



Economic Commission for Europe
Conference of European Statisticians
Group of Experts on Population and Housing Censuses**Twenty-first Meeting**

Geneva, 18–20 September 2019

Item 2 of the provisional agenda

**Results of tests with regard to methodology, technology,
participation, and other aspects****Review of results of tests carried out to improve the quality of
population and housing census****Note by the Central Statistics Office of Poland*****Summary*

As part of the preparations for the census in the 2020 census round, a number of organizational, methodological and technological tests were carried out in Poland aimed at developing the best solutions for use in the census, which will be implemented in 2021. The following were taken into account:

- (a) Optimization of the data collection methods and tools used;
- (b) Amount of resources and funds (including financial) both on the part of census executors from the side of official statistics, as well as entities obliged to provide census data;
- (c) User-friendliness of the applied solutions for users;
- (d) Communication and technological efficiency of designed information technology (IT) applications;
- (e) Safety of collected and processed census data;
- (f) Optimization of deadlines for the implementation of individual stages of collecting data from respondents;
- (g) Completeness and quality of the census results;
- (h) Optimization, transparency and efficiency of the popularization message targeted at census entities.

Each of these was considered in terms of maximum simplification of the method of census implementation while maintaining high quality of obtained results. Test results and conclusions are presented in this document.

* Reissued for technical reasons on 14 August 2019.

** Prepared by Janusz Dygaszewicz, Magdalena and Janczur-Knappek.



I. Introduction

1. The population and housing census is a statistical survey of exceptional scale and importance. It is a survey incomparable with any other statistical survey, because it concerns the entire population of the country, and its purpose is to collect and develop complete information on the condition and socio-demographic structure of the population and housing resources of the country. It is implemented every 10 years. During the intervening time, both the methodology of censuses and the methodology of collecting data from respondents evolve. The market of technological tools and information and communications technology (ICT) systems, which significantly influence the organization of data collection methods and the participation of respondents in the census, change dynamically. A widening circle of possibilities to direct information messages to respondents, including via digital tools, helps to shape pro-social and “pro-statistics” attitudes (affecting the public image of official statistics).

2. Taking into account the aforementioned arguments, it can be assumed that in the population and housing censuses over the years there are constant changes that can improve the quality of censuses. Only the principle of preserving the comparability of results and continuing the creation of time series remains unchanged.

II. Analysis of the situation

3. Ever since the census of the population has been 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 has been enriched. Over the decades of census implementation, the information scope of censuses has evolved, thanks to which there is now an international set of data to be collected. This set is modified if necessary by a team of international experts working in task forces and working groups. From the respondent's point of view, however, the reduction of the burden resulting from the obligation to participate in the census has so far been less significant. But this has now changed. 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.

4. In countries with developed administrative registers, the benefits of using data in registers to replace data obtained directly from respondents were noticed. This has been a gradual process carried out over years, starting from the partial elimination of some data by collecting them from registers and then filling in the remaining data with direct interviews, and leading up to the case of the full census being carried out exclusively on the basis of registers. The necessary condition is 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. The effect of such an activity in the countries implementing the census purely on the basis of administrative sources is the complete elimination of the respondents' participation, i.e. the abolition of the obligation to participate in the census. Thus, an additional, very significant benefit was achieved by minimizing the costs of the census – another driving force for changes in the implementation of censuses.

5. Countries that do not have registers that can be used for census purposes or whose registers are at such a stage of development that they do not guarantee full substitution of census data, must look for other ways to limit the content of electronic census (e-census) forms, and thus reduce the number of questions to result in a smaller burden on respondents. It should always be remembered that efforts to support respondents must not affect the quality of the census.

III. Observation of census conditions in Poland

6. Administrative registers operating in Poland are not yet of such quality that the results of the census can be based only on them. However, their improvement is constantly observed. Therefore, it was already decided in 2011 to partly use data from registers and to complete the missing data with information obtained from respondents. To be sure about the quality of the census results, all the census questions were included in the census form, even those which had coverage in the administrative sources. The purpose of this activity was to show the respondents what information about them was contained in the registers (in the census form entered in the form of Master Record (MR)). Their obligation was 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.

7. After the census in 2011, 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 always carried out within the framework of the official statistics system 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 was strengthened.

8. In the 2021 census, the system described in paragraph 6 will be maintained, but in a modified form. This will involve comparing data collected from registers with data collected from respondents. However, due to General Data Protection Regulation (GDPR) regulations, data from registers will not be presented to respondents in the electronic form which they fill out. As a result of the continuous process of improving data quality in administrative sources, it will be finally possible to eliminate questions containing data from registers in subsequent censuses, and then to completely abandon the need to involve respondents.

