Statistical Journal of the IAOS 36 (2020) 165–175
DOI 10.3233/SJI-190566
IOS Press

Transition from traditional census to combined and registers based census

Janusz Dygaszewicz
Statistics Poland, 00-925 Warsaw, Aleja Niepodległości 208, Poland
E-mail: j.dygaszewicz@stat.gov.pl

Abstract. This paper discusses some pre-conditions for the transition from traditional census to combined census and to register-based census. It also sheds light on the practical use of certain methodological experiences and quality aspects in the 2020 census round. The paper provides an overview of potential use of administrative registers in population and housing censuses and formulates basic recommendations for their use. It focuses on the case study of Poland, which carried out a population census using combined methods of linking administrative registers with ad hoc surveys.

Keywords: Combined census, register based census, GIS, data transformation, quality

1. Introduction and justification

Since the census of the population was implemented, the primary goal has always been to count the population and learn about its socio-demographic structure. Over time, the thematic scope of information collected from the public was enriched, and there is now an international set of data to be collected. This set is modified, if necessary, by a team of international experts participating in task forces and working groups. Although the reduction of the burden related to the obligation for the respondents to participate in the census used to be less important, it has been now changing from their perspective. The society has become more demanding and does not always accept an absolute requirement to submit to a census without any facility to reduce the time spent providing information and without being able to choose modern forms of participation in the census.

In countries with developed administrative registers, the benefits of using data in registers to replace data obtained directly from respondents were noticed. It was a gradual process, carried out over the years, starting from the partial elimination of some data by collecting them from registers and then filling in the remaining data with those acquired from direct interviews until the full census could be carried out exclusively based on registers. The necessary condition was absolute assurance of the reliability of administrative data, i.e. definitional compatibility of concepts in registers and statistics, validity of data at the moment of the census, and an exhaustive scope of data consistent with the scope of information collected in the census.

Under the register-based approach, there is no direct collection of data from the population, and the traditional enumeration is replaced by the use of administrative data from various registers (population register, building/address register, social security, etc.) through a matching process, normally making use of personal identification numbers (PINs), if available.

Countries which do not have registers that can be used for census purposes, or which are at such a stage of development that they do not guarantee the full replacement of census data, must look for other ways to reduce the burden on respondents, for example by reducing the information content of the forms and thus the number of questions. It should always be borne in mind that efforts to reduce the burden on respondents must not have a negative impact on the quality of the census.

The quality of administrative registers operating in Poland is still insufficient to be the sole base for the results of the register based census. However, it has been improving constantly. Therefore, it was already

decided in 2011 to partly use data from the registers and complete the missing data with information obtained from respondents. To be sure about the quality of the census results, all census questions were included in the census form, even those covered in administrative sources. The purpose was to show the respondents what personal information was contained in the registers and to oblige them to correct the data if they were out of date. The number of changes made by respondents was used to assess the quality of data in the registers.

After the 2011 census, the work continued to improve the quality of data in the registers in order to make the data more consistent with the requirements of official statistics. The work is carried out within the framework of statistics and in cooperation with the register administrators with the support of the Ministry of Digitization. Representatives of official statistics participating in the work of inter-ministerial working teams try to point out the very important role of registers and the need for their coherence. Moreover, in the amended act on public statistics, the role of the President of Statistics Poland and his influence on changes made in the public registers has been strengthened.

For the 2021 census, the data collection system will be maintained, but in a modified form. This will involve comparing data collected from registers with data collected from respondents. However, to comply with GDPR[1] regulations and the statistical confidentiality, data from registers will not be presented in the respondents' electronic form. As a result of the continuous process of improving data quality in administrative sources, it will be finally possible to eliminate questions that can be collected from registers in subsequent censuses, and then completely abandon the need to involve respondents.

Over the last two decades, some countries in UNECE region developed innovative methods to conduct the combined census by using administrative data linked with a full enumeration or sample survey for specific variables [7]. Usually, this approach is a basis for the transition from a traditional to a register-based census.

