Economic and Social Commission for Western Asia (ESCWA)

Undertaking Population and Housing Censuses

Register-based and combined methodologies

United Nations
Beirut, 2018

Acknowledgements

Executive summary

The present report outlines the potential use of administrative registers in population and housing censuses. It focuses on the case study of Poland, which conducted a population census in 2011 using combined administrative registers. The report discusses some preconditions for the transition from traditional census methodologies to combined census methodologies. It also considers the experience of some Arab countries in terms of the potential use of registers for the 2020 census round.

CONTENTS

Introduction and definitions

Population censuses have a history of thousands of years, but a widely used approach to census taking was established in Europe during the eighteenth century, which became known as the traditional approach. The register-based census approach emerged during the past 30 years.

The register-based approach does not collect data directly from the population. Instead, administrative data is obtained from various registers (including population registers, building/address registers, and social security registers) through a matching process, making use of personal identification numbers (PINs) if available.

During the past 20 years, some countries in the United Nations Economic Commission for Europe (UNECE) region[1] developed innovative methods to conduct combined census taking by using administrative data linked to a full or sample survey of specific variables. Those methods can be used to make the transition from a traditional to a register-based census.

In the 2010 round,[2] there was a significant increase in the number of countries conducting a register-based census (from 3 to 9) or a combined census (from 5 to 10) and a corresponding decrease in the number of countries conducting a traditional census (from 40 to 34).

According to plans for the 2020 round, of the 48 UNECE countries for which information is available, 14 countries plan to conduct a register-based census (29 per cent), 12 countries a combined census (25 per cent) and only 22 countries a traditional census (46 per cent). If the 32 countries that are members of the European Union (EU) and the European Free Trade Association (EFTA) are considered, the results are that, in the 2020 round, 13 countries plan a register-based census (41 per cent), 9 countries a combined census (28 per cent) and only 10 countries a traditional census (31 per cent).

In the Arab region, only one country (Bahrain) undertook a population and housing census using administrative registers in 2010. The Gulf Cooperation Council (GCC) countries intend to use registers for the 2020 census round.

Figure 1. Number of UNECE countries by census method in the 2000-2020 census rounds
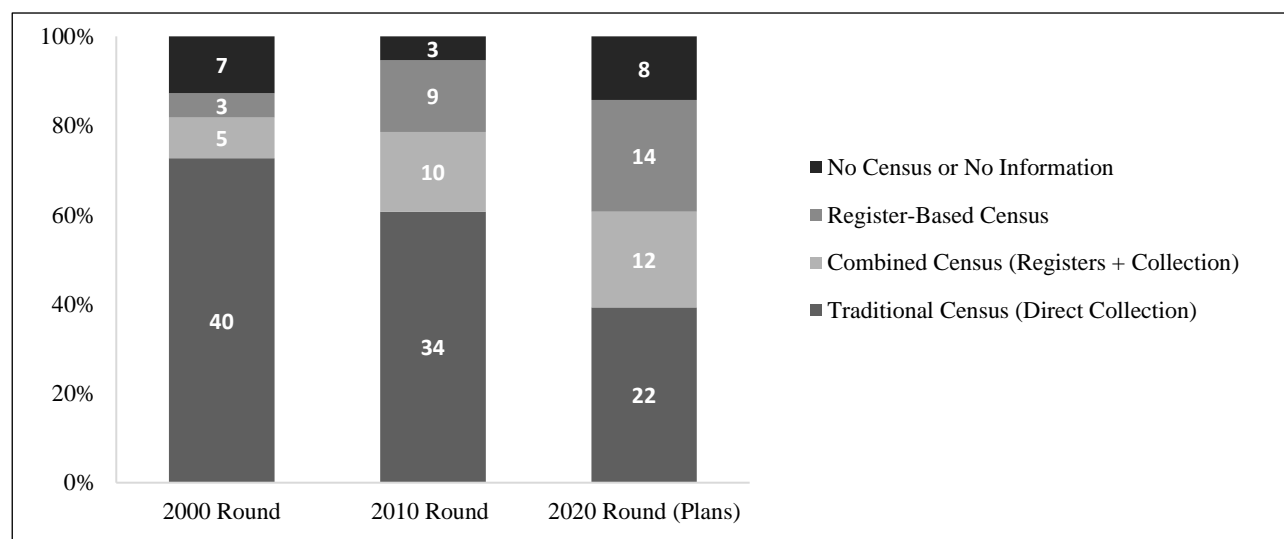
Source: Valente, 2015.

Note: In 2006, Montenegro became independent, increasing the number of UNECE countries from 55 to 56.

---

[1]   A list of UNECE members is available from https://www.unece.org/oes/nutshell/member_states_representatives.html.

[2]   Valente, 2015.

Regardless of the approach used, a population census can be defined as follows: "(T)he total process of collecting, compiling, evaluating, analysing and publishing or otherwise disseminating demographic, economic and social data pertaining, at a specified time, to all persons in a country or in a well-delimited part of a country".[3]

Similarly, a housing census can be defined as: "(T)he total process of collecting, compiling, evaluating, analysing and publishing or otherwise disseminating statistical data pertaining, at a specified time, to all living quarters and occupants thereof in a country or in a well-delimited part of a country".[4]

Regarding the applied methods of census taking, there are four main possibilities:

- Traditional full field enumeration;

- Rolling census (continuous surveys);

- Combined census by integration of administrative sources with existing surveys or ad hoc sample surveys;

- Fully register-based census.