9. In recent years, the use of alternative sources, including Big Data, has become increasingly important. Experimental work is being carried out on the availability and quality of these sources as well as the areas of Big Data application for the population and housing census.

10. Regardless of the use of registers and alternative sources, work is underway on the modernization of data collection from respondents, including the use of methods based on ICT systems.

IV. Identifying areas of change

A. Methodology for building tools for collecting data with “CAxI” channels: CAWI (Computer-Assisted Web Interview); CATI (Computer-Assisted Telephone Interview); CAPI (Computer-Assisted Personal Interview)

11. The methodology for building data collection tools entails matching surveys to the data collection mode. This includes consideration of the length of questionnaire, blocks of questions, order of questions, filtering questions, and functionality.

12. Implementation of the census using CAxI methods depends on the possibility of:

(a) The respondent independently filling in the census form, where the respondent will only be given a personal view of the content of the questions and possible answer options;

(b) Tips given by the census enumerator during the telephone interview, where the respondent will not see the content of questions or possible variants of the answer;

(c) Prompts and explanations provided by the census enumerator during the direct interview.

13. The electronic census form should be a universal tool, adjusted to the intellectual abilities of the average respondent and to the basic skills of using a computer and Internet applications.

14. There are several main aspects to look out for when creating an electronic form (e-form) for the census. These are:

(a) The formulation of questions in the e-form: questions should be arranged in a transparent and legible way for the person filling in the questionnaire. They should also be formulated in a simple and understandable way so that the respondent knows exactly what is being asked. Otherwise it may happen that the respondent, not knowing what the question is about, will not give a correct answer or will skip the question;

(b) The length of the e-form: an excessive number of questions should be avoided in the e-form. It is unlikely that the respondent would reliably complete, for example, a twenty-page document. A too-large e-form will tend to discourage respondents from completing it. The e-form should be possible to fill out in about 15 minutes, with the number of questions limited to the necessary minimum - an average of 30 questions;

(c) Blocks of questions: it is recommended to put questions into thematic blocks due to the fact that they are then a compact and logical whole, and the e-form is orderly, which makes it more communicative. Each block should start with a general question and then go to more detailed ones;

(d) Order of questions: questions in the questionnaire should be arranged in the right order. General questions should precede detailed questions. The simpler questions should appear before the more difficult ones;

(e) Filtering: filtering questions are very useful, as they allow elimination of topics that do not concern specific people. They make it easier for the respondent to fill in the questionnaire by avoiding questions that do not concern him/her;

(f) Hints, explanations and dictionaries: the hints and explanations used in the survey enable the respondent to get acquainted with the given issue and facilitate the selection of the right category, thus improving the quality of the results obtained;

(g) Functionality: the layout and graphical presentation of the questionnaire should facilitate the provision of responses. The e-form should be clear, eye-catching and should encourage respondents to take part in the survey.

15. For the purpose of tests to improve the quality of the census, the following assumptions were made:

(a) Data are collected through mixed interview techniques, i.e. CAWI, CATI, CAPI;

(b) The target group is the full population;

(c) The information content of the survey satisfies national and Eurostat needs;

(d) The information range is the same in all data collection techniques;

(e) The optimal length of the e-form: it is possible to fill it in within 15 minutes, and the number of questions on the e-census is limited to the necessary minimum (average 30);

(f) Questions are grouped into thematic blocks;

(g) Good, attractive graphics are used in the e-forms, above all in the CAWI technique;

(h) Filtering questions: data are supplemented in the fields indicated for completion by the smart e-form when using CAWI techniques;

(i) Entering data into active fields indicated for completion by a smart e-form or block/question/person selection by an interviewer using the CATI and CAPI techniques;

(j) In all CAxI techniques, especially in CATI and CAPI, the possibility of departing from the service path provided in the e-form is ensured with the possibility of automatic return to the interview point (the possibility of supplementing the selected information during the interview in the previous thematic blocks);

(k) Questions in the e-form are formulated in a way that is understandable to the average respondent (not requiring additional explanations);

(l) Names and definitions used in the e-census form are readable and understandable for each respondent (not requiring additional explanations);

(m) Dictionaries and hints are used, showing up after a sequence of characters is entered, in those questions for which this is possible and where it will ensure that good quality data are obtained;

(n) E-form uses data from administrative sources (for example: building, dwelling, registered persons).