The register-based population census system is built around a set of basic registers that contain comprehensive data on the units that are to be described in the population and housing census. Some register-based census countries lack some of the census variables in the available registers and choose to support their census with data from already existing sample surveys. A common feature of register-based censuses is that no census questionnaires are used to collect information about the population. Therefore, with reference to censuses based on registers there is no need to use of paper or even enumerators, and they are generally much cheaper than combined censuses and are particularly cheaper than traditional censuses.

[1] General Data Protection Regulation.

2. Preconditions for using administrative registers in censuses

Before deciding whether to use administrative registers in censuses, National Statistical Offices (NSOs) need to develop methods for assessing the quality of registers, their metadata and data. The starting point for the quality assessment is the common statistical quality framework which has been developed at the level of the European Statistical System and the United Nations Statistical Commission and implemented in quality reports at national and regional levels. It contains such indicators as relevance, accuracy, timeliness, punctuality, comparability, coherence, accessibility and clarity. From the perspective of the process quality, some other indicators like best methods, cost efficiency and low response burden should also be considered [8]. Among many basic requirements, the most important ones resulting from the Polish experience are as follows:

2.1. Requirements for populations

For statistical purposes, administrative registers can be divided into base registers and specialised registers. The three base registers that can be subject to national laws are:

- The whole resident population
- All houses and dwellings in the country
- All active businesses in the country

Additionally, for geocoding and location purposes, some countries (including Poland) establish a fourth base register (spatial register) consisting of all territory division units with geometric boundaries of administrative and statistics units and x, y coordinates of each address points. Geographic Information Systems (GIS) can be fully introduced to register based census.

Other registers called "specialised registers" cover the whole or part of the population under specific laws, contain all additional necessary data to describe so-

cial, economic and environmental activity in a country and can be used for the evaluation of specific statistical phenomena.

2.2. Requirements for Identifiers in registers

Identifiers have to be unique, universal and stable:

- Only one ID per object;
- No two objects may have the same ID;
- The same ID in many (optimally: all) systems;
- Should remain attached to the subject forever (even after the subject's death);
- Should be subject to validation by digit check or checks against base registers.

2.3. Requirements for administrative variables

All administrative variables have to be understandable, properly described and published with:

- Well-defined concepts;
- Clarity and comprehensibility of definitions (mutually exclusive classes);
- Classifications according to the existing taxonomy;
- Congruence with the law (therefore laws should be simple and logical);
- Truth and verification;
- Known (published) quality.

3. The latest techniques for cleaning and preparation of administrative data

Administrative registers are designed for administrative needs and not statistical ones. This means that to enhance data quality, administrative registers have to be cleaned and transformed before being used in censuses or current statistics. If the registry owners clean data in their own database before sending them to the statistical office, then it is important to document this process in the metadata to avoid the duplication of data processing and data cleaning.

Based on experiences derived from the last census in Poland, a comprehensive model for transformation of data from administrative sources into statistical datasets can be proposed also for preparation of the 2021 census. The model comprises seven phases: preliminary preparation of an administrative register, data transition, validation and adjustment, integration, complex de-duplication, selection of best statistical value and creation of final statistical data set.

A classification procedure based on a model of quality assessment is needed before the selected registers are converted into the statistical register. In Statistics Poland, assessing the quality of administrative registers includes:

- timeliness of data,
- methodological compatibility,
- completeness,
- identification standards used in the registry,
- usefulness,
- compatibility of data in administrative sources with data obtained in the study/survey,
- possession of an identifier allowing integration of data from various sources.

After the quality assessment, the preliminary preparation of an administrative register data procedure is initiated. It covers importing, mapping, simple de-duplication and re-normalisation procedure. In the case of Poland, the importing procedure included:

- Consolidation of data from various sources;
- Extraction of data into the production environment based on the SAS software;
- Conversion of data into one format that was suitable for processing – SAS tables;
- Validation of imported data structure that was an integral part of this process.

In the preliminary preparation phase, redundant variables (i.e. those that will not be used) are excluded from further processing, and the remaining ones are named according to a fixed standard. Then identical records are removed from the dataset (simple de-duplication). The last step of this stage is to create a flat table containing specific subjects.

