

NATIONAL INSTITUTE OF STATISTICS

Quality Report of the 2011 Population and Housing Census

Bucharest

CONTENT

1. OVERVIEW	3
1.1. Legal background	3
1.2. Bodies responsible	4
2. DATA SOURCES	5
2.1. Classification of the data sources according to Article 4(1) of Regulation (EC) No. 763/2008	5
2.2. List of the data sources used for the 2011 census	5
2.3. Matrix „Data sources x Topics”	7
2.4. Extent to which the data sources meet the essential features [article 4 (4) of Regulation (EC) no. 763/2008]	8
2.4.1. Individual enumeration	8
2.4.2. Simultaneity	8
2.4.3. Universality within the defined territory	8
2.4.4. Availability of small-area data	9
2.4.5. Defined periodicity	9
3. CENSUS LIFE CYCLE	9
3.1. Reference date according to Article 5(1) of Regulation (EC) No 763/2008	9
3.2. Preparation and execution of data collection	10
3.2.1. Questionnaire-based data	10
3.2.1.1. Design and testing of questionnaires	10
3.2.1.2. Preparation of any address lists, preparation of the field work, mapping, publicity	11
3.2.1.3. Data collection	12
3.3. Processing and evaluation	13
3.3.1. Data processing	13
3.3.2. Quality and coverage assessment, post-enumeration survey(s), final data validation	18
3.4. Dissemination	22
3.5. Measures to ensure cost effectiveness	24

1. OVERVIEW

1.1. Legal background

The legal basis for conducting the Population and Housing Census is provided by:

- the Conference of European Statisticians – Recommendations for the 2010 Round of Population and Housing Censuses;
- Law No. 677/2001 on the protection of individuals with regard to the processing of personal data and the free movement of such data;
- Regulation (EC) No. 763/2008 of the European Parliament and of the Council of 9 July 2008 on population and housing censuses;
- Commission Regulation (EC) No. 1201/2009 of 30 November 2009 implementing Regulation (EC) No. 763/2008 of the European Parliament and of the Council on population and housing censuses as regards the technical specifications of the topics and of their breakdowns;
- Commission Regulation (EU) No. 519/2010 of 16 June 2010 adopting the program of the statistical data and of the metadata for population and housing censuses provided by Regulation (EC) No. 763/2008 of the European Parliament and of the Council;
- Commission Regulation (EU) No. 1151/2010 of 8 December 2010 implementing Regulation (EC) No. 763/2008 of the European Parliament and of the Council on population and housing censuses, as regards the modalities and structure of the quality reports and the technical format for data transmission;
- Government Ordinance No. 36/2007 on the conducting of the 2011 Population and Housing Census in Romania approved with amendments by Law No. 5/2008;
- Government Decision No. 1502/2009 on the organization and conduct of the 2011 Population and Housing Census in Romania;
- Government Decision No. 1251/2010 amending Government Decision No. 1502/2009 on the organization and conduct of the 2011 Population and Housing Census in Romania;
- Government Decision No. 1324/2010 amending Government Decision No. 1502/2009 on the organization and conduct of the 2011 Population and Housing Census in Romania;
- Government Decision No. 922/2011 amending Government Decision No. 1502/2009 on the organization and conduct of the 2011 Population and Housing Census in Romania;
- Emergency Ordinance No. 34/2011 amending and supplementing Government Ordinance No. 36/2007 on the conducting of the 2011 Population and Housing Census in Romania.
- Law No. 170/2011 approving Government Emergency Ordinance No. 34/2011 amending and supplementing the Government Ordinance No. 36/2007 on the conducting of the 2011 Population and Housing Census in Romania;
- NIS rules on statistical confidentiality;
- Confidentiality agreements.

1.2. Bodies responsible

Pursuant to Emergency Ordinance No. 36/2007, the body responsible with the preparation, conduct, coordination and monitoring of the work needed for carrying out the 2011 Population and Housing Census was the **Central Commission for Population and Housing Census**. The main responsibilities of this Commission were as follows:

- a) to approve the overall Census organization program, schedule and methodology for data recording and processing, as well as the schedule for publishing the results;
- b) to make recommendations following the consultations undertaken to develop the draft Government Decision on the Census reference moment, the data collection period, the Census budget, the organizational structure and staffing necessary for the activities program to carry out the Census and process the Census results;
- c) to approve the registration forms and instructions for filling thereof;
- d) to approve the classifications and nomenclatures required for the registration and processing of Census data, except those approved by Government decisions;
- e) to approve the instructions and the organization manner and to guide the actions and operations to divide the territory of the administrative-territorial units into Census sectors;
- f) to coordinate the recruitment and training of all personnel involved in conducting the Census works;
- g) to approve the Census publicity program;
- h) to present to the Government and publish the Census preliminary results.

Pursuant to Article 5 of Law No. 5/2008, “The organization and carrying out of the Census were performed in technical and methodological terms by the National Institute of Statistics, under the coordination and control of the Population and Housing Census’s Central Commission, according to the Recommendations of the European Union, and the United Nations Economic Commission for Europe rules in the field of Population and Housing Census.”.

