

Supporting effective use of information and communication technology in
population census

Practical-Advisory Workshop
Moscow, April 16-19 апреля, 2012

MATERIAL ON THE REPUBLIC OF ARMENIA

The latest Population Census in the Republic of Armenia has been conducted in October 2011. The phases that have already been accomplished and the phases that should be implemented after the data entry are described in this report.

Phase 3 – Build

3.1 Build data collection instrument

The interview method was used for the 2011 population census data collection. In order to realize this method, a special questionnaire was built, that was filled out by the instructed enumerators. The questionnaire was tested during the pilot population census conducted in October 2010.

Connection of metadata and data at the point of data entry was assured by the data description dictionary that was built by the CPro software.

The Division of planning and program-methodological works and the Division of information technology and programming (within the Department of population census) were responsible for this task.

The questionnaire was build using the software CoreIDRAW and was ready two months prior to the commencement of the population census. This questionnaire is used exclusively for the population census.

3.2 Build or enhance process components

The software CPro was used to process the 2001 population census data. It proved its value at almost all stages of the data processing, except for the final output generation. However, in the new version of CPro all downsides have been eliminated and the output tables can be received now in the required formats. Therefore, it was decided to use the same software CPro in 2011. All applications are ready and already used in the data processing.

The division of the information technology and computer programming is responsible for this sub-process.

3.3 Configure workflows

The census activities can be divided into 3 major stages:

Preparatory activities

Conducting the census

- Preliminary visitation
- Data collection
- Post-census sample control visitation

Data acceptance and processing

- data acceptance
- coding
- data entry
- obtaining the outputs

The sequence of these operations was defined and realized in accordance with the strict schedule.

All five divisions of the Department of population census are responsible for the process flow as a whole.

3.4 Test production system

The testing of the computer system and applications commenced yet before conducting the pilot census. Basic testing of the new version of CSPro (CSPro 4.1) and applications, built on CSPro, occurred during the pilot census. Immediately prior to the general census, the hardware and network (the main instruments for the data entry) technical testing had been done.

Responsible for this sub-process was the Division of the information technology and computer programming.

3.5 Test statistical business process

In 2010, the pilot census was conducted in Armenia, during which all the population census phases were tested. The data collection was performed from 1% of population. The major purpose of testing was to identify the errors in order to take measures to eliminate those errors prior to 2011 census.

All the divisions of the Population census department actively participated in conducting the pilot census.

3.6 Finalize production systems

For each phase of the population census, the documents and instructions were prepared and finalized at least one month prior to the launch of a given phase. Before the launch of each phase, the training was provided for the production personnel.

Phase 4. Collect data

4.1 Select sample

In Armenia the population census is the only survey that covers all the country; therefore, for the population census, the sample selection is not performed.

4.2 Set up collection

This sub-process includes a number of interconnected activities. To organize the works on population census, in the first place, the availability of the schematic maps of cities and large villages was assured, as well as and to the extent possible, the street names and the numeration of blocks and buildings were regularized, numbering and name plates were installed. The next step was the registration of houses. Then, temporary personnel were hired. After the preparatory cartographical works have been completed in general, and the questionnaire was sent to be printed, the special trainings started, for which the Division of planning and program-methodological works and the Organizational division are responsible.

Learning activities were accompanied by the practical filling out the questionnaires and other census documents.

The population was periodically informed regarding the preparation and conduction of the population census through the press, radio, TV and also text messaging.

The clause on collected data confidentiality was included into the contracts with the temporary personnel.

Protection of the collected data is guaranteed by the population census law.

Responsible for this sub-process were the Division of planning and program-methodological works and the Organizational division of the Population census department.

4.3 Run collection

The process of population census in the Republic of Armenia includes the following stages:

1. preliminary visitation
2. conducting the census
3. control visitation

During the preliminary visitation the enumerator became familiar with and clarified the boundaries of his/her area, visited all the houses and buildings in his/her enumeration area, informed the dwellers on the conducted population census, explained, if necessary, the purposes and tasks of its conduction and clarified the appropriate time to visit the given household to enumerate its members.

Population census was conducted from 12 to 21 October, 2011 using the interview method and filling out the questionnaire.