The next phase is data transition providing cleaning procedure for two kinds of data. The first is for IDs and address data variables and the second for variables describing statistical characteristics. In this phase, the data processing in the production environment consist of the following steps:

- Up-case – standardise all entries in the selected column of text to standard uppercase letters. Standard up-case only applies to columns in which raising capital letters will not alter the information;
- Profiling – create a report on data quality;
- Parsing (separation) or combining variables – applied in the case of several pieces of information joined in a single variable; for example, the division of the address into city, street ID, street name, house number and apartment number;

- Unification/standardisation of data according to firm rules;
- Standardisation with schemes – this is the correction of incorrect recording by the imposition of appropriate schemas. The schema is a table with two columns. One column contains wrong names and the other contains corresponding correct names (standard). Schemas are used to correct erroneous entries for names like province, county, municipality, town, street, street prefixes and country names. Schemas can also be used to replace outdated code names for new municipalities;
- Conversion – for the variables in the registers that are hallmarks of the entity (e.g. gender, education, marital status). The conversion means replacing differently stored descriptive values by the same information, i.e. transforming registry variable to the statistical standard.

The result of the data transition is a dataset in which the data are consistent, and the variables are standardised and have substantively correct values.

After the data cleaning, the validation and adjustment phase is applied. This stage is intended to check the results of processing the relevant rules, that is, to verify compliance with the assumptions of data consistency, accuracy and standard. The validation consists of checking the data, correcting abnormal values according to the algorithms prepared by methodologists, eventually excluding the unimprovable records from further processing. The stage ends up generating a report with data quality improvement, and on the quality basis the set goes to further processing or back to the stage of cleaning.

The next phase is integration. The dataset from the register is integrated with a reference dataset, which can be a pre-prepared list of statistical units like people, real estate, agriculture farms, business entities or another referenced administrative register. The integration also assumes the existence of a common variable/group of variables that define the connection. The integration process is related to the quality indicators on the coverage of objects. Incomplete coverage (under-coverage) – can be measured by the percentage of missing objects in the collection with respect to the reference collection. Over-coverage can be measured by the percentage of objects' source missing in the reference data set.

After integration the complex de-duplication stage is applied for the elimination of redundant units – equivalent but not identical. It applies in cases where the same variables for the repeated units contain other values. Multiple occurrences of units in the set are often the result of high-detailed data collection. The result of de-duplication is one record with all the possible and unique information for an entity without losing information from others.

As a result of the integration phase, a number of transformed registries are attached to a set of reference variables of many records, so there arises the problem of choosing the most appropriate value. The phase of the selection of statistical variable value from many registers solves that problem. There are some approaches like downloading of information only from one register or for a specific variable combining information from several registers in the specified order (selection of the most appropriate source of the second, third order etc.). It is also possible to use several ways to select best values:

- from the register covering the largest set of information,
- from the most recent register,
- related to the degree of filling of the variable,
- for quality/utility stored value for the variable.

When choosing the best value of a variable from a number of records, indicators of the following quality of the variables may be helpful:

- degree of integration (e.g., the percentage of objects in an integrated set, with respect to the Reference Collection),
- degree of filling (e.g., the percentage of objects in an integrated set for which the value of the variable is not empty),
- utility (e.g., the percentage of objects in an integrated set for which the value of the variable is not empty, minus the value of the variable outside of the scope),
- timeliness (the choice of a variable determined by the date 'status to' data).

It should also be noted that variables can be repeated not only between registers, but also within a single registry. In such cases, the methods for selection of values may be the same as described above. The selection of variable value from multiple-registers yields a single variable statistics with best value.

The result of the above steps is the creation of a statistical data set containing specific entity and assignments to the statistical variables. Such statistical register is available for use by analysts for multivariate analysis. Such process transfers data from the production environment to the target environment (analyti-

cal), based on quick data loading. According to the local law in some countries, it is necessary to anonymise data for further analysis. In such case, the statistical data base should be anonymised. For example in Poland, the following variables are removed from the analytical data base:

- Surname;
- Names;
- Street;
- House number;
- Apartment number;
- Social Security ID or other PIN;
- Phone number;
- All other variables containing data that could identify a specific person.