The Commission for Population and Housing Census consisted of:

- Chairperson: - Ministry of Interior and Administrative Reform
- Vice Presidents: - President of the National Institute of Statistics
 - State Secretary for the Ministry of Interior and Administrative Reform
 - State Secretary for the Ministry of Development, Public Works and Housing
 - State Secretary for the Ministry of Economy and Finance
 - State Secretary for the Ministry of Labor, Family and Equal Opportunities
- Secretary: - Vice-President of the National Institute of Statistics
- Members: - State Secretary for the Ministry of Education, Research and Youth
 - State Secretary for the Ministry of Defense
 - State Secretary for the Ministry of Foreign Affairs
 - State Secretary for the Ministry of Public Health
 - State Secretary for the Ministry of Justice
 - State Secretary for the Ministry of Culture and Religious Affairs
 - State Secretary for the Government General Secretariat
 - State Secretary for the Interethnic Relations Department
 - representative of the Romanian Academy
 - representative of the local public authorities’ associative structures
 - president of National Agency of Roma

For managing all methodological and technical activities a Technical Secretariat of the Central Commission for Population and Housing Census was established in the National Institute of Statistics.

2. DATA SOURCES

2.1. Classification of the data sources according to Article 4(1) of Regulation (EC) No. 763/2008

2011 Population and Housing Census in Romania was planned as a **conventional census**. This was the only solution as the administrative sources were not available in the period this census was planned.

Once a part of the administrative data sources became available (in the beginning of 2012), it was possible to use them for two main reasons:

- data reconciliation
- record imputation (indirect data collection).

The whole territory of Romania was divided into census districts, sections and sectors.

Between 20 - 31 of October 2011 around 100¹ thousands enumerators (one enumerator per a census sector) visited the census sectors and made individual “face-to-face” interviews with all persons having usual residence in that specific geographical area. Data about households (private or institutional), dwellings and buildings (residential or not-residential in which at least one person had usual residence) were also recorded in the Census paper questionnaires.

After data collection period the data entry, error corrections (including record deletion, total and partial imputation) and data editing were done.

On the same database other data corrections, record and item imputation were made. Therefore, **we had a single data source** for 2011 Population and Housing Census in Romania used as the main base for the enumeration of the census population, which will be called the “census microdata database”.

2.2. List of the data sources used for the 2011 census

All people it should have to be interviewed by an enumerator during data collection period. Depending on each individual situation, people should be classified in an unique category and his/her personal information should be fulfilled on a specific type of questionnaire. The four categories are:

- present person (P) or
- temporarily absent person (TA) or
- temporarily present person (TP) or
- person left for a long period of time (PPI)

Present person was considered the person (irrespective of citizenship and country of birth) who have lived in their place of usual residence in a specific geographical area inside Romania for a continuous period of at least 12 months before the census reference date or he/she arrived in their place of usual residence in a specific geographical area inside Romania during the 12 months before the census reference date with the intention to stay there for at least one year.

In order to be considered as present, a person should be present in the census reference date (or he/she was at the job, school, hospital and so on, but he/she would come home at the end of the day/week).

A P questionnaire was fulfilled for each homeless people enumerated. They were considered present in their original household or in a collective unit quarters were the enumerator found them.

¹ The exact number of enumerators was 103,429 persons.

Temporarily absent person was considered the person (irrespective of citizenship and country of birth) who have lived in their place of usual residence in a specific geographical area inside Romania for a continuous period of at least 12 months before the census reference date or he/she arrived in their place of usual residence in a specific geographical area inside Romania during the 12 months before the census reference date with the intention to stay there for at least one year.

The persons included in this category were not present at the census reference date at his/her usual resident address, but their absence was less than 12 months or with the intention to be absent for less than 12 months. The absence could have several reasons and the person could be inside or outside Romanian territory: for work, job related reasons, holiday, medical treatment or health care.

Temporarily present person was considered the person (irrespective of citizenship and country of birth) which, at the census reference date, he/she had came in the enumerating place for a period of time smaller than 12 months or having the intention to stay for less than 12 months in a private household or in a collective unit (hotel, hospital, hostel and so on) for different reasons: holiday, job related reasons, visiting the relatives or friends, health care or medical treatment.

Person left for a long period of time was considered the person (irrespective of citizenship and country of birth) which, at the census reference date, was left from the locality where he/she had the usual residence in another locality inside or outside Romania, for a period of at least 12 months or he/she left the locality by less than 12 months, but he/she had the intention to be absent for at least 12 months.

The students studying in another locality were recorded as persons left for a long period of in the parents' household and as present persons in the collective living quarter (or in a renting dwelling) in the locality where they were studying at the census reference date.

Individual information for present and temporarily absent persons were fulfilled in P type questionnaire.

Individual information for temporarily present persons were fulfilled in TP type questionnaire.

Individual information for person left for a long period of time were fulfilled in PPI type questionnaire.

In the usual resident population were included only present and temporarily absent persons.

Our data source, the "census micro-data database" was organized on components, each component having the structure of a specific Census questionnaire. The link between the components is a unique code, relate to the census sector and records hierarchy inside the census sector.

So, the components of "census micro-data database" are:

- (1) P file, containing individual records for all persons having usual residence in Romania for at least 12 months at the Census reference date;
- (2) TP file, containing records for all temporarily present persons;
- (3) PPI file, containing records for all persons left for a long period of time;
- (4) M file, containing data for all persons 15 years old who were not present in the Census reference date in Romania, being abroad for less than 12 months or more than 12 months. Information from this file is related with information from P file or PPI file;
- (5) LC file, containing records for all dwellings (conventional and non-conventional) and information about the building in which the dwelling is situated;
- (6) G file, containing records for all private households;
- (7) SC file, containing records for all collective living quarters and institutional households.

Collective living quarters consisted of one or more distinct buildings located at the same address and intended for living for different time periods by relatively large groups of people generally unrelated by kinship, which usually gathered for common purposes and interests and could share certain places and facilities (bathrooms, kitchens, dining rooms, etc.)

For people living in collective living quarters a P or PPI individual questionnaire was fulfilled, depending on each person's situation.