The concept of register varies across countries depending on local law, practice, governmental information systems and pragmatics. Some common characteristics can be described as follows:

- Register: a systematic collection of identifiable unit-level data with the ability of updating, correcting or extending the established data set;

- Administrative register: a register primarily used for public administrative purposes in a governmental information system;

- Non-administrative register: a register operated by private organisations for business purposes fulfilling non-public administration needs;

- Statistical register: a register created by processing data from administrative and non-administrative registers for statistical purposes.

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.[5] Some register-based census countries are missing some of the census variables in all the available registers and choose to support their census with data from already existing sample surveys. What all register-based census countries have in common is that no census questionnaires are used to collect information about the population. Therefore, register-based censuses are in general much cheaper than combined censuses and especially cheaper than traditional censuses.[6]

---

[3] United Nations Statistics Division, 2008.

[4] Ibid.

[5] United Nations Economic Commission for Europe (UNECE), 2015.

[6] UNECE, 2016.

I. PRECONDITIONS FOR USING ADMINISTRATIVE REGISTERS IN CENSUSES

A. QUALITY OF REGISTERS, METADATA AND DATA

Before deciding whether or not to use administrative registers in censuses, National Statistics Offices (NSOs) need to develop methods for assessing the quality of registers, their metadata and data. The starting point for quality assessment is the common statistical quality framework 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 comprises such product indicators as relevance, accuracy, timeliness, punctuality, comparability, coherence, accessibility and clarity. From the perspective of process quality, some other indicators should also be included, including best methods, cost efficiency and low-response burden.

B. 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 localisation purposes, some countries 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 censuses.

Other registers known as "specialised registers" cover all or part of the population under specific laws and can be used for the evaluation of specific statistical phenomena.

C. REQUIREMENTS FOR IDENTIFIERS

Identifiers must be unique, universal and stable:

- Each object should have only one ID;
- No two objects should have the same ID;
- The same ID should be used in many (if possible, all) systems involved;
- The ID should remain attached to the subject forever (even after death);
- The ID should be subject to validation by digital checks or checks against base registers.

D. REQUIREMENTS FOR ADMINISTRATIVE VARIABLES

All administrative variables must be understandable, properly described and published. They must have:

- Well-defined concepts;
- Clarity and comprehensibility of definitions, with mutually exclusive classes;
- Classifications according to the existing taxonomy;
- Congruence with the law;
- Truth and verification.

II. LATEST TECHNIQUES FOR CLEANING AND PREPARATION

Administrative registers are designed for administrative needs rather than for statistical use. To enhance data quality, administrative registers must be cleaned and transformed before being used in censuses

or statistics generally. If register personnel clean data in their own databases before sending that data to statistics sections, it is important to flag and document that the data have been cleaned, so that resources are not wasted duplicating data processing and data cleaning.

On the basis of experiences derived from the most recent census in Poland, the author of the present report proposed a comprehensive model for transformation of data from administrative sources into statistical data sets. The model consists of seven phases: preliminary preparation of an administrative register, data transition, validation and adjustment, integration, complex deduplication, selection of best statistical value, and creation of final statistical data set.

Figure 2. The generic model for transformation of data from administrative sources into statistical datasets
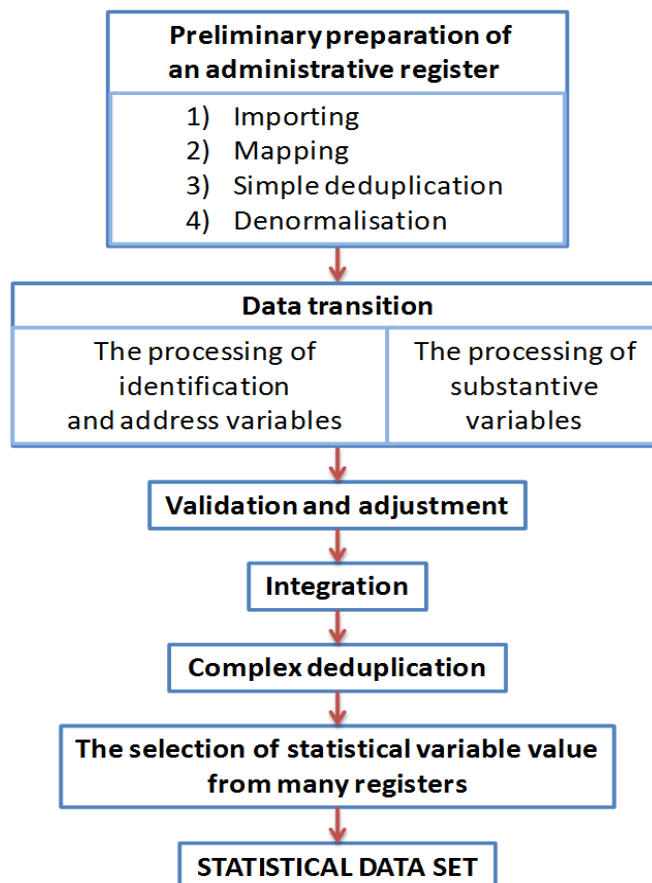
Source: Author.