After conducting all the above operations, it is necessary to measure the quality of administrative data processing. At minimum, this should consist of:

- over-coverage error rate,
- under-coverage error rate – subjective indicator of completeness,
- objective indicator of completeness,
- imputation rate,
- data correction index.

All the above data processing stages were invited and checked by the Polish statistics during the 2011 census and constitute a valuable resource for the Polish statistics. Some elements of this experience later contributed to the UNECE Guidelines developed by the Task Force on the use of registers and administrative data for population and housing censuses [8].

4. The Polish case study: Population and housing census 2011

The national combined population and housing census conducted in Poland in 2011 was designed and implemented with the application of a mixed model, i.e. using data from administrative registers and data obtained from respondents (20% ad-hoc sample survey), with the use of electronic questionnaires.

The 2011 Act on the National Census of Population and Housing provided for the widest possible use of public administration information systems. Data not included in the public administration's information system or data not eligible for statistical quality were collected by means of the Internet application (CAWI) or by hand held (HH) electronic devices equipped with the electronic form application (CAPI mode). As a result, paper questionnaires were completely eliminated.

4.1. The use of administrative sources

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

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

- minimisation of the costs of statistics production,
- risk of an increased non-response in statistical surveys, including censuses,
- intensive development of IT systems of public administration, based on advanced technologies.

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

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

Using data from administrative sources required an in-depth understanding of the information resources, which were found in these sources. An analysis of all the sources and variables potentially useful for the censuses was carried out. The necessary metadata on approximately 300 administrative registers were collected, of which almost 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 classifications, with the existing Polish and EU statistics dictionaries. Appropriate weights

were determined both for the variables and administrative registers from which these variables came, 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.

Finally, 28 sources were used from Government and Local-Government administration, and from administrators outside public administration such as building administrators, housing co-operatives, power distribution plants and telecommunication operators. All the administrators of databases provided access to their information resources for the purposes of the population and housing census in 2011.

Data from administrative systems were used in the census:

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

To enable the administrators to transfer data from regionally dispersed systems via telecommunication channels, Statistics Poland developed 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 (encrypted channels). These solutions were also applied for collecting data from over 2500 local communities (LAU2).

The unit data obtained from registers were converted into statistical registers, simultaneously being subject to the process of cleaning, de-duplication and standardisation. The process was carried out in the Data Quality System (DQS) in the 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.

A "meta-information repository" was created to collect methodological, technical and operational meta-information. This ensured process control of data processing as well as monitoring of the course of processes, including measurement and collection of meta-information concerning quality at all stages of the process of data development, i.e. at data collection, processing, analysis and dissemination stages.

4.2. Other data acquisition methods used in the combined census

Poland was one of the first countries in the world to use a wholly innovative method consisting of several of the most modern techniques for collecting census data simultaneously. Apart from the use of IT systems and registers 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/Web Interview) – an online self-administered questionnaire, which entails checking the respondent data obtained from administrative sources, within a specified time frame, and, if needed, correcting them and providing missing information (self-enumeration);
- 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, among other things, if the questionnaire had been filled in accurately through logical and accounting controls.

The appropriate census architecture had to be constructed to enable the optimal application of advanced IT and telecommunication technologies in censuses. For the purposes of census design and implementation, Statistics Poland integrated various technologies in the IT Census System (from applications installed on mobile terminals, through applications managing and assisting in telephone interviews, to specialist bases, data warehouses and analytical and reporting tools).

4.3. GIS technology

For the first time in Poland's census history, GIS technology was used in 2011 to implement and monitor the enumerators' fieldwork.

With the use of various reference materials and registers containing spatial information, Statistics Poland

created spatial data for statistical address points and boundaries of statistical division of the country. All the data were collected together with x, y coordinates. Owing to that, digital maps used by census enumerators were an indispensable data source (to navigate and verify dwelling locations in the field), by local leaders (for on-line census monitoring within the LAU2 level administrative units), and by regional (NUTS2) and central supervisors (for on-line census monitoring on regional or global level). Digital maps were used to monitor on-line census progress in a defined area or for a specific enumerator (an on-demand location or daily route could be visualised on the map).