After the end of all statistical actions of data correction and data completion, the assembly of all records contained in P, TP and PPI files represents all enumerated people.

2.3. Matrix „Data sources x Topics”

	Data source	Topics	
		Non-derived topics	Derived topics
1	Population P – Person PPI - Person left for a long period of time TP - Temporarily present person M - Person 15 years and over living abroad	<ul style="list-style-type: none"> - Place of usual residence - Sex - Age - Legal marital status - Country/Place of birth - Country of citizenship - Previous place of usual residence and date of arrival in the current place; or place of usual residence one year prior to the census - Relationships between household members - Location of place of work - Current activity status - Occupation - Industry (branch of economic activity) - Status in employment - Educational attainment - Ever resided abroad and year of arrival in the country (from 1980) - Tenure status of households - Ethnic group - Religion - Mother tongue - Number of live-birth children - Year of marriage (first one and actual one) - Year of consensual union's formation - Domicile - Internet use by individuals - Work time (total number of hour and hours worked in the main activity) - Work sector - Social protection benefits (for unemployed) - Difficulties in daily activities 	<ul style="list-style-type: none"> - Total population - Locality
	G – Household questionnaire		<ul style="list-style-type: none"> - Household status - Family status - Type of family nucleus - Size of family nucleus - Type of private household - Size of private household
2	Housing LC – Dwelling,	<ul style="list-style-type: none"> - Type of living quarters - Location of living quarters - Occupancy status of conventional dwellings 	<ul style="list-style-type: none"> - Density standard

	Building SC – Collective Living Quarter	<ul style="list-style-type: none"> - Number of occupants - Useful floor space and/or number of rooms of housing units - Housing arrangements - Type of ownership - Water supply system - Toilet facilities - Bathing facilities - Type of heating - Dwellings by type of building - Dwellings by period of construction - Dwelling position in the building - Number of rooms (living rooms and rooms used only for professional purposes) - Thermic rehabilitation of dwellings - Material of external walls of building - The place where the building is located 	
--	--	--	--

2.4. Extent to which the data sources meet the essential features [article 4 (4) of Regulation (EC) no. 763/2008]

The census micro-data database meets the essential features of the Census, defined by Regulation no. 763/2008 on population and dwelling censuses. The reasons are described below.

2.4.1. Individual enumeration

The information about each enumerated persons was obtained by personal interview (or proxy interview with a person who knew the situation of a specific person), and, for the under-coverage, by indirect data collection from administrative data sources (record imputation). The administrative data source used included individual records about persons for which the personal interviews were not done during the data collection period. For the direct data collection, information obtained by “face-to-face” interviews were recorded on paper questionnaires and included later in census micro-data database. For the records obtained by indirect data collection from administrative data sources, the individual characteristics were taking over directly in the census micro-data database, using the link on individual level by personal numeric code.

2.4.2. Simultaneity

Information obtained on individuals and housing in the census refers to the same point in time, namely the Census reference date. This was the midnight of the 20-th October 2011.

For under-coverage, the records found in administrative data sources were imputed in census micro-data database only if it was determinate from several administrative data sources the 12 months continuously presence on the Romania’s territory.

2.4.3. Universality within the defined territory

The 2011 population and dwelling census provide data about all statistical units (persons, dwellings, buildings, family nuclei) within the territory of Romania.

The usual resident population measured by the 2011 population and dwelling census included all persons that have their usual residence in Romania, irrespective other characteristics (legal residence, citizenship, country of birth), at the Census reference date.

The indirect data collection was based on the principle of usual residence in the sense that a record missing from the initial census database and present in an administrative data source was included in the census micro-data database **only if it was enough evidence** that person had the usual residence for at least 12 months inside Romania around the census reference date.

2.4.4. Availability of small-area data

The data are available for all small geographic areas and for small groups of statistical units.

Due to the necessity to assure the data quality, the data are disseminated at the level of territorial - administrative units – counties, towns, communes, villages inside communes and census districts².

2.4.5. Defined periodicity

The population and housing census in Romania was a long historical tradition. Romania carried-out population censuses in 1899, 1912, 1930, 1941, 1948, 1956, 1966, 1977, 1992, 2002 and 2011. For the last 3 decades, the population census was done in the beginning of each decade.

The population and housing census will continue to be carried-out regularly in the future in a new form, adapted to the specific conditions at that time being. For the sake of comparability, we hope that specific country economic conditions will allow to carry-out the census at the beginning of every decade.

3. CENSUS LIFE CYCLE

3.1. Reference date according to Article 5(1) of Regulation (EC) No 763/2008

The 2011 population and housing census reference date was the midnight of the 20-th October 2011.

The main milestones for 2011 population and housing census in Romania were:

Activity	Period/ Ending date
Setting-up the specific census related legislation	2007-2012
Setting up the Central Commission for Population and Housing Census	Beginning of 2008
The approval of Census General Budget by the Romanian Government	December 2009 (the budget was available in the beginning of 2011). It was supplemented in 2012.
Completion of census materials (questionnaires, manual, nomenclatures and classifications, other useful materials) and approval by the Central Commission for Population and Housing Census	July – August 2011
Census simulation	7-16 May 2011
Data collection period for 2011 Population and Housing Census	20-31 October 2011
The post-enumeration survey	3-14 November 2011
Data entry for summarization questionnaires	December 2011 – January 2012

² A census district contains around 60 census sectors.