A classification procedure based on a model of quality assessment is needed before the selected registers are converted into a statistical register. Assessing the quality of administrative registers includes the following factors:

- Timeliness of data;
- Methodological compatibility;
- Completeness;
- Identification standards used in the registry;
- Usefulness;
- Compatibility of data in administrative sources with data obtained in a study/survey;
- Having an identifier allowing integration of data from various sources.

After the quality assessment, the preliminary preparation of an administrative register data procedure is initiated. That covers importing, mapping, simple deduplication and renormalisation procedure. In the case of Poland in 2011, the importing procedure entailed:

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

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

The next phase is data transition, providing a cleaning procedure for two types of data. The first is for IDs and address data variables. The second is for substantive variables describing statistical phenomena. In that phase, data processing in the production environment comprises the following steps:

- Up-case – standardise all entries in the selected column of text to standard uppercase letters. Standard up-case applies only to columns in which raising capital letters will not help to change the substantive entries, with standard capital letters of feature columns containing the house number, apartment number and a variable containing identification numbers such as social security ID;

- 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, such as 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 schemas – that is, 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 (standards). 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 substantive variables considered factual information in the register that are hallmarks of the entity (for example, gender, education and marital status).The conversion means replacing differently stored descriptive values by the same information, transforming registry variable to the statistical standard.

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

After data cleaning comes the validation and adjustment phase. That stage is intended to check the results of processing the relevant rules, in order to verify compliance with the assumptions of data consistency, accuracy and a given substantive standard. The validation consists of checking the data, correcting abnormal values according to the algorithms prepared by methodologists, eventually excluding from further processing records that cannot be improved. The stage ends up generating a report with data quality improvement. If the quality is acceptable, the set proceeds to further processing. If the quality is unacceptable, more cleaning must be done.

The next phase is integration. The data set from the register is integrated with a set of references that might be a pre-prepared list of statistical units such as people, real estate, agriculture farms, business entities or another reference administrative register. Integration 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 cover 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 being included that were not in the reference data set.

After integration, the complex deduplication stage is applied to eliminate 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 highly detailed data collection. The result of deduplication is one record with all the possible and unique information for an entity without losing information from other entities.

As a result of the integration phase, a number of transformed registries are attached to a set of reference variables of many records. The problem arises of how to choose the most appropriate value. The phase of selecting statistical variable values from many registers solves that problem. Some approaches include downloading information only from one register or for a specific variable combining information from several registers in a specified order.

It is possible to use several ways to select best values, as follows:

- Using the register covering subjective reference collection that is greatest or where the register is integrated to the greatest extent with a set of references;

- Using the most recent register;

- Using the register where the degree of filling of the variable is greatest;

- Using the register with the best quality/utility stored values for the variable.

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

- Degree of integration (for example, the percentage of objects in an integrated set, with respect to the reference collection);

- Degree of filling (for example, the percentage of objects in an integrated set for which the value of the variable is not empty);

- Utility (for example, 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);

- Actuality (the choice of a variable determined by the date 'status to' data).

It should be noted that variables can be repeated not only between registers but also within a single registry. Methods for selecting values might in such cases be the same as described above. The selection of a variable value from multiple registers yields a single variable statistic with best value.

The result of the above steps is the creation of a statistical data set containing specific entity and assignments to statistical variables. Such a statistical register is available for use by analysts for multivariate analysis. The process transfers data from the production environment to the target analytical environment, based on quick data loading. According to local law in some countries, it is necessary to anonymise data for

further analysis. In such cases, the statistical data base should be anonymised. For example, the following variables can be removed from the data base:

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

After conducting all of the above, it is necessary to measure the quality of administrative data processing. At a minimum, that should include evaluating:

- The over-coverage error rate;
- The under-coverage error rate, using a subjective indicator of completeness;
- An objective indicator of completeness;
- Imputation rate;
- Data correction index.

III. A CASE STUDY: OVERVIEW OF THE USE OF POPULATION REGISTERS IN POPULATION AND HOUSING CENSUS

A. POLAND: POPULATION AND HOUSING CENSUS 2011

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

The Act concerning the 2011 national population and housing census stipulated that the information systems of public administration should be used as widely as possible. Data 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, the use of paper questionnaires was eliminated.

1. *The use of administrative sources*

The starting point was the use of administrative sources already existing within 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 according to the scope and time specified.

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

- The minimisation of the costs of statistics production;
- The risk of an increased non-response in statistical surveys, including censuses;
- The 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:

- Effective use of administrative and non-administrative systems;
- Reduced census costs;
- Reduced social burden connected with data transfer;

- Improved data safety;
- Guaranteed survey harmonisation;
- Availability of information for future annual censuses based on registers;
- Availability of data from administrative registers for levels of territorial disaggregation;
- Possibility to identify double-entry errors (over-counting);
- Creation of a micro database supporting indirect estimation, modelling at the unit level;
- Improved estimation for small areas;
- Improved coherence and reliability of statistical data.