4.4. The Geo-statistics Portal

Statistics Poland developed and implemented a GIS portal (https://geo.stat.gov.pl) for dissemination of the census data spatially. The Geostatistics Portal is a platform for interactive cartographic presentation and the publication of data acquired in censuses. It serves to store, present and share information with a broad group of recipients.

The interface of the Geostatistics Portal allows quick and easy access to the 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. Apart 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.

5. Main achievements of the last combined census in Poland

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

- data were simultaneously collected, 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)), so paper questionnaires were completely eliminated, and were replaced by digital solutions,
- data from 28 administrative registers and 3 non-administrative systems were effectively integrated,
- the use of GIS technology helped conduct the census preparatory work and an on-going census process monitoring and made it possible 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 that have been developed will have a considerable influence on the methodology of future statistical surveys.

To guarantee progressive solutions, considerable efforts need to be expended with a view to developing a new census strategy. Attempts should be made to:

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

6. The construction of the list for census 2021 in Poland

The purpose of building a list for census 2021 is to create a broad, comprehensive and subjectively compilation set describing the population for Census 2021 and annual censuses after 2021. The construction of the list of the persons was based on the integration of data from many heterogeneous sources to obtain a consistent image of the collected data, It was assumed that the integration key would be a personal identification number (PESEL), because it is the main identifier found in most polish administrative sources. For building a list of persons, an IT system was created that includes not only determining the number of persons to be examined in the census, it also allows calculating census and other characteristics for statistical surveys.

The construction of the list for census 2021 will be implemented in three main stages:

- Preparation of the list of the address and housing based on the address identification system of

streets, real estate, buildings and flats (NOBC), including inhabited and uninhabited flats located in residential and non-residential buildings, and connecting spatial addresses;
- Preparation of the list of persons with assigned home addresses;
- Integration of lists constituting the basis for the implementation of the population and housing census.

The object-oriented approach was adopted in the construction of the address, buildings and dwellings list. The list consists of four tables:

- Buildings;
- Flats;
- Collective living quarters;
- Persons.

with a set of variables and connecting keys between individual Tables

Necessary data sources for constructing the addresses, buildings and dwellings list as follows:

- TERYT national official register of territorial division and NOBC address identification system for streets, real estate, buildings and flats;
- Public administration information systems and official registers:
- Universal Electronic Population Register System (PESEL);
- Central Register of Entities – National Register of Taxpayers (CRP-KEP);
- Central List of Insured – National Health Fund (NFZ);
- Comprehensive IT System – Central Register of Insured (KSI CRU) / ZUS;
- a new Insurance IT System (nSIU) – Agricultural Social Insurance Fund (KRUS);
- Agency for Restructuring and Modernization of Agriculture (ARiMR);
- Records of cities, streets and addresses (EMUiA).

Non-public information systems:

- Information systems of enterprises operating in the field of electricity sales (set of Energetyka ZE);
- Statistical survey for units conducting economic activity, i.e. the Statistical Units Database (BJS).

Necessary steps for building of the population data set are as follows:

- Collecting and merging unique PESEL numbers from the administrative registers;
- Validation of the PESEL number consists in checking the required number of identifier characters, compliance of the check digit and the possibility of occurrence of the date of birth in the calendar encoded in the number;
- The PESEL number validation algorithm includes automatic checking of back dates as well as future dates, including leap years;
- In the list of persons holding the personal ID number, it is necessary to calculate the variables; sex and age due to the need for further validation necessary for the construction of this list;
- Verification and designation of specific groups of persons due to specific characteristics.

The last stage of integration of the list of persons with the NOBC database, consisting of the address identification system, was built on the so-called full address key, which it consists of:

- name of the commune (gmina),
- city name,
- street name,
- TERYT commune code,
- TERYT city identifier,
- TERYT street identifier,
- number of the building,
- flat number.

In conclusion, it is worth noting that the presented way of building the list is based solely on data from administrative sources and will allow to cover the whole population and all addresses. This method also makes it possible to:

- full integration with other subjects, ensures high quality of data,
- fully automatic and repeatable process,
- provide the quality control at every stage of processing.

The developed method allows to use it in the next census and will be suitable for the post 2021 censuses.