Data entry for individual questionnaires	December 2011 – February 2013
Insurance of data completeness (indirect data collection)	March- June 2013
Dissemination of provisional results	2-th February 2012
Dissemination of preliminary results	24-th August 2012
Dissemination of definitive results, by:	
- Press release	4-th July 2013
- paper definitive results presentation (4 thematic volumes)	last quarter of 2013
- hypercube web application	31-th March 2014
Transmission of 2011 Census Quality Report to Eurostat	31-th March 2014

3.2. Preparation and execution of data collection

3.2.1. Questionnaire-based data

We used the following types of census questionnaires:

A. QUESTIONNAIRES FOR REGISTRATION

1. **LC** – Dwelling, Building
2. **G** – Household
3. **P** – Person (individual form)
4. **PPI** – Person left for a long period of time
5. **TP** – Temporal present person
6. **M** – Person 15 years and over living abroad
7. **SC** – Colective living quarters

B. SUMARIZATION QUESTIONNAIRES

1. **CL** – Sumarization table for census sector's buildings and dwellings
2. **CP** – Sumarization table for persons from census sector
3. **CELR** – Sumarization table for population according to ethnicity, mother tongue and religion from the census sector
4. **CM** – Sumarization table for persons living abroad

C. AUXILIARY QUESTIONNAIRES

1. **PR** – Fact sheet of census staff
2. **BM** – Folder bordereau which comprises the composition of census sector and folder contents

3.2.1.1. Design and testing of questionnaires

The design of 2011 census questionnaires took into consideration the following main aspects:

- a. the necessity to collect all topics included in the Annex of Regulation no.763/2008 to be collect in the Population and Housing Censuses;

- b. the need to cover without overlap all population, i.e. the four population categories described at point 2.2.;
- c. obligation to collect all core characteristics/ variables described in Appendix 1: List of proposed core and non-core topics for the 2010 round of population and housing censuses from the United Nations Economic Commission for Europe and Eurostat Recommendations for the 2010 censuses of population and housing;
- d. the need to ensure international comparability by using international classifications of occupation, industry, country of citizenship/birth, international standard classification of education and so on;
- e. possibility to collect some of non-core topics described in Appendix 1 of the above document;
- f. the need to ensure the time series comparability between population censuses;
- g. the necessity to provide preliminary results as soon as possible after data collection;
- h. the need to capture the specificity and special characteristics of Romanian population;
- i. national data requests from:
 - **National Bank of Romania** – request to include in the census questionnaires a special module addressed to persons 15 years or more who leave abroad for more or less than 12 months; the module was focused on the remittances from abroad and social and demographic characteristics of people living abroad.
 - **League for Rights' Promotion and Defense of Persons with disabilities in Romania**– request to collect data on disability accordingly to International Classification of Functioning Disability and Health. The census questionnaire included a minimum set of 6 questions elaborated by the Washington Group on Disability Statistics.
 - **Non-governmental organizations for promoting the Roma rights** had a great implication in Census preparations, in order to obtain a good participation of this ethnic group in Census activities.
 - **Ministry of Regional Development and Tourism** – request for data describing the number and characteristics of thermic rehabilitated dwellings.
- k. other specific issues related to a survey addressed to population: convenient length of the interview, easy to understand questions, not sensitive questions and others.

3.2.1.2. *Preparation of any address lists, preparation of the field work, mapping, publicity*

The National Agency for Cadastre and Land Registration was in charge with preparation of address lists and delimitation of census sectors. For this issue, all Romania's territory was divided in 103.429 census sectors, for each of them existing a list of all addresses which should be visited by enumerators in order to identify all dwellings (conventional or other type) and buildings in which at least one person has his/her usual residence.

The address lists and the content of each census sector were obtained automatically, based on the information available at each locality's level.

Due to some lack in the available information, the offered address lists were updated by the counties' technical secretariats staff by visiting the assigned census sectors before the start of data collection period.

Finally, each enumerator had a list of addresses which he/she should visit in order to collect census data.

3.2.1.3. *Data collection*

Preliminary visit: On 19-th October 2011 each enumerator had to visit all territory of his/her census sector in order to:

- know the addresses which should be visited;
- identify all buildings and finding if they have living destination or they are firms, enterprises, schools and so on;
- identifying all dwellings in which is possible to live somebody, especially those who hadn't living destination;
- identifying all buildings, including abandoned, un-occupied buildings and unfinished buildings;
- distribute the publicity materials and people announcement about census's launching;
- make a planning of the visits during the data collection period in order to have enough time to carry-out all the visits;
- finding buildings/dwellings at the sector's limit which can be double included in two separate sectors or building/dwellings not included at all anywhere.

Visits during the data collection period: in the period 20-31 October 2011 each enumerator had to visit all addresses from his/her census sector in order to make individual interviews with all persons and to record in Census specific questionnaire the characteristics of dwellings and buildings.

The enumerators had the obligation to visit an address as many times as it was necessary to found the household / persons living there. Proxy interviews were allowed only in case of persons who were difficult to be found.

The rhythm of interviews was monitored using daily sms. Each enumerator should report each morning of the data collection days the number of persons and dwellings enumerated and recorded in the previous day. The monitoring was done at counties and national level.

During the last days of the second data collection week, analyzing the rhythm of registration, some supplementary measures were adopted, namely:

- each City hall organized a special place where persons not visited yet by an enumerator could come in order to be interviewed;
- City halls in big cities defined a special e-mail address for receiving Census self-fulfilled questionnaires from persons not enumerated yet;
- each City hall sent supplementary enumerators in some census sectors where monitoring exercise showed low enumeration figures; in some cases, the enumerator visited some addresses together with his/her supervisor in order to convince population to accept census interviews;

During whole data collection period, two population information channels functioned, namely:

- at each City hall population can find the name of the enumerator responsible for a specific census sector, in order to verify the fake enumerators;
- at INS level a Call Center functioned, in order to offer any kind of information related to Population and Housing Census, 24 hours from 24, all data collection days and some days after.