All population of the republic was registered as for the well-defined point of time – midnight from 11th to 12th of October.

As the result of the population census RA 2011, the data on two categories of population will be received: permanent population (permanently present and temporary absent) and present population (permanently and temporary present).

After the completion of census enumeration, the instructors-controllers together with the enumerators performed sample control visitation in their areas. This control visitation was performed in all enumeration areas, covering in each of them 10% units (houses).

Responsible for this task was the division of planning and program-methodological works of the Population census department.

4.4 Finalize collection

After the portfolios acceptance and registration, the process of the data entry from the paper questionnaire was launched. This year the data entry from the keyboard was also used. The duration of this process is 6 months. Responsible for this process is the Information technologies and programming division.

Phase 5 – Process

5.1. Integrate data.

On the reason that in Armenia the census data are collected from one source, there is no necessity to integrate data collected from different sources. The data integration is performed after the data coding and data entry. As the result of data entry, the text files are created, each of them corresponds to one enumeration area. After the data entry completion, all the files are combined, and the database is created for further cleaning and analysis. The data integration can be performed also before the data entry completion for the processing software testing. For the data entry as well as integration the software CPro is used. Duration of the data integration is very few seconds. Responsible for this sub-process is the Division of information technology and programming of the Population census department. The sub-process of data integration is also used in other processes of data production.

5.2. Classify and code

The works on classification and coding were launched in parallel with the information acceptance at the Department of census. Each portfolio corresponds to one enumeration area; therefore, each portfolio was assigned with the corresponding geographical code, i.e. during the information acceptance the accompanying document was filled in. Then the data of census enumeration form are classified and coded. From 33 individual questions, 7 are text questions that needs to be coded accordingly. Four manuals for coding were produced:

- Classification of administrative-territorial division of the RA
- Classification of countries

-Classification of economic activities (based on the international EC classification NACE)

-Classification of occupation (based on the international classification ISCO)

The main difficulty rises when the question “occupation” is coded. Because of the non-standard replies on this question, we selected manual coding again. If the code is not in line with the dictionary, during the data entry the operator enters the text of reply that is processed further by the appropriate subject matter experts.

Duration of all data coding takes 6 months. Responsible for this sub-process is the Division of data processing and classification. The sub-process of classification and coding is used also in other processes of data production.

5.3. Review, validate and edit

Editing the population census database is performed exclusively using the application for automated editing. The application program is written in CSpPro. It verifies the presence of the responses on questions, established boundaries and also logical connection between the enumeration form questions. In addition to the identifying the errors, the program of automated editing performs the function of errors correction.

The program is made in such a way that in parallel with the errors correction the messages on errors and their percentage is generated.

To correct the data, the program of automated editing is used after the data integration. However, to test the program itself, it is also used before the data integration.

The data correction process takes about a month; however, after the output data is produced, sometimes there is a necessity for additional data editing.

Responsible for this sub-process is the Division of information technology and programming of the Population census department. The sub-process of data editing is used also in other processes of data production.

5.4. Impute

As our data control program not only checks, but also correct the data, it can be said that the program function is a kind of imputation. But in addition to the actions, based on logical verifications, the well-known imputation methods are used, such as “hot-deck” or random values.

The age of the household member is imputed by method “hot-deck”. The age value is taken from the data cube, in which the age is given under the certain values of sex and relationship. This data cube is constantly updated, if the sex, age or relationship of the next member is within the normal limits.

The sex of the respondent is imputed using the random values in case if the other methods cannot be applied.

5.5. Derive new variables and statistical units

The database description vocabulary consists of variables that correspond to the questions of the census form. In order to derive the new variables that are required for the output data, the new variables are added to the database description vocabulary. The automated editing application program, in addition to the data correction, creates the values for these variables.

Creation of the variable on respondent dwelling category is an example of the derivation of the new variable. The respondent belongs to the permanent population, i.e. the new variable is equal to 1, if he/she permanently lives on a given address, or if he/she is absent but the duration of absence is less than 12 months.

Another new variable is derived by the data aggregation and is based on the newly created variable. This is number of [household] members which permanently live in the household.

Responsible for this sub-process is the Division of information technology and programming of the Population census department.

The sub-process is used also in other processes of data production.