Using data from administrative sources required an in-depth understanding of information resources that were found in those 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 those registers, separate records were opened and all variables from the sources subjected to a utility analysis. The variables were evaluated with regard to their conformity, in terms of definitions and classifications, with the existing Polish and European Union statistics dictionaries. Appropriate weights were determined both for the variables and administrative registers from which those variables came, taking into consideration their utility and quality. The knowledge concerning the quality and utility of variables from different registers formed a basis for the rules of merging data, plus their estimation and imputation in the operational base of microdata.

In the most recent census in Poland, 28 sources were used from government and local government administrations, and from administrators outside public administration, such as building administrators, housing cooperatives, energy 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, plus a sampling frame.

To enable administators to transfer data from regionally dispersed systems via telecommunication channels, the Central Statistical Office (CSO) of Poland constructed an electronic platform for data collection and processing, together with a net-based application for direct data transfer via electronic means in a secure connection, using encrypted channels. Those methods were also applied to collecting data from more than 2,500 local self-governments (LAU2).

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.

The Master Record was the set of variables derived from the registers containing information that was introduced to the census forms (pre-filling personalised electronic questionnaires) in order for respondents to verify (confirm or update).

2. *Quality evaluation of data from administrative sources*

For census purposes, there were four stages of quality assessment evaluation of data sources (systems, metadata, data sets and final statistical products), plus evaluation of resulting data.

With regard to the data sets from administrative data sources, quality assessment of raw data sets provided by the administrators and sets after a transformation (that is, after adjusting them for use in the census) was carried out. Quality assessment was also carried out at all stages of development of administrative data and the data combined with data from other sources. The schema below shows a graphical representation of the processes in the census, including quality assessment.

A metainformation repository was created to collect methodological, technical and operational metainformation. That 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: data collection, data processing, data analysis and data dissemination.

Figure 3. Generic Statistical Production Model for Combined Census
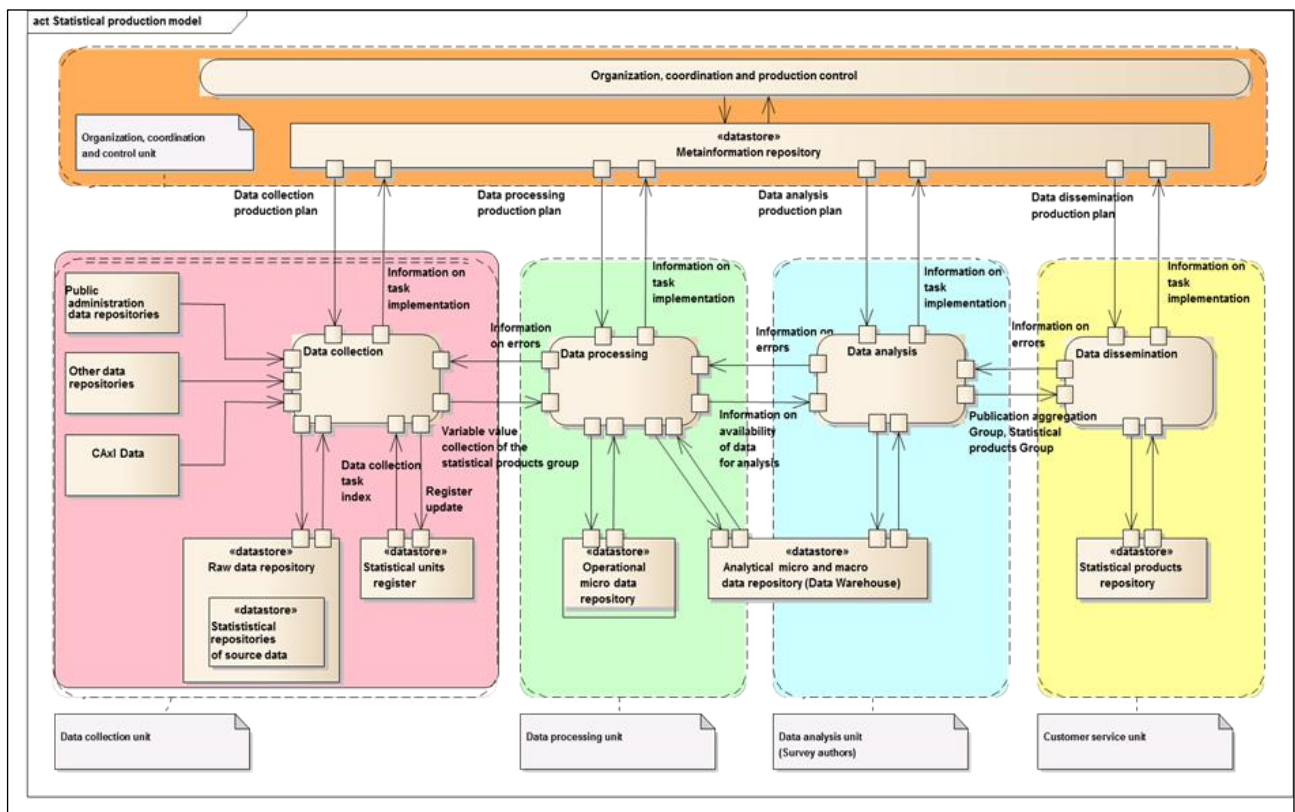
Source: Author.