Enumerators' activities after data collection period

Immediately after data collection period, each enumerator should carry-out three important activities.

First of them consisted in verifying the order of questionnaires, completeness and correctness of information included in questionnaires.

The second duty was to fulfill two auxiliary questionnaires and five summarization questionnaires for his/her census sector, using data recorded in each individual questionnaire. Processing these summarization questionnaires for all 103.429 census sectors it was possible to publish first preliminary results, on 2-th February 2012, three months after end of data collection.

The last main activity was to deliver all census materials (fulfilled and un-used questionnaires, manual and list of addresses, summarization questionnaires and so on) to the County's Technical Secretariat of Population and Housing Census.

Census advertising was done through the following activities:

- posters presenting the scope, objectives and data collection period of Population and Housing Census elaborated in main mother tongue present in Romania; the posters were posted at City Halls and in public places with great visibility;
- letters for households with main information related to census;
- INS's staff participation in several TV and radio broadcasts presenting the main features and useful information about organization and coverage of 2011 Population and Housing Census;
- audio and video advertisement presented on local and national TV and radio channels;
- publicity made in public transportation (buses and metro) and so on.

3.3. Processing and evaluation

3.3.1. Data processing

- Capturing, coding

Activity of data entry begun after all types of questionnaires were delivered to the County's Technical Secretariat of Population and Housing Census.

Before data entry starting, special category of census staff - the persons in charged with coding – performed the activity of codification (manual) for all variables defined as open questions in the fulfilled questionnaires, using predefined classification and nomenclatures. It was the case of variables related to: citizenship, ethnic group, mother tongue, religion, localities (usual residence, place of birth, domicile, previous residence, place of work), country of previous residence, school of highest level graduated, school attained, occupation and industry.

Data entry was done in a traditional way, using P.C. operators.

- Identifying variables

A special identification code was used for the linkage between different questionnaires. For example, for linkage between individual questionnaires of household's members and household or dwelling questionnaire, the components of this special identification code had the following meaning:

- Folder number (census sector code)
- Order number of the building within the folder
- Order number of the dwelling within the building
- Number of the LC form within the folder
- Number of the SC form under which the dwelling was recorded
- Number of G form of the dwelling

- Person's order number

For identification of missing individual records in census micro-data database and administrative sources used in the indirect data collection the personal numerical code (CNP) and persons' addresses were used.

- Record editing

For each kind of questionnaire, a special list of logical conditions was prepared and included in the data entry and data checking software. There are around 1000 logical conditions which it checked:

- presence or absence of answers at the compulsory variables;
- logical correlation between answers of several variables;
- presence of outliers and unusual cases;
- normal limits for quantitative variables;
- logical correlation between answers in different questionnaires (for example, between information regarding "de facto" or "de jure" marital status and the existence of the husband/partner in the same household) and so on.

A set of checking was done in the same time with data editing. The second set of checking was done after data editing, in the phase of data validation. These activities were done at each county's level.

When data editing and data validation were finished, a second package of operations were done at INS level, namely:

- identification of invalid personal numeric codes (CNP);
- looking for and cleaning of double records (two records for the same person) in the same file (P, TP or PPI file);
- identification of missing records;

These errors were solved at county's level too, based on lists with errors.

- Record imputation

After processing the individual forms, the under-registration found in 2011 Census's provisional results processing stage was confirmed. Therefore, methods have been applied to ensure the census data completeness, using both the method of indirect collection from administrative sources and statistical techniques for imputation of data. There were 1,183 thousand persons not enumerated in the Population and Housing Census of October 2011 (hereinafter, PHC 2011) and identified at the administrative data sources level. The administrative data sources found to have useful information for finalizing the results of 2011 Census are contained in:

- The National Register of Personal Data (RNEP) – managed by the Directorate for Persons Record and Databases Management;
- Statement on obligations to pay social security contributions and income tax, and the nominal records on insured persons - D112 – managed by the National Agency for Fiscal Administration;
- Record of Employees - IM – managed by the Labor Inspectorate;
- CNPP database – managed by the National House of Public Pensions (CNPP);
- CNAS database – managed by the National Health Insurance House (CNAS);
- Tax Registration Statement / Statement for individuals who carry out economic activities independently or liberal professions – D70 - managed by the National Agency for Fiscal Administration;

- Record of beneficiaries of state child allowance, family allowance and help the guaranteed minimum aid – managed by the National Agency for Payments and Social Inspection (ANPSI);
- Database of students enrolled in the 2011-2012 school year – managed by the Ministry of National Education.

The starting point of the indirect collection from administrative sources procedure was the comparison of records in the 2011 Census database (data recorded on the individual forms completed by the enumerators during the field collection stage) with the existing entries from the Population Register database. The set of records found in the Population Register and missing from the PHC 2011 database (recorded on all P, PPI or TP forms³) were searched in other administrative sources for the October 2011 month and the preceding and subsequent months of the same year, in accordance with the concept of usual residence in Romania for at least 12 months; this concept was applied during the 2011 PHC to measure the usual resident population. Just for persons identified in the administrative sources used, for which there was clear evidence that they were in Romania during the Census and most of 2011, a record was added (totally imputed) in the PHC 2011 database for which identifiers and variables' values were subsequently filled for the Census variables. Thus, all records obtained from indirect data collection had the same structure with the rest of records obtained from interviewing people during the field data collection period and refer to the same reference point, enabling the aggregation of information for the country's entire usual resident population, whether or not interviewed by enumerators.