1. The impact of new tools to collect data on the quality of results

16. CAxI techniques used to collect data have a huge impact on the quality of results. Regardless of the channel in which the census is conducted, the user-friendliness and communication of the e-form increases the respondents' interest and the willingness to participate in the census. The unambiguity of questions avoids mistakes resulting from misunderstanding of questions. Clear answers to clearly defined questions also have an influence on the number of errors that have to be corrected during data processing, as well as on the rate at which results are obtained, developed and made available. The preparation of the census results is the culmination of the entire process of census implementation. Therefore, it should be done in a perfect and fast manner. The use of modern methods of data collection significantly speeds up their processing and development. Dissemination of census data in the form of microaggregates, aggregates and result sets should be implemented in an innovative way, by granting access to the analytical database. People who, for various reasons, use the results of censuses, are more likely to be interested in them, the more their curiosity is stimulated through the user-friendly preparation of an e-census form. The information obtained by CAxI methods allow for conducting a variety of analyses (including spatial analysis) and enables recipients to create individual reports. These reports will also be better, the more precise and unambiguous the wording used in the census form.

(a) *The influence of data collection methods*

17. This section looks at the how the different data collection methods influence the response indicators for key variables and the quality of input data.

18. When assessing the influence of data collection methods, i.e. the mixed techniques for interviewing, on the response indicators for key variables and on the quality of input data, the following assumptions were made:

(a) The data collection mode affects:

(i) The quality of input data: type of errors, scale of errors, accuracy, completeness, timeliness of data collection;

(ii) The cost of data collection: on the part of the producer, on the part of the respondent;

(iii) The organization of data collection;

(iv) The design of the e-form;

(v) the level of stability of data collection;

(b) A compromise between the quality of collected data and the costs of data collection is necessary.

19. In the assessment of the influence of data collection methods - mixed techniques for interviewing, on the key response indicators the following issues are relevant:

- (a) Determining which data collection modes can be used to obtain the combination of a high response rate with a low cost of obtaining data (collection of data via the Internet, CAWI, cannot be the only mode of data collection due to the low response rate)
- (b) Determining the sequence of data collection modes
- (c) Matching the mode of data collection to the various subpopulations planned to be covered by the census.

B. Methods for measuring the influence of data collection methods on the quality of input data and methods for mitigating the undesirable impacts of data collection modes

20. The following assumptions were made when developing the methods for measuring the influence of data collection methods on the quality of input data:

- (a) An assessment is made of the impact of data collection methods on the quality of input data:
 - (i) Taking into account the standard method determined in the Quality Framework (point C);
 - (ii) With an established reference point for comparisons (e.g. to data obtained in traditional surveys using statistical methods);
 - (iii) With a determination of the correction methods of errors (e.g. in the case of sample selection);
- (b) It should be possible to quantify the impact of data collection modes on their quality. That is, within each mode the number and type of errors made for analogous questions should be determined.

21. When developing methods to mitigate the undesirable impact of data collection modes, it is first necessary to review all the factors that can be eliminated regardless of the data collection mode. These are:

- (a) “communicativeness” of the e-form (independent of the data collection mode). A lack of communicativeness of the e-form is equally burdensome across all modes of data collection. Regardless of the data collection channel (and even if the questions are formulated in an unambiguous manner), where necessary, an e-form should include hints that will help the respondent or enumerator to fill in the e-form, as well as to shorten the length of time needed for completion. In so doing, the quality of the data will be automatically increased;
- (b) efficiency of the mobile application (e-form). In each of the modes of data collection the functionality of the mobile application is very important, allowing for quick transition between questions, controlling the logical correctness of the response, validation of responses, displaying error messages, etc. The application’s user-friendliness affects the quality of the data and the attitude of the person completing the e-form regardless of whether it is a census enumerator, a statistical interviewer or the respondent him/herself;
- (c) attributes of the ICT environment in which data collection takes place. For individual channels, the ICT environment should be constructed in such a way that all operations carried out with the e-form (most often in the background, i.e. invisible to the user) take place in the shortest possible time and in the most understandable, intuitive way. Problems with handling a mobile application or waiting for the opening or closing of an e-form, the appearance of further questions, etc. cause a discouragement of the user and the desire to abandon the activities performed.

22. After completing the recommendations from points a) to c), it will surely be found that most of the causes that reduce the quality of the data have been eliminated. There will remain reasons that are related to the person filling in the e-form, e.g. the effect of the enumerator, or reasons that cannot be predicted, e.g. time pressure, indisposition at the time

of completing the e-form (stress, illness, etc.). If these effects can be lessened by the communicativeness of the e-form, it is possible to be properly prepared for unforeseen situations.