3. *Other data acquisition methods used in the combined census*

Poland was one of the first countries in the world to use a totally innovative method comprising 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 that 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;

9

- 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 on an internet browser) verified if the questionnaire had been filled in accurately, among other things, through logical and accounting control.

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, the CSO of Poland used the IT Census System (ISS). The system, comprising more than 10 components implemented by different contractors, provided IT assistance for all operations within the census. The ISS integrated various technologies (from applications installed on mobile terminals, through applications managing and assisting telephone interviews, to specialist bases, data warehouses, plus analytical and reporting tools).

Pursuant to the National Census Act of 2011, the Operational Microdata Base (OMB) was developed, prepared, and implemented by the CSO. The system created included hardware system tool infrastructure (computer hardware, system software and tool software) and application software (computer programs).

Figure 4. Combined Census architecture – The IT Census System and three main stages
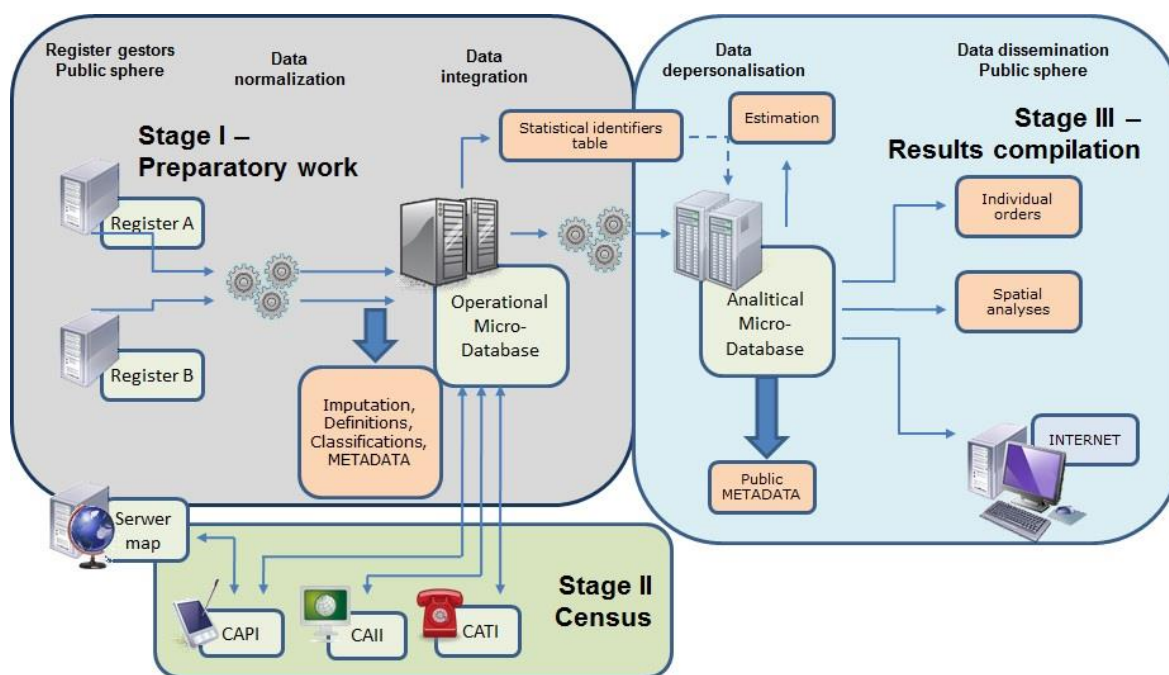
Source: Author.

The ISS employed various solutions to ensure 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 by the CSO, which were later implemented in practice. They included:

- The preparation of normalisation, control and correction rules for data sets from administrative systems, including data imputation (in administrative data sets);

- The preparation of rules for the synchronisation of data from administrative systems (the harmonisation of base periods);

- 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 and clear linkage;

- 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 and creating new objects, such as households and families;

- The preparation of a data-estimation model/method using data from administrative systems and statistical surveys;

- The preparation of data anonymization rules.

The OMB enabled the inclusion of data transmitted in electronic form through four information channels by entities obliged to do so under the Act. The entities included those maintaining administrative registers and persons covered by the census. Data was obtained via the internet, telephone interviews conducted by statistical interviewers or face-to-face interviews conducted by census enumerators. Complete cleansing was achieved by processes connected with the control, correction and linking of data. In the following step, 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. The subsystem constitutes the Central Metadata Repository (CMR).

The role of the AMB is to store depersonalised census data in their final form. In that data set, relevant statistical analyses are carried out to provide results for publication of census information (census products). The AMB allows recipients of statistical information to acquire data quickly 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 cartograms (choropleth maps) and cartodiagrams.

4. *Geographic Information System (GIS) Technology*

For the first time in Poland's census history, Geographic Information System (GIS) Technology was used in 2011 to implement and monitor field work by enumerators.