The under-aged pertaining to the identified adults (mother and father) for whom information were taken from the administrative sources above, as well as for adults (mother and father) who were counted during the PHC 2011, were searched and identified in the Population Register.

Based on the information related to the address of the enumerated persons, data on dwellings were also completed and the households within those housing were reconstituted.

- Item imputation

The principle of item imputation was to use the best source for imputation and best criteria to find a good proxy for the missing information. Therefore, item imputation can be classified in two main categories:

- A. Some variables were imputed from the administrative data sources where we found it; we consider this an indirect collection of data, not an imputation, because these data are information declared by individuals, so it represent valid values, not artificial ones (as it is the case of item imputation).
- B. The second category refers to statistical imputations. Depending on type of variables, we applied the following kind of item imputation:
 - for variables referring to individuals, we used the hot-deck donor method;
 - for the qualitative variables related to dwellings, we used the method of most frequent value from the cell a specific dwelling is part of ;
 - for the quantitative variables related to dwellings, we used the method of cell's average for that cell a specific dwelling is part of ;

The administrative data sources were used “in cascade”, one after one. The information found in administrative data sources were use for item imputation of census's variables of the total imputed records.

³ P – Person (present or temporarily absent); PPI – Person left for a long period of time (inside or outside the country); TP – person temporarily present .

If the information was available in administrative data sources for the non-enumerated persons, the following variables were collected from administrative data sources by indirect data collection for imputed records: sex and date of birth derived from CNP, legal marital status, domicile, previous residence, citizenship, highest level of school graduated, educational institution the person is attending, current activity status, work time, occupation, employment status, industry and type of sector in which the person is working.

The automatic corrections applied solved the data inconsistencies, like for example:

- age - highest level of school graduated – current activity status;
- occupation – industry - type of sector in which the person is working;
- age - educational institution the person is attending - highest level of school graduated;
- age – sex - legal marital status ;
- age – number of live-born children;

Due to special character of some variables and due to non-existence of this information in the administrative data sources, we leave the answer “Not available” for the following variables:

1. in P questionnaire

- year of marriage
- year when the consensual union was constituted
- number of live-born children (only for women)
- ethnic group
- mother tongue
- religion
- difficulties in performing the daily activities

2. in LC questionnaire

- dwelling’s type of ownership
- number of rooms used only for professional purposes
- dwelling’s endowment with air conditioning installation
- year of building’s construction
- construction material for external walls of the building

• Record deletion

We delete the records considered to be duplicates (it were present in the census micro-data database more than once) of the same person/ household or dwelling.

In the step of individual data checking, the existence of several records with the same individual numerical code were found (most of the cases we found pairs of records). These were classified as:

- a. pairs of records with the same CNP enumerated within the same county and
- b. pairs of records with the same CNP enumerated in different counties

The modality of these cases was different between the two categories.

Pairs of records with the same CNP enumerated within the same county

For each pair of multiple records with the same CNP, enumerated in the same county, the staff of counties’s Technical Secretaries of Counties Commissions for Census had the obligation to identify the corresponding fulfilled paper questionnaires and compare from the correctness point of view the information fulfilled in questionnaires with the one included in the census micro-data database.

- if, by comparing the name and first name, date of birth and other information on the questionnaires, the staff concluded the two questionnaires refer to the same person (one person for which two separate questionnaires were fulfilled), then one of the two records was deleted from the census micro-data database;
- on the contrary, the conclusion of the comparison was the fact that one individual numerical code fulfilled one paper questionnaire was wrong and the other CNP one was correct. In this case, based on the information fulfilled on the questionnaires, especially the name, first name and date of birth, the correct individual numerical code was identified for the second record present in the census micro-data database;
- if the correction of CNP was not possible, this code was deleted and the record was kept in the census micro-data database.

Pairs of records with the same CNP enumerated in different counties

The pairs of records with the same CNP enumerated in different counties were firstly checked from the point of view of identity of answers on the majority of the variables. The deletion of one of the two records was done using the following criteria:

- from two records having the age greater or equal with 18 years and student as the current activity status, it was kept the record enumerated in a locality in which academic institutions were present;
- from two records having fulfilled the locality where the job was located, one of the records having the job in the same locality of enumeration and the other record in a different locality (other than the enumeration one), it was kept the record for which the job's locality and the locality of enumeration were the same;
- from two records, one fulfilled on a LC questionnaire and the other one in a SC questionnaire⁴, the record fulfilled on SC questionnaire was kept in the census database;
- from a pair of records, one associated with a present person in the enumeration locality and the other one associated with a person temporarily absent, it was kept the record associated with the present person.

Once all the persons associated with one household were deleted, the records associated with the corresponding household and dwelling were deleted, too.

• *Generation of households and families*

Starting with the individual imputed records, households were determined based on address (all persons living in the same dwelling) and family relationships.

Two different cases were taken into consideration:

1. individual imputed records associated with a dwelling already recorded in the census micro-data database – in this case the individual imputed records were included in the household already recorded in the database and the family relationships were made over;
2. individual imputed records associated with a dwelling (address) missing in the census micro-data database – in this case a household record and a dwelling record were totally imputed in the G and LC files.

The remaining individual imputed records, where no family association was found, were considered households with unrelated members and un-known type of household.

⁴ LC – Dwelling, building; SC – Collective living quarters.