1. Strengths and weaknesses of multi-channel data collection

23. Multi-channel data collection consists of simultaneously obtaining data in the census by various channels, where the simultaneity in this case means the time frame of the data collection phase in the census, and not a reference to a specific point in time (specific date). Therefore, various ways of organizing the activation of individual channels can be adopted:

(a) Simultaneous start at the beginning of the census and closing at the end. This means the simultaneous activity of all data collection channels. Such a data collection organization is effective only when the respondents are very well profiled, so that they can be assigned to the right channel with a high probability. There must also be a very good motivation of the respondents to complete the self-enumeration, otherwise it may happen that many people refuse the telephone or direct interview justifying by choosing self-enumeration, and as a result they will not do it and there will be no time to send them the enumerator again, which will result in an incomplete census. This method of channel activation is not recommended, as it does not result in high quality of data;

(b) Sequential launching of individual channels without duplication. This happens when the channels are completely separated by time, which means that only one channel is active at a time. The last one to close should be the CAPI channel, since this one offers the highest control over the respondent. This is a very good method of channel activation, conducive to the high completeness of the census. However, it has the disadvantage that the duration of individual channels should be extended, meaning the length of the entire data collection period is longer than in other organizational variants. As a consequence, although it has no negative impact on the quality of data (indeed an improvement of quality may occur), the costs of the census increase;

(c) Sequential launching of channels with allowed partial overlapping of activity. It is important to preserve the order of activation of individual channels and the duration of their exclusivity. It is particularly important that the first channel launched should be the online self-enumeration, as it is the most economical and permits respondents unrestricted ability to give information about themselves (in the case of sensitive questions), especially in the case of introverts. It is also important that in the initial period of data collection, the CAWI channel should be the only way to implement the census. It is possible for the remaining channels to be open in parallel, if a set of telephone calls to the respondents is available and the persons from this set can be assigned to the CATI method. Thus, census enumerators can be sent to other respondents. If the census has been well popularized, including the CAWI channel, it can be assumed with a high probability that the majority of people will perform a self-enumeration in the first two weeks of the census. With this assumption, after two weeks, it is advisable to activate further channels. While running them simultaneously there is no need to extend the deadline of the census. This will allow for optimal management of the financial resources allocated to the census. It is also necessary to remember about the earliest deactivation of the self-enumeration so as not to block the census with other methods. The scenario of activation of the data collection channels described in this point is the most advantageous due to the quality and completeness of the data and the profitability of the census.

24. Multi-channel data collection also gives the respondent the ability to choose the most suitable option for providing information about him/herself. Each of the modes has specific advantages as well as disadvantages. However, the advantages and disadvantages are not the same for all users. It is important that everyone chooses a channel that is most comfortable for them. The following describes the features of data collection channels, which may determine or influence the decision of the respondent to choose a channel:

(a) *Online self-enumeration.* The respondent has a designated framework date in which s/he should fill in the e-form. S/he also has a mobile application at his/her disposal, with which self-enumeration runs quickly and without difficulty. The respondent should choose the right moment to work with the application, regardless of the time of day or

night. Problems will arise if it turns out that the application has technical or substantive defects that make it difficult to work with. If the application works properly, an additional advantage is a sense of “anonymity”, which is missing if the respondent needs to provide information to the enumerator, i.e. to a stranger, which some respondents would like to avoid. In addition, the online self-enumeration is an application that can, after proper preparation, enable people with functional disabilities to participate in the census, for example people who are deaf, mute or blind. This mode can also be chosen by people who have concerns about letting a stranger (enumerator) into their home for security reasons;

(b) *Telephone interview.* This is a convenient method of taking part in the census for people who do not tolerate the presence of strangers in their home (enumerators), and who do not have the opportunity to perform a self-enumeration. Their possible concerns about the identity of the statistical interviewer can be dispelled by putting at their disposal an Infoline, where all information on the census will be provided. The telephone interview, with its lack of visual contact with the interlocutor, is for some people a barrier (because they need physical contact), and for others an advantage, because they remain in some way unrecognizable. A prerequisite for the success of the telephone interview method is the preparation of an e-form so that its completion does not take longer than 15 minutes. Extending the interview results in tiredness and discouragement of the respondent;

(c) *Direct interview with an enumerator.* This method is especially recommended for digitally excluded people or lonely people, for whom contact with another person, even a stranger (enumerator), is valuable. It is also a good method for respondents who find it difficult to mobilize themselves to conduct the self-enumeration: for the enumerator they must provide information regardless of whether or not the moment in which the enumerator came is convenient for the respondent.