7. Plans for 2021 census

The implementation of the census will be based on the methods implemented in the previous edition of the census in 2011. Rapid technological progress has a significant impact on the scale of modern solutions currently adaptable to the needs of the census. The experience gained in 2011, improvements in individual processes and conclusions drawn became the basis for the decision to implement such methodological, organiza-

tional and IT solutions that allow collecting information from every inhabitant of Poland subject to the census, without the need to incur increased financial outlays during the implementation of this project.

Census 2021 in Poland will be implemented as a full survey carried out from April 1 to June 30, 2021. Data from natural persons will be collected primarily by means of the Internet self-enumeration method. The self-enumeration will consist of providing data about oneself in accordance with a set of questions, via an interactive web application available on the Statistics Poland's website.

Particular emphasis will be placed on ensuring the highest possible response rate of respondents through the web-based self-enumeration system. Persons who will not have access to electronic devices enabling internet self-enumeration, digitally excluded persons, the elderly (who are not able to make internet self-enumeration) and those unable to make self-assessment due to their health condition, will be provided with free access to dedicated rooms supplied with the computer equipment and the software necessary to conduct the self-enumeration. A suitable room – meeting the conditions of confidentiality and freedom to fill in questions in the census application – will be available in every commune (e.g. commune office), statistical office, voivodship office and other public places. At the request of a natural person, the necessary assistance will also be provided with regard to the handling of an interactive application.

The census 2021 adopts a mixed census implementation model by combining the data from registers with the census survey. Moreover, the data collected from the respondents will be compared with the data from the registers at the stage of preparing the results. The result of the comparison will be the basis for the verification of the quality of the register data.

The population and housing censuses shall provide for the following data collection methods:

a. Internet self-enumeration (CAWI), carried out by the person included in the census between April 1, 2021 and May 16, 2021.

In particularly justified cases, if a natural person is not able to fulfil the obligation of the self-enumeration census, data shall be collected using the following methods:

b. a telephone interview (CATI), conducted by a computer-aided telephone interviewer, consisting of the collection of data from natural persons included in the census,
c. direct interview (CAPI), conducted by a field enumerator using a mobile device equipped with software dedicated to conduct a census involving the collection of data from individuals included in the census. The number of field enumerators will be determined for particular voivodships (NUTS2 level) and communes (gminas). This will take into account factors related to the availability of the Internet, the age structure of the population in the given area, population density, terrain circumstances, etc. A larger number of enumerators will be directed to communes with "more challenging" conditions.

In order to test the adopted methodological, organisational, technical and publicity related solutions, two trial censuses are planned to be carried out before the real one. It is assumed that, during the trial censuses, the following topics, among others, will be tested:

a. the quality of the address-and-residential and personal lists, estimation of the list of unavailable units,
b. verification of the informational scope of the application,
c. functionality of electronic data collection channels,
d. evaluation of the questionnaire and individual questions (interview time, difficulty level of the formulated questions, logical control of the electronic form),
e. efficiency of data acquisition on portable electronic devices,
f. technical issues related to the transmission of completed electronic forms to the relevant servers,
g. operation of an integrated IT system, i.e. effectiveness of the Internet self-enumeration system and the telephone surveys and an interview with the enumerator,
h. effectiveness of popularization solutions (including activities in traditional and on-line media and sending letters of the President of Statistics Poland to the citizens).

Immediately after the census, if necessary, a complementary research will be conducted to collect information that could not be collected in the census and that should be collected to ensure the completeness of the census, based on a sample drawn from the statistical survey frame, updated with data from the census frame.

It is also planned that a control survey will be carried out to check the quality of census.

8. Plans for post-2021 census

For many years, Polish Statistics has been collecting data from administrative sources, social networks, sensors, GPS technology, etc. The censuses after 2021 will undoubtedly be based on the full use of administrative and non-administrative registers. Annual updates after 2024 will require full automation of data processing from administrative sources and other alternative sources such as Big Data combined with statistical data.

Data integration, administrative data and geospatial information are becoming increasingly important in our statistical production process. However, data silos have been a hindrance to data integrity, data integration and a barrier to production efficiency.