3.3.2. Quality and coverage assessment, post-enumeration survey(s), final data validation

Coverage assessment

In the period between 2002 and 2011 censuses the measure of usual residence population in Romania was over-evaluated, because no data sources on emigration according to usual residence criterion were available. In this way, the usual resident population was over-evaluated due to under-evaluation of international migration (i.e. the totality of persons leaving Romania for more than 12 months, establishing usual residence abroad, but maintaining the legal residence – domicile - in Romania).

After data entry and data validation, a total of 18,938,636 individual records were present in the census micro-data database. Comparing with current population measure, the under-coverage detected in the step of provisional results was proved. Finding methods for census data's completeness was needed. This method was to look for not-enumerated records in administrative data sources, using time, usual residence and twelve months criteria.

- *Under-coverage*

The existence of **under-coverage** was detected by several means:

A. Post-enumeration survey

The main objective of the post-enumeration survey was to measure both the quality of the answers given on individuals and households, and the amount of information collected, in order to increase user confidence in the results of the census.

The control survey sought to estimate the errors that have occurred in the census data regarding:

- *the completeness* – checking all households / housing units throughout the country were enumerated;
- *coverage* - checking whether the population of the entire country was enumerated;
- *quality of responses* recorded at the Population and Housing Census – by recording into the survey forms the responses corresponding to a limited set of characteristics of those covered by the population and housing census program.

By comparing the responses recorded in the survey with the Census ones, we were able to determine the omissions and the extra entries, the frequency and the size of random or systematic observation errors, their accumulation trend in terms of certain characteristics, and the relationship of dependence between the correctness of responses and the responses source: the person concerned, another person from the household or a person outside the household.

All dwellings and individuals in residential and non-residential buildings, as well as housing units in collective living spaces (student hostels and workers' dormitories for unmarried, nursing homes, child care institutions, hotels, etc.) which also had to be recorded at the Population and Housing Census have fallen within the scope of the survey.

The following were not included: dormitories, hotel or camp-type *collective living spaces* and *people in them*, recorded on separate Census forms.

Post-enumeration survey sample

The sample was established in view of the two basic principles of selection: **maximizing representativeness** and **minimizing costs**.

Noting that, as early as the Census preparation phase, a selection base was created by dividing the entire territory of the country into Census sectors of approximately equal size, the sample of post-enumeration survey was established from that selection base, by means of a two-stage probabilistic selection: sectors were chosen in **the first stage** of the survey and the dwellings were chosen in the **second stage**. Each chosen sector was assigned to a controller.

According to the scope of the survey, the sampling base did not cover the Census sectors established in collective housing units.

Considering that surveying a number of about 32 dwelling in each census sector (one third of the total number of dwellings in a sector) allows an accurate assessment of the enumerator's activity and provided that the sample size ensures a maximum error selection of 0.3% - for errors whose frequency is lower than 2% - guaranteed with a probability of 0.95, it resulted that, in the first stage, the sample should contain 504 Census sectors (about 5% of the total number of sectors), hereinafter referred to as survey sectors.

The post-enumeration survey sample selection was done in **two stages**:

- **centrally**, at the National Institute of Statistics, by *choosing the survey sectors* from the list of Census sectors, *by means of unrepeated random selection*; all residential buildings and dwellings related were visited in the selected sectors, a minimum number of indicators being recorded to assess the Census *completeness and coverage degree*;
- **in each county**, by *selecting a third of the dwellings included in the survey*, for which essential features relating to all households and individuals that make them up were recorded in order to verify the *quality of Census data*.

Post-enumeration survey forms

To ensure high quality of results, the post-enumeration survey form included a small number of features, both for individuals and for housing units, studying only the characteristics defining the main census results, namely:

a. for households in the dwellings

- number of present persons
- number of temporarily absent persons
- number of persons left for a long period of time
- number of temporarily present persons

b. for dwelling

- number of households in the dwelling
- type of dwelling
- type of ownership
- kitchen / kitchenette inside the dwelling
- bathroom (bath / shower) in the dwelling
- water closet (WC) in the dwelling
- running water in the dwelling
- heating (including own central heating)

c. for present and temporarily absent persons from the household

- gender
- date of birth
- legal marital status
- situation of the person on the Census day
- for how many months has the person been away from the enumeration locality?
- place of birth (mother's residence at the moment of the enumerated person's birth)
- domicile
- has the person ever had a different residence than the enumeration residence?
- previous residence
- date on which the person settled in the enumeration locality
- citizenship
- what ethnic group does the person consider he/she belongs to?
- what is the person's mother tongue?
- highest educational institution from which the person graduated (receiving a certificate or a diploma) - name and profile (the master's and PhD. studies were not included)
- current activity status (performed during the reference week 13 to 19 October 2011)
- occupation in the main activity
- workplace
- main activity of the unit/subunit where the person was employed

Questions were formulated identically to those of the Census forms (including the serial number of the characteristic); the definition, the contents and the recording method remained also unchanged.

Recording method

Recording the information was performed just as it was at the Census, i.e. by "face-to-face" interviews, based on the statements of persons inside the housing unit, as much as possible, as at the Census.

Key findings of the post-enumeration survey:

By comparing the post-enumeration survey results with those obtained at the Census, it was found that 128 people have been omitted in some households – representing 3.4 ‰ of the persons registered in the survey; however, 64 extra persons were incorrectly entered in other households, that is 1.7 ‰ of all persons in the sample, resulting in a net error of omission of 64 persons (1.7 ‰) and a coverage degree of 998.3 ‰. Dwellings that have been omitted had a negative impact on the Census results, i.e. a negative error of 1.1 ‰ as against the post-enumeration survey: 0.8 ‰ in urban areas, and 1.3 ‰ in rural areas.