2. Strengths and weaknesses of multi-channel data collection: analysis

25. Multi-channel data collection undoubtedly both has a lot of advantages and is fraught with disadvantages. It is based on the desire to make the census more user-friendly and less burdensome for respondents. A very strong point in favour of multi-channelling based on ICT solutions is guaranteed improvement of the quality of data through algorithms that check the correctness on regular basis (validation when filling in the e-form). Immediate verification of data also creates the possibility of much faster development of census results than, for example, using traditional writing methods, i.e. paper forms. Collecting data through various ICT solutions reduces the costs of the census, at least until it becomes possible to perform the census only using registers or by online self-enumeration.

26. The weakness of multichannel data collection is the need to build a costly IT system to enable the collection of data in various ICT environments, and then the combination of data acquired in different channels. Nevertheless, expenditures for the construction of an IT system are borne once, and then can be used not only to implement censuses, but also to conduct current surveys.

C. Quality Framework for multi-channel data collection

27. The Quality Framework for multi-channel data collection should provide a standard method for assessing the quality of data collected by various channels and in particular should take into account the following aspects of quality:

- (a) Accuracy;
- (b) Timeliness;
- (c) The cost of collecting data;
- (d) The burden on respondents.

28. In the Quality Frameworks adopted for application in the 2021 census in Poland, a compromise between the aforementioned quality aspects should be obtained. The selection of the most important aspects and their hierarchy should be made during the design phase

of the census. According to the generic statistical business process model (GSBPM), the design of the data collection methodology begins with the identification of users' needs. The needs of users and their priorities are diverse. The persons responsible for designing the data collection take into account the identified needs of users, in particular in terms of accuracy and timeliness, taking into account the limitations including costs, respondent burden, technologies and innovative methods of data collection. An important issue is the impact of increasing rates of non-response on the precision of estimating data based on the sample, and ultimately the accuracy of the resulting data. The collection mode affects types of errors (coverage, measurement, non-response) and the size of errors. Therefore, quality tests and evaluation should be conducted separately for each channel.

29. The Quality Framework for multi-channel data collection determines the quality requirements and is necessary for quality management as part of the multi-channel data collection management system. The Quality Framework covers the design phase of the census and the data collection phase and should contain the following for each channel:

- (a) Challenges;
- (b) Quality design: goals, quality dimensions, error types, threshold values;
- (c) Standards, quality requirements;
- (d) Quality control tools: evaluation and measurement;
- (e) Quality control techniques;
- (f) Methods of correcting errors, deviations from accepted norms;
- (g) Standardized and systematic multi-stage control of the quality of collected data: quality monitoring (checklists, quality reports, information on errors);
- (h) Assessment of quality, analysis and interpretation of the value of indicators;
- (i) Effectiveness of data collection processes;
- (ii) The effectiveness of actions taken after errors;
- (iii) Comparison of quality indicators for data collected by various channels;
- (iv) Assessment of the impact of data collection methods on the quality of collected data.

1. Evaluation of the quality of multi-channel data collection

30. For the possibility of evaluating the quality of multi-channel data collection, there is a need to build a multi-channel data collection checklist covering quality dimensions, indicators and their threshold values.

31. The checklist for multi-channel data collection should contain threshold values for individual indicator values. As the mixed data collection method is used in the 2021 census, it is necessary to determine minimum quality standards for each channel. The threshold value will play the role of the minimum quality standard. The established standard should be used in the data collection assessment process. Standards should be known, identified at the design stage, data collection stage, including e-form design and organization of the entire process. Based on the Quality Framework, a quality assessment questionnaire should be designed taking into account established standards for all factors that affect the quality of collected data. In the questionnaire, the requirements should be ranked by their importance. In the process of monitoring the quality of multi-channel data collection, risks should be taken into account and the consequences of non-compliance with established quality standards should be determined.

D. Data collection management system

32. The purpose of the multichannel data collection management system is:

- (a) Collecting data with the desired quality, within the set time and budget available;
- (b) Ensuring high efficiency of data collection processes;
- (c) Orientation to respondents' needs;
- (d) Focus on challenges, including innovation in data collection.

33. Issues related to data collection must be dealt with the due care, and management must take place in a wider context. When developing a multi-channel data collection management system, planning, organization and monitoring processes should be taken into account.

34. As part of the planning process, the following important issues were identified:

- (a) Challenges:
 - (i) Methodological;
 - (ii) Technical; implementation of new technologies;
 - (iii) Organizational;
- (b) A way to respond to challenges; tasks to be carried out;
- (c) Risks and reactions to them;
- (d) Principles of data collection: quality, real costs and costs perceived by respondents; building an atmosphere of cooperation and trust in the relationships between statistical producers and respondents;
- (e) Resource planning;
- (f) Assumptions for building a repository, for collecting metadata about the data collection process. Information is needed to monitor and solve problems. The information is also necessary to assess the course of work in accordance with the adopted schedules and assumptions. For this there is a need for a well-functioning system.