5.6. Calculate weights

We do not use weights during the data processing as the census is an all-country event. However, if the relevant decision on conducting the post-census survey to verify the census quality will be made, the methodology for statistical data processing and weights to extrapolate the survey data on general population will be developed.

5.7. Calculate aggregates

The aggregates for the entire republic are produced after the completion of the data entry. In general, these data are related to the population number for permanent and present population – as for the entire republic and also for the different geographical levels. The data aggregates are produced also before the completion of data entry to define the characteristics of concrete administrative-territorial units.

The process itself runs very fast, but the analysis of results, corrections and re-aggregation can take a lot of time.

Responsible for this sub-process is the Division of information technology and programming.

The sub-process is used also in other processes of data production.

5.8. Finalize data files

The database file, produced in sub-process of data integration, after the sub-processes of data processing, becomes the database. However, this is not the final file as it can be changed in the process of data analysis.

This sub-process is used also in other processes of data production.

Phase 6 – Analyze

6.1. Prepare draft outputs

Apart from the variables that are created during data editing, the additional indicators are produced when the output results are produced (for example, the number of nuclear families in the household). These indicators facilitate the process of data analysis.

The output data are produced on the basis of the maquettes designed in advance. This sub-process takes several months.

Responsible for this sub-process is the Division of information technology and programming of the Population census department.

The sub-process is used also in other processes of data production.

6.2. Validate outputs

The produced output data are validated from the side of all interested organizations. On the first priority basis, the response percentage on different questions in accordance with the required levels is validated. The logical connection between the data, based on which statistics is produced, is also validated.

For dissemination of 2011 population census output results, the layouts of about 70 tables were prepared yet on the stage of design. When the data file for one marz (administrative-territorial unit) is ready, the works on verification and interpretation of results will commence.

For some indicators, the results are compared with the results of other processes of data production. For example, the population size is compared with the results of the current statistics that is produced by the Demography division of the National Statistical Services of RA.

As this process is implemented in parallel with the results production, it can take several months.

Responsible for this sub-process are: the Division of the data processing and classification, the Division of methodology, the Division of information technology and programming of the Population census Department, and also other divisions concerned.

The sub-process is used also in other processes of data production.

6.3. Scrutinize and explain

The output data are produced in such a way that all the users, in particular, statisticians, understand the substance of these results.

In turn, the statisticians analyze and evaluate statistics before the final result is produced.

6.4. Apply disclosure control

Prior to the data dissemination, the compliance with the established confidentiality rules is verified. In principle, when these rules are violated, the data are closed. However, if there are a lot of such violations in the table, it is necessary to decline this table or change the format of the table.

Responsible for this sub-process are: the Division of the data processing and classification and the Division of information technology and programming of the Population census department.

6.5. Finalize outputs

This sub-process finalizes all activities on preparation of final statistics – the logical control results and preparation of metadata are finalized, the final result is discussed with the appropriate subject matter experts and, finally, the State Committee on Statistics approves the statistical content to publish.

Phase 7 – Disseminate

7.1. Update output systems

The data and metadata formatting is performed in advance.

During this sub-process the data are loaded into output database, and also the data readiness for dissemination is checked and updated.

Responsible for this sub-process is the Division of the data processing and classification of the population census department.

The process can take several months. It is used also in other data production processes.

7.2. Produce dissemination products

For dissemination of 2011 population census output results, the printed publications and websites will be used.

All typical steps of the dissemination products production will be performed. Namely: preparation of tables, charts, maps, editing the above mentioned products and formatting them in accordance with the publishing standards.

Responsible for this sub-process are: the Division of the data processing and classification and the Division of information technology and programming of the Population census department.

7.3. Manage release of dissemination products

The scale of the activities, related to this sub-process, is not known yet, because it depends on financing.

7.4. Promote dissemination products

The websites, blogs, etc. will be used to promote the statistical products, but the usage of other instruments will depend on financing.

7.5. Manage user support

The users' queries always were registered and reviewed in national statistical service of Armenia. As the result, because of these queries, the changes in the enumeration form and, accordingly, in the output data take place. For example, as a result of a numerous queries on the subject of agriculture, four questions, related to the agriculture have been added to the 2011 census enumeration form.