Various reference materials and registers containing spatial information were used to create spatial data for statistical address points and borders of statistical divisions of the country. Digital maps used by census enumerators were an indispensable data source (to navigate and verify dwelling locations in the field) by local

leaders (for online census monitoring within the LAU2 level administrative units), and by province (NUTS2) and central supervisors (for on-line census monitoring by province or at a global level). Maps were used to monitor online the census progress in a defined area or for specific enumerator positioning. For example, an on-demand location or daily route could be imposed on a map.

5.   *The Geostatistics Portal*

The Geostatistics Portal is a tool 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 portal functions on two levels. Internal users (statisticians) have access to both unit and aggregate data. External users have access only to aggregate data, which are published so as to provide 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 maps) and cartodiagrams. It is possible to set one's own parameters for the visualisation of a thematic area for a given cartogram. Parameters include measure, aggregation level (territorial division unit) and the number of intervals. Aside from the possibility of using readymade spatial analyses in the Geostatistics Portal, internal users can create customised thematic maps based on a selected feature of the data model, using dynamic spatial analyses such as linear or distance analyses, or object buffering.

6.   *Main achievements of the most recent combined census in Poland*

The census in Poland was innovative worldwide as well as countrywide because of the following factors:

- Data were collected simultaneously, without paper, from four different channels: administrative registers, internet self-enumeration (CAII), direct interviews conducted by census enumerators using electronic questionnaires (CAPI), and telephone interviews conducted by statistical interviewers (CATI);

- Data from 28 administrative registers and 3 non-administrative systems were integrated effectively;

- Paper questionnaires were eliminated and replaced by ICT solutions;

- GIS technology helped preparatory work for the census and provided on-going process monitoring. The technology made possible the compilation and presentation of results based on multi-dimensional spatial analyses;

- The IT Census System ensured a high level of security for processed data;

- The modern statistical data processing technologies developed 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.

The new technology used in the 2011 census can also be implemented in questionnaire-based surveys. It is cheaper, employs up-to-date control mechanisms, enhances the quality of the material collected and reduces the burden on respondents.

The time frame proved correct. The deadlines set for the data collection stage (specified in the Act) did not have to be extended, and neither did the census budget.

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

In the next census round, Poland expects to use newer IT technology that will have evolved by then.

Considerable efforts are needed to develop a new census strategy that will provide progressive improvements. Attempts should be made to:

- Reduce census costs;
- Use administrative sources in an even more effective way;
- Reduce social burdens connected with data transfer;
- Improve the safety of transferred data;
- Improve the coherence and reliability of statistical data.

B. CENSUSES IN JORDAN AND BAHRAIN

The development of a register-based population census system is a long process that can take years or even decades. In some Arab countries, however, electronic questionnaires replaced paper questionnaires fairly rapidly. The most spectacular example is the modern electronic 2015 census conducted in Jordan.

Several statistical offices in Arab countries, mainly Gulf Cooperation Council (GCC) member states, are considering using administrative registers in preparation for the 2020 census round. All GCC member countries agreed to conduct their censuses in the same year and month (April 2010) and repeat them every 10 years. The most advanced example is Bahrain, with the register-based census conducted in 2010.

The examples of Jordan and Bahrain are inspirational for other Arab countries in connection with the 2020 census round. Jordan's example could be suitable for countries that lack good quality, stable administrative registers. Bahrain's example could be followed by countries that have more advanced governmental information systems and high-quality administrative registers. Arab countries have different levels of administrative system maturity, but cooperation with relevant administrative bodies to develop reliable statistical registers as data sources for censuses and other statistical projects is crucial.

The use of modern technologies can reduce costs, the dissemination time and the burden on the public. However, applying the technologies is time consuming. It also requires staff expertise and financial resources.

C. THE POTENTIAL USE OF REGISTERS IN POPULATION AND HOUSING CENSUSES IN SELECTED ARAB COUNTRIES

The first data items to be taken from registers might include addresses, basic demographic data items, civil registration information and income data. It is essential, however, that the countries introduce a comprehensive and high-quality population register and a system of common identification numbers before attempting to link data from different administrative sources. The registers might include data maintained in a population register and a register of buildings and dwellings, as well as data from a business register. Such registers should cover all people resident in the country.

All statistical units should be linked to one another by means of identification systems. There should be the capacity to link persons to household dwelling units and to the dwellings and buildings in which they live, and for employees to be linked to their employers. Employers and buildings also need to be linked, in order to determine workplace. All units should be geographically located by using local area codes or map coordinates.

1. *Current situation*

The most recent census in Bahrain was conducted in 2010. The reference census day was 27 April 2010. The main scope of the census was population, housing, buildings, establishments and agriculture statistics. In the pre-census preparation work, all necessary stages were planned and executed:

- Understanding the legal framework;
- Defining variables needed for the census;
- Defining census variables and their ownership;
- Defining census data codes;
- Establishing the availability of administrative census variables;
- Establishing the update mechanism for census variables;
- Ensuring the quality of census administrative variables;
- Evaluating infrastructure readiness;
- Evaluating marketing, feedback and trust.

All the steps described above are crucial for a register-based approach to censuses and are similar to the steps described in the case study of Poland.