35. As part of the organization process, the following important issues were identified:

- (a) Staff involved in the data collection process. Organizational structure of units involved in the data collection process:
 - (i) Roles;
 - (ii) Tasks;
 - (iii) Prioritization of tasks;
- (b) Communication with enumerators and personnel involved in the data collection process;
- (c) Types and sequences of data collection techniques;
- (d) Tools for collecting data: e-forms;
- (e) Collecting data from external administrative systems, intended for use in the census, as a data source;
- (f) Supplying the e-form with data from external administration systems;
- (g) Development of address/housing/person lists for the census;
- (h) Rules for collecting data by individual techniques;
- (i) Field operations management system;
- (j) Rules for the selection of enumerators;
- (k) Training and preparation of auxiliary materials for training;
- (l) Standardization of methods and methods of implementation;
- (m) Methods for eliminating errors, insufficient coverage;

- (n) Strategy of reminders and actions in the situation of low response rates; the approach to take in the situation of lack of response or lack of contact with the respondent;
- (o) Data collection activities:
 - (i) Number and content of reminders about the census, method of dissemination of information: telephone, e-mail, letter (traditional mail), visit;
 - (ii) Schedule of visits, interviews between the enumerator/interviewer and the respondent (number, date);
 - (iii) The number of re-visits, contact attempts, in the situation of lack of response during the planned collection period;
 - (iv) Methods for dealing with respondents who are difficult to make contact with;
 - (v) The treatment of refusals to provide information;
 - (vi) Communication with respondents;
 - (vii) Duration of the interview;
 - (viii) Time of data collection by each channel;
- (p) ICT technical infrastructure supporting the data collection process and adapted to the management of data collection, management of enumerators and interviewers, monitoring, reporting, data collecting;
- (q) Monitoring of data collection.

36. As part of the data collection monitoring process, the following key issues were identified:

- (a) Monitoring techniques (including tracking enumerators route using GIS tools);
- (b) Monitoring tools and reports:
 - (i) Types of reports: on the results of work (quality reports according to channels), on the course of work (indicators of completion of interviewers' work);
 - (ii) The structure of reports;
- (c) Analysis of reports and evaluation of data collection processes;
- (d) Problem solving: procedures for dealing with specific situations, problem solving methods, optimal solutions;
- (e) Corrective actions;
- (f) Evaluation of the effects of corrective actions.

E. New solutions for collecting data using new techniques resulting from the development of ICT

37. The purpose of developing new organizational solutions for data collection based on CAxI methods is:

- (a) Defining the conditions under which it will be possible to effectively use modern ICT solutions to collect data;
- (b) Defining all risks that could potentially have a negative impact on the result of data collection;
- (c) Defining a minimum threshold for achieving census results, which can be considered sufficient to accept as correct;
- (d) Specification of any corrective actions that should be implemented immediately in the event of a threat preventing the continuation or completion of the census as planned.

38. Furthermore, when planning an organization, it is important to take into account:

(a) *Sequence of activities and their duration*

The development of implementation schedules (framework and detailed one) is one of the basic preparatory activities for the census, regardless of the method used to implement the data collection stage. However, when using multi-channel data collection based on CAxI methods, the schedule is much more complex because it has to take into account all interdependencies within and between individual actions (regarding strictly data collection channels) undertaken within the census. The framework schedule captures in general all the milestones of the census and the main tasks occurring within them. It gives a general view of the course of the census, which allows preparation of the sequence of activities to be carried out in a smooth, time-optimized way that eliminates excessive accumulation of tasks or unnecessary downtime. Detailed schedules permit planning of all tasks and subtasks inside milestones. In all schedules it is also necessary to specify task performers, especially if the structure of the team of people involved in the census is complex;

(b) *Assigning activities to performers in a way that does not raise any doubts*

The use of modern technological solutions for collecting data in the census requires the preparation of many processes (often also composed of subprocesses) in a comprehensive manner that cover individual tasks to be performed at each stage of census work. The more complex and diverse the technology used, the more processes there are to handle. Each of the processes must have an owner responsible for its implementation. It must also have performers. To optimize the implementation of tasks, a clear division of responsibilities and tasks must be made. Moreover, detailed procedures should be developed for use in standard situations as well as in the event of unusual situations. The procedures must include a description of the situation and the individual steps to be taken to handle the incident, together with the assignment of roles and responsibilities;