B. Ad-hoc registration of addresses where the enumerators found non-response

As a response to the monitoring exercise of registration rhythm during data collection period, we asked enumerators to make a summarization of the dwellings where people did not responded to interviews. The information collected was referring at:

1. occupied dwellings (the individuals were enumerated and the interview took place)
2. dwellings that changed the living destination (it has no human living destination; it is shop, firm etc.)
3. conventional dwellings where occupants' explicit refuse of participation
4. not-contacted conventional dwellings (nobody answered to enumerator's assays to contact them, but external information showed somebody is living there)

5. un-occupied conventional dwellings (the enumerator found from another person in that dwelling nobody is living).

Unfortunately, this method did not showed the entirely dimension of under-coverage because some of these dwellings where enumerators could not make interviews were already recorded as un-occupied dwellings.

C. Mass media messages during data collection period

During data collection period, mass-media showed critical aspects regarding Census. One of them was related to existence of some town's areas where no enumerators visited or people declaring that they were not visited by enumerators.

According to the Population and Housing Census Law, in the last days of data collection period, people not visited yet by enumerators had the obligation to go to the City hall to make the census registration. Mass media highlighted some cases in which big number of people visited City hall's special offices in order to be registered.

D. Comparison with Romanian population register, administrative sources and Eurostat mirror statistics

Another criteria used was the comparison of census statistics with the similar ones from the Romanian population register. Also this register uses the domicile and the existence of Romanian citizenship as criteria for persons' registration, it gave us a measure of the number of people who should be recorded in all types of questionnaires (P, TP, TA and PPI).

If all needed information would exist, the number of census population (usual resident population) would be equal with the number of records from the Romanian population register, adding the number of foreign citizens having the usual residence in Romania at the census reference date and subtracting the number of emigrants according to the usual residence criteria. This check was done and, following it, the solution for solving the under-coverage was found. More exactly, from the comparison with data from administrative data sources, a usual residence population of around 20.1 million persons would result.

The under-coverage was solved by method of indirect data collection.

A number of 1,183,005 individual records were totally imputed in census micro-data database.

The share of adults for whom data were indirectly collected from the administrative sources was: 64.9% from the D112 statement, 0.5% from the Record of Employees, 2.1% from the D070 statement, 16.1% from the Record of beneficiaries of state child allowance, family allowance and help the guaranteed minimum aid, 4.6% from the database of persons insured in the public health insurance system and 11.8% from combinations of multiple sources.

Of the total 1,183,005 persons, 47.2% are men and one fifth was less than 10 years of age at the time the Census was conducted. The degree of under-coverage was significantly higher in urban areas than in the rural ones. Of the total records indirectly collected from the administrative sources, the share of those found in cities and towns was 71.9%, and the difference (28.1%) was at the commune level.

Bucharest recorded the highest under-registration rate. Of the 1,183 thousands persons indirectly registered from the administrative sources, over 17% were found in Bucharest. Top seven counties includes: Constanta (4.8%), Iasi (4.4%), Timis (3.4%), Bacau (3.0%), Cluj, Dolj and Galati (each with 2.9%). The enumeration was conducted to a better extent in the small counties. Counties with the

lowest weights of records collected indirectly from the administrative sources were: Covasna (0.5%), Harghita (0.7%), Salaj (0.8%) and Bistrita-Nasaud (1.0%).

- ***Over-coverage***

The existence of over-coverage was measured after the data entry:

- for persons using the Personal Numeric Code (CNP) by finding more than one unique CNP in the P, TP and PPI files;
- for dwellings, using the addresses;
- for buildings, using the unique building code;

The over-coverage was solved deleting the second, third etc. record from the census micro-data database for each identification variable present more than one time in the census micro-data database.

The criterion of eliminating one from the second, third etc. record for the same individual number code were describing in the section 3.3.1. **Data processing, Record deletion.**

Final data validation

After data entry and record and item imputations, a new data validation phase was carry-out on the whole dataset (directly or indirectly collected). Errors detected were solved and the whole set of tabulation was prepared. Last errors showed by tabulation were solved and hypercubes were executed on the clean census micro-data database.

3.4. Dissemination

The 2011 Population and Housing Census results are disseminated using the following methods:

Press releases

There were 3 Press releases for the dissemination of results:

- on 2-th February 2012 for the provisional results;
- on 24-th August 2012 for the preliminary results;
- on 4-th July 2013 for the definitive results.

Each Press release was accompanied by a set of results tabulation, presenting the main results from each step, available on the census dedicated sections on the website www.recensamantromania.ro.

A Press Conference was organized on 4-th July for the presentation of first definitive results.

Tabulation

It was planned and carried-out 4 printed definitive results volumes, namely:

Volume I : *Usual resident population –Demographic structure* – 62 tables included

Volume II : *Usual resident population – Ethnic and confessional structure* - 34 tables included

Volume III : *Usual resident population – Socio-economic structure* - 58 tables included

Volume IV : *Buildings, dwellings, households* - 53 tables included

A fifth printed volume was elaborated, an English version of the final 2011 Population and housing census results, named *General results population, households, dwellings* (45 tables included).

Micro-data access for scientific purposes

Micro-data access is allowed by request, for scientific scope, based on a legal Contract for micro-data use. The contract should identify the contractant, legal framework, Contract's object, beneficiary obligations, conditions when the Contract is canceled and ways for solving the litigations.

On the National Institute of Statistics website – www.insse.ro – a dedicated web application for Census historical data is be available. The users have access at a 10 percent data sample from all census data in order to have information about the coverage and the statistics available from the last three population and housing censuses (1992, 2002 and 2011).

Special requests

After the