Small differences include the lack of distinction between the quality of administrative variables and the quality of each administrative register. Also, some solutions for cleaning and transforming administrative registers to a statistical register such as the National Data Set (NDS) need to be clearly mentioned in the preparation phase.

Regarding the availability of census administrative variables, it was estimated that 88 per cent of data could be obtained directly from registers. Such good availability was possible because of the solid administrative information infrastructure established in Bahrain between 1979 and 2009.

In Bahrain's Government Computer Data Centre (which includes central servers and a mainframe), core administrative registers were created. They included a Central Population Register, a Central Establishment Register and, finally, a National Data Set (NDS) that incorporated update mechanisms for data ownership by most government entities. The infrastructure was supported by the Government Data Network (GDN) and its legal framework, plus a Ministerial resolution for terms, national classifications and coding systems.

The information system was enhanced by security regulations and policies plus direct update links with the civil service and regulatory bodies responsible for immigration, electricity and pension funds. The steps described above were prepared and executed along the lines of work conducted in Europe for register-based or combined censuses.

In Poland, the OMB was established in 2009 for census purposes (linking and processing data obtained from administrative registers and other electronic channels for data collection) and for supporting statistical surveys in general.

In Bahrain, the NDS collected all population data required and was the main source for census operations. During the cleaning process, census data were extracted, copied and filtered by the Central Population staff for further processing.

The Census Team analysed the census data prepared and, where necessary and possible, improved data quality. Some errors were corrected automatically by means of a script prepared by the Development Team. The results were returned to the NDS for updating. Other errors were corrected manually or were verified and corrected by census field staff as a result of ad hoc field surveys. The data verified and corrected that way were

also sent to the NDS for updating. As a result, the quality of the NDS improved significantly and became a stronger reference source for the register-based population census.

2.   *Challenges and opportunities*

The methods described above were implemented effectively. It should be noted, however, that Bahrain, with its small population of 1.5 million, was able to use procedures similar to those used by small countries in the European Union for register-based censuses. In larger countries, as with Jordan and Poland, the relevant administrative data are spread across hundreds of registers. It is impossible to keep in one database the data for the whole population with combined characteristics contained in one unit record referring to each individual.

For larger countries, other methods (such as the multilayer approach used by the OMD in Poland) are more suitable for linking and processing data from a large amount of dispersed and distributed registers and from other computer-assisted channels of data collection (CAxI). The author of the present report recommends that Bahrain consider implementing a multilayer operational microdata base architecture for the 2020 census.

In Bahrain, as in Poland and other countries, the use of administrative registers encountered obstacles. The main challenges included data missing from administrative records, data security issues, plus insufficient quality of variables and records. As a result, the transformation and validation processes from administrative registers to statistical registers were adversely affected. Other challenges included variance in classifications and definitions across entities producing data, plus a lack of skilled human resources.

Another issue affected field operation and phone call activity introduced during the most recent census in Bahrain. Field survey and telephone data collection only occasionally obtained missed data caused by error or low quality. About 12 per cent of data variables initially were not available.

Despite the above obstacles, the new approaches and solutions prepared and implemented during the 2010 Bahrain census were impressive, deserving high praise and attention.

3.   *Transition from traditional census to register-based census in Jordan.*

Jordan's 2015 census was an excellent example of how up-to-date technology can be utilised to compile statistics. The census used all possible electronic channels for data collection based on computer-assisted personal/telephone interview methods (CAPI and CATI). Because of the unsatisfactory quality of administrative data, the administrative registers were used only on a limited scale, mainly for preparing population census frames. Despite that, some goals of modern electronic census were achieved, including:

- Reducing respondent burden;
- Using existing information;
- Improving timeliness;
- Reducing costs;
- Adhering to government directives;
- Taking a modern census approach.

Taking into consideration the experience in Poland, the author of the present report recommends that Jordan consider the following steps in order to make the transition from the current state of census methodology to a combined or fully register-based census. Because of the assessment of the inadequate quality of administrative data sources and variables contained in the registers, it is necessary to prepare and execute cleaning processes for census purposes. The main requirements are:

- Developing quality indicators for data sets at each stage of data processing and the procedures for calculating their values;

- Developing procedures for bringing data from administrative sources to achieve full compliance or minimum discrepancy, using appropriate statistical methodology;

- Developing procedures for normalisation, editing data sets from the administrative systems, including data input from administrative data sets;

- Developing procedures for synchronisation of data from administrative systems;

- Developing rules for linking data from different administrative systems;

- Developing rules for linking data from administrative systems with data from computer assisted web/personal/telephone interview (CAWI, CAPI and CATI);

- Developing rules for the calculation of final unit records with best values of census variables;

- Developing rules for the anonymization of final unit records of census data.

The author of the present report also recommends that Jordan assess the quality of data processing and census results. The main quality aspects are relevance, accuracy, comparability and coherence. The main requirements are:

- Developing rules for missing data completion, involving imputation and calibration;

- Developing rules for creating derived objects, such as households and families;

- Developing a model/method of data estimation with the use of data from administrative systems and sample surveys;

- Developing rules for calculating data outputs;

- Developing quality indicators at each stage of the census, plus procedures for calculating their values.