(c) *Defined personal and material resources*

The great role in designing a data collection organization based on modern IT technologies entails - just as with every method of census implementation - accurately identifying the number of people needed to be involved (including distinguishing roles and assigning numbers of people to them), the quantity of IT equipment to ensure efficient census implementation and calculation of the necessary financial resources. However, what distinguishes a modern technological census from a traditional one is the maintenance of appropriate proportions between expenditures related to the construction of data collection channels and other costs. This is particularly important when initializing new methods, also taking into account the use of IT infrastructure acquired in the census – which can be used after the census to carry out current statistical surveys. In the next census, where data collection channels are already built, the proportion of expenditures changes again, because at this point it is no longer about building channels, but only about their development related to technological progress in the statistical environment;

(d) *Strengthening the competences of people planned to be involved at every stage of the established organization*

Changing data collection technologies requires the acquisition of new skills and competences. Therefore, it is necessary to design a series of training for people involved in the census, with the training being dedicated to the type of roles in the census. It should also be remembered that the training should end with verification of the degree to which the participants have assimilated the new knowledge, because the efficiency of using the new technology will condition the success of the census implementation.

1. Description of new organizational solutions

39. *Determining the probability of effectiveness and the scope of methods.* When designing new technological solutions to be used in data collection, especially if it is a multi-mode project, the effectiveness of individual channels should be assessed. This is the starting point for designing further organizational work. Each data collection channel is subject to some specific restrictions that must be taken into account. These are restrictions such as:

- (a) For CAWI: Internet access, digital exclusion;

(b) For CATI: length of the census form (number of questions), access to the telephone database;

(c) For CAPI: respondents inhabiting closed, protected housing estates; difficulty in finding a respondent at home due to work at different times of the day, etc.

Taking the above into account and a number of other factors, whether socio-technical, cultural or related to the condition of households depending on the inhabited region of the country, the expected return of answers via the online self-enumeration should be estimated. The selection of CATI and CAPI channels, in turn, will be more influenced by financial conditions (the amount of money in the census budget) and the number of people eligible to be used as telephone interviewers, if telephone interviews are carried out using statistical office resources.

40. *Profiling of respondents for pre-assignment to the channel.* This aims to support the estimation of the effectiveness of data collection within individual modes, as well as the development of an effective method of reaching respondents with properly prepared census promotion. Groups of respondents who are most likely to choose an online enumeration should be selected. In this case, information provided by information sources on the use of IT services, age group, profession etc. may provide clues. Next, households should be selected to be reached by telephone where there is little chance that they will choose self-enumeration (e.g. elderly, chronically ill, etc.). The remaining persons will be enumerated by census enumerators, although profiling can already indicate these people through the elimination of CAWI and CATI channels.

41. *Acquiring data.* One of the key actions supporting data collection using CAWI methods is to obtain relevant data, e.g. telephone numbers or e-mail addresses. If such action is not generally permitted by law, efforts should be made to ensure a legal exemption specifically for the census, taking into account the importance of the census and its usefulness for the broad and important recipients of the census results.

42. *Inventory of actions that can be performed by the statistical office's own staff and the scope of outsourcing.* Using multi-mode data collection requires having employees with appropriate competences, or else outsourcing all or part of the tasks. There is no doubt that entrusting the census in its entirety to an outside contractor would deprive public statistics producers of the possibility of acquiring knowledge, skills and experience. It is also more difficult to control the course of individual processes and to maintain statistical confidentiality, as well as to create or maintain the public image of official statistics. This is why it is worth performing at least part of census tasks in-house, in order to be able to become completely independent of external contractors in the future. An important task is to determine the proportion of workmanship. Moreover, a necessary condition, regardless of the scale of outsourcing, is the maintenance of responsibility for the course and results of the census and the control and management function in the census within the framework of public statistics.

43. *Designing the IT Census System (ICS).* Technologically advanced ICT solutions, especially if more than one data collection channel is used, require the development and construction of an appropriate IT environment. Often these are separated subsystems that must be linked into a single, coherent system so that there is an absolute guarantee of receiving complete, unique and excellent quality census results. Therefore, it is worth paying attention not only to the correct operation of individual subsystems or applications, but also to communication within the ICS. A reliably designed ICS allows comprehensive support of all census activities, as well as their automation where possible. Within designated limits (on the basis of assumptions built into the system) it is possible to perform part of the tasks manually or to correct effects resulting from automatic actions (e.g. allocation of addresses to enumerators in the CAPI method or change of data collection channel for a given address during the census).

