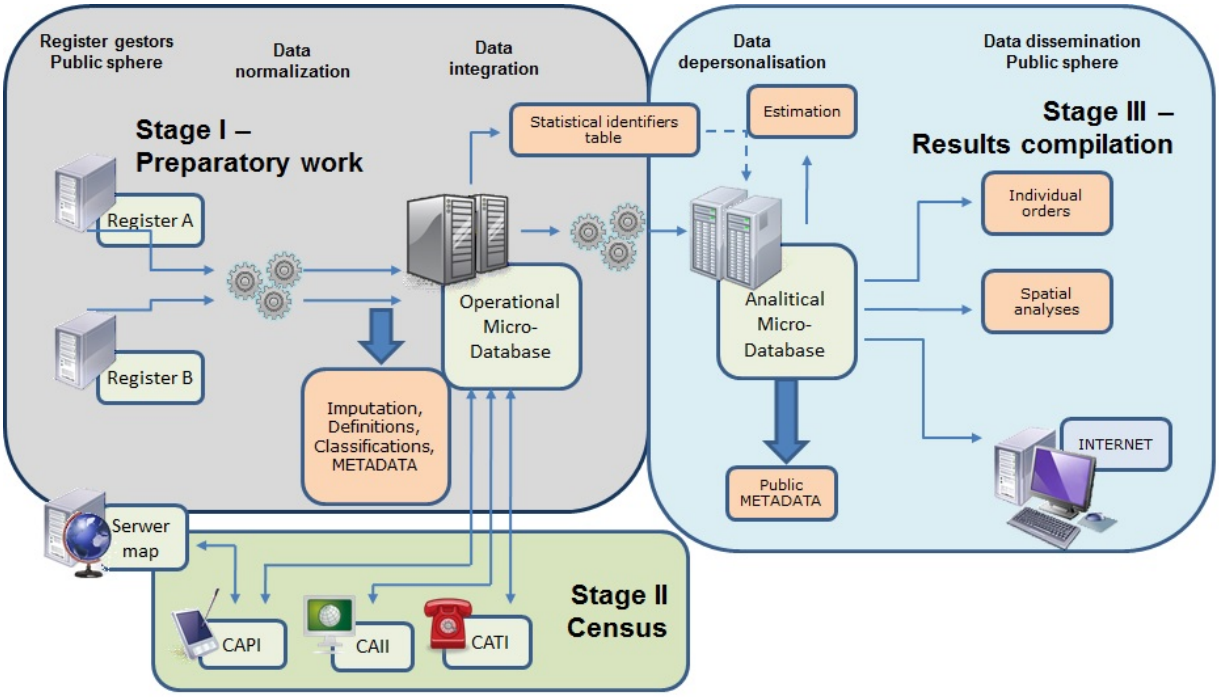
Others data acquisition methods used in the census

Poland was one of the first countries in the world which prepared a totally innovative method consisting of using several of the most modern techniques for collecting census data simultaneously. 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):

- 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.
- CATI (Computer Assisted Telephone Interview) – a computer assisted telephone interview, conducted by a statistical interviewer.
- CAPI (Computer Assisted Personal Interview) – an interview conducted by a census enumerator, registered on a hand-held device.

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.

Census architecture – The IT Census System



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).

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.

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

The solutions and rules included:

- the preparation of normalisation, control, and correction rules for datasets
- from administrative systems, including data imputation (in administrative datasets),
- 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,
- the preparation of rules for supplementing missing census data – imputation and calibration,
- the preparation of rules for linking data from various administrative systems
    - the methods of precise/clear linking,
- the determination of the values of variables included in censuses (data-source rules),
- the preparation of the rules for calculating the values of census variables,
- the preparation of rules for creating derived objects – creating new objects (households, families),
- the preparation of a data-estimation model/method using data
- from administrative systems and statistical surveys,
- the preparation of data anonymisation rules.

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).

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).

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).

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.

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

GIS Technology

For the first time in census history, GIS (geographic information systems) has been used in conducting and monitoring the survey.

With use of various reference materials and registers containing spatial information polish official statistics has created spatial data for statistical address points and borders of statistical division of the country. Digital maps used by census enumerators were an indispensable data sources (to navigate and verify dwelling locations in the field), by gmina leaders (for census monitoring within the gmina), and by voivodship and central supervisors (for census monitoring on voivodship or global level). Maps were used to monitor the census progress in a defined area or for specific enumerator (an on-demand location or daily route could be visualised on the map).

The Geostatistic Portal

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

- storing,
- presenting,
- sharing information for a broad group of recipients.

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.

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.

Summary

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

- 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,
- data from 28 administrative registers and 3 non-administrative systems were effectively integrated,
- paper questionnaires were completely eliminated, and were replaced by ICT solutions,
- 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,
- 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,
- A comprehensive tele-information structure was established, considerably increasing the automation of statistical data processing.

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.

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.

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

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.

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:

- reducing census costs,
- using administrative sources in an effective way,
- reducing social burdens connected with data transfer,
- improving the safety of transferred data,
- improving the coherence and reliability of statistical data.