Good relations between the people responsible for registers and the people responsible for statistics in administrative agencies are crucial for achieving high-quality statistics.

Administrative agencies should have a good understanding of how data will be used statistically, plus a clear idea of agency roles and responsibilities. Requirements from Jordan's Department of Statistics must be clear and transparent, with relationships between the department and agencies built on effective communication and trust.

Evolving from traditional (even modern) census methods to a fully register-based census is a time-consuming task that requires considerable investment, new skills, organizational culture change, management focus, plus effective cooperation with other agencies. Administrative registers present both challenges and opportunities. Challenges include improving confidentiality and quality. Opportunities include reduced non-response rates, reduced time for publication of final census data, plus a lower spatial level of data dissemination. The advantages far outweigh the disadvantages.

IV. CONCLUSIONS AND RECOMMENDATIONS

Registers and other administrative sources are increasingly becoming a viable alternative to traditional censuses.

Several statistical offices in Arab countries, mainly GCC members, are considering using administrative registers in preparation for the 2020 population and housing census round.

Recommendations to stakeholders in the Arab region to facilitate the use of administrative registers in the 2020 census round (and later rounds) include the following:

- Review or prepare a national legislation system as a basis for the creation of population and housing registers, with permission to use the data contained in those registers for statistical purposes;

- Establish a universal personal identification (unique identity) system to facilitate proper data linkages. Identifiers must be unique, universal and stable;

- Establish and maintain a transparent partnership with administrative agencies;

- Identify, together with users, the key requirements for the census;

- Review current registers to understand strengths and gaps in order to build statistical registers from different administrative sources;

- Establish quality frameworks and assessment processes for data/metadata before using registers in censuses;

- Choose effective methodologies for data linkage, managing missing and inconsistent data and eventually linking administrative data with traditional data;

- Use new methods applied to data collection and data processing to establish organisational structures designed for register-based censuses suitable for use in different circumstances;

- Prioritise the quality of registers and the quality of data/metadata when choosing the administrative registers to produce statistics for censuses;

- Assess the quality of data sources and their impact on related output;

- Improve data quality from administrative registers by cleaning and transforming them prior for use in censuses or statistics generally;

- Distinguish between indicators describing quality of sources, quality of metadata and quality of data;

- Use the three base registers that are subject to national laws. They include the whole resident population register, the building and housing registers and the statistical business register, plus their complementary registers. Consider all possibilities to compile data from auxiliary administrative and non-administrative sources, including Big Data;

- Establish a fourth base register, a spatial register, for geocoding and localisation purposes. The register should comprise all territory division units, with geometric boundaries of administrative and statistics units plus *x,y* coordinates of address points;

- Implement GIS to enable linkage of statistical data with spatial data in order to provide complete geographical coverage and conceptual detail of all variables, whether available in registers or not;

- Understand and publish accurate descriptions of all administrative variables;

- Communicate to all stakeholders that statistics must not enable individual entities to be identified publicly.

Bibliography

Dygaszewicz, Janusz (2011). Combining data sources from census, surveys and administrative sources in Poland. Paper presented to the Directors General of the National Statistical Institutes Conference. Wiesbaden, Germany, September.

_____ (2012). Modern census in Poland. Paper presented to the United Nations International Seminar on Population and Housing Censuses: Beyond the 2010 Round. Seoul, November.

Economic and Social Commission for Western Asia (ESCWA) (2013). *A study of Age Reporting in Selected Arab Censuses of Population*. E/ESCWA/SD/2013/5.

United Nations Economic Commission for Europe (UNECE) (2007). *Register-based Statistics in the Nordic Countries: Review of Best Practices with Focus on Population and Social Statistics*. Geneva. Available from www.unece.org/fileadmin/DAM/stats/publications/Register_based_statistics_in_Nordic_countries.pdf.

_____ (2011). *Using Administrative and Secondary Sources for Official Statistics: A Handbook of Principles and Practices*. New York; Geneva.

_____ (2015). *Conference of European Statisticians Recommendations for the 2020 Censuses of Population and Housing*. New York; Geneva. ECE/CES/41.

_____ (2016). *Outline for UNECE Guidelines on the Use of Registers and Administrative Data for Population and Housing Censuses: Note by the Task Force on Register-Based and Combined Censuses*. Geneva. ECE/CES/ GE.41/2016/21.

United Nations Statistics Division (2008). *Principles and Recommendations for Population and Housing Censuses*, Revision 2. ST/ESA/STAT/SER.M/67/Rev.2.

_____ (2013). *Overview of National Experiences for Population and Housing Censuses of the 2010 Round*. New York. Available from https://unstats.un.org/unsd/censuskb20/KnowledgebaseArticle10706.aspx.

_____ (2017). *Principles and Recommendations for Population and Housing Censuses*, Revision 3. ST/ESA/STAT/SER.M/67/Rev.3.

Valente, Paolo (2015). From the 2010 to the 2020 census round in the UNECE region – plans by countries on census methodology and technology. Working Paper, No. 24. Geneva: Economic Commission for Europe. ECE/CES/GE.41/2015/WP.24.