44. *The order of activation of data collection channels.* This is one of the strategic decisions conditioning specific organizational activities. With the knowledge about the impact of data collection methods on the quality of results and the expected range of individual channels and their assumed effectiveness, it is possible to determine the optimal date of activity of each channel with high probability. What is more, having the profiled

groups of respondents at the point of directing them to the appropriate data collection channel, the sequence of the methods should be determined and the overlapping (partial or total) or separation of individual modes should be determined. Assuming that the second, most desirable effect of data collection – apart from high quality results as the main effect – is to obtain a high response rate in the CAWI, the online self-enumeration should be the starting channel and in the initial period it should be exclusive.

45. *Internet access points for respondents.* For respondents willing to enumerate via the Internet, and not having access to the Internet, so-called access points in places widely available should be created, with guarantees for the safety of enumeration both for respondents (here in the sense of preserving their privacy) and for their data (maintaining statistical confidentiality). The best places for this are public administration offices (community offices, statistical offices), institutions and centres providing social assistance, post offices, schools, etc. This is an activity that can significantly increase the rate of responses submitted by the Internet.

46. *Designing management, monitoring, control and reporting activities.* Collecting data with many channels, additionally complicated by the fact that during the census it is possible to change the data collection channel for the same respondent (in case if the original channel is ineffective) requires:

(a) Establishing management structures for channels in which tasks are allocated, i.e. for CATI and CAPI;

(b) Close cooperation of dispatchers in both channels in order to exchange information on a current basis;

(c) Having appropriate tools to control the flow of e-forms between channels;

(d) Having appropriate tools supporting management, monitoring and control. The current tracking of the course of the census in individual channels allows identification of undesirable situations and their immediate elimination;

(e) Having the appropriate reporting tools to track both the progress of data collection by channel and in total, and the current verification of threat coefficients for census implementation at voivodship and country level.

47. *Environment of data collection: promotion.* A special task during the implementation of the census is the dissemination of knowledge about the census and encouraging participation in it. Skilful promotion can “impose” on the respondents the channel of data collection most preferred by the organizers of the census. However, special care should be taken in the selection of appropriate media and its content for the targeted groups of recipients. It is equally important to develop an information campaign in which the terms of popularization relevant to the census work phase and the intensity of the campaign will be preserved. The point is for the respondent to feel encouraged by popularization, not tired of its intensity, which can have the opposite effect.

48. *Anticipation of activities accompanying the collection of data resulting from the needs of respondents.* Experience shows that even the most perfectly prepared organization of census work does not exhaust the range of events that may arise during the data collection stage in the field. That is why it is a good idea to anticipate solutions even for unnecessary cases, if such actions do not generate excessive financial expenses. One such solution is to create a possibility of enumerating a respondent in CATI in response to his/her phonecall made to the Infoline number. This may be a situation where the respondent, due to their type of work, family situation, planned long journey, etc. will not be able to fulfil the obligation at any other moment of census. In this case, the CATI variant should be used, which consists in redirecting the call from the Infoline to the CATI channel.

2. Impact of new organizational solutions on the quality of results

49. At each stage of organizational activities, regardless of the method of census implementation, their effects should be checked and reacted to if it is found that this influence is undesirable. This is particularly important in the case of new methods, such as CAWI methods of data collection.

50. CAxI methods of data collection may affect the quality of the census results indirectly through the impact on respondents. However, the quality design of the multi-channel data collection described in point C: “Quality framework for multichannel data collection” is not taken into consideration. It is about the potential reception by respondents of the new data collection organization. People tend to be distrustful of change. Public perception is one of the determinants of the success of the undertaking, which is why it is so important to pay close attention to the respondents’ reactions. They can be expressed by:

- (a) Response indicators in individual channels;
- (b) Feedback provided on the Infoline;
- (c) Responding to interviewers’ telephones and visits of census enumerators;
- (d) Frequency of visits to places where computers are made available for self-enumeration;
- (e) The nature of press articles and other media widely available and used;
- (f) Positive or negative attitude to official statistics.

51. Each of the above factors should be carefully analyzed in order to formulate conclusions and develop strategies to maintain correct trends or counteract negative situations. The role of teams responsible for popularizing the census in consultation with organizational teams is particularly important.

V. Conclusion

52. In the age of the information society, it seems that it would be a mistake not to take advantage of the opportunity that modern technologies offer. Nevertheless, further experimental work should be carried out to improve the quality of the census.

53. The ICT world is developing very fast. What is new today will be obsolete in a few years. However, constant work on improving the quality of the census will allow for the creation of patterns and models that can be used to develop the quality of censuses in the future, when new tools and data sources for the implementation of censuses will appear.