Regarding post-2021 census, Statistics Poland will create a reference architecture framework (RAF) based on a process-oriented model of a statistical production mainly:

- To enable departure from a stove-pipe production, based on data silos;
- Pointing out how iterative loops with and between the GSBPM phases, various data flows should be supported by IT;
- Defining the framework for secure data flows, according to the GSBPM;
- To assure of process and data quality management.

In the Polish Statistics, there are plans to address the following census-related challenges for post-2021 census:

- New data sources:
  - collecting and managing large amounts of data sets including Big Data,
  - analysis and assessment of the quality of collections and data in data sets,
  - assessment of the impact of the quality of the data sources used on the quality of the resulting data from censuses.
- New techniques and technologies;
- Developing Smart Statistics – new methodologies relating to the integration of data from various online sources.

9. Conclusions and recommendations

According to the Polish experiences and UNECE guidance [8], the following recommendations should be considered in order to facilitate the use of administrative registers in censuses:

- Review or prepare national legislation system as a basis for the creation of a population and housing registers and permission to use the data contained in those registers for statistical purposes;
- Establish a universal personal identification (unique identity) system to facilitate proper linking of data. Identifiers have to be unique, universal and stable;
- Keep transparent partnership with administrative agencies;
- Identify with users the key requirements for the census. Review current registers to understand strengths and gaps and build statistical registers from different administrative sources;
- Establish quality frameworks and assessment processes before using registers in censuses;
- The quality of registers and quality of metadata and data from registers is the most important element that should be considered while deciding about the use of administrative registers in the production of statistics;
- Choose proper methodologies for data linking, managing the missing and inconsistent data and eventually linking administrative data with traditional data;
- Establish organisational structure designed for register-based census suitable for different circumstances derived from new methods applied to data collection and data processing;
- Improve data quality, the administrative registers have to be cleaned and transformed before being used for census or current statistics;
- Regarding the administrative register sources, it is worth distinguishing between indicators describing quality of sources, quality of metadata and quality of data;
- Use the three base registers that can be subject to national laws: the whole resident population register, the building and housing registers and the statistical business register, and their complementary registers. Consider all possibilities to compile it from auxiliary administrative and non-administrative sources, including Big Data;
- For geocoding and location in space purposes, establish a fourth base register (spatial register) including all territory division units with geometric boundaries of administrative and statistics units and x, y coordinates of address points;
- Implement Geographic Information Systems (GIS) for linking statistical data with spatial data for complete geographical coverage and concep-

tual detail of all the variables, whether available in the registers or not;
- All administrative variables have to be understandable, properly described and published;
- Communicate to all stakeholders that statistics act as the one-way road. The data that comes in must never come out in an identifiable manner.

References

[1] UNSD: Overview of national experiences for PHC of the 2010 round. http://unstats.un.org/unsd/demographic/sources/census/wphc/UNSD/overview.pd.

[2] United Nations, Statistics Division. Principles and Recommendations for Population and Housing Censuses, Revision 3. 2015. ST/ESA/STAT/SER.M/67/Rev.2.

[3] UNECE. (2007) Register-based statistics in the Nordic Countries. Review of best practices with focus on population and social statistics, New York and Geneva. Unable to open.

[4] UNECE. (2011) Using Administrative and Secondary Sources for Official Statistics: A Handbook of Principles and Practices, New York and Geneva. Unable to open.

[5] United Nations, Statistics Division, Overview of National Experiences for Population and Housing Censuses of the 2010 Round. June 2013.

[6] Conference of European Statisticians Recommendations for the 2020 Censuses of Population and Housing, UNECE, Geneva, October 2015 http://www.unece.org/ publications/2020recomm.html.

[7] Valente Paolo, 2015, From the 2010 to the 2020 census round in the UNECE region – Plans by countries on census methodology and technology. Paper submitted to the Meeting of the UNECE-Eurostat Group of Experts on Population and Housing Censuses, Geneva, 30 September to 2 October 2015.

[8] UNECE Guidelines on the use of registers and administrative data for population and housing censuses, Geneva, December 2018. http://www.unece.org/index.php?id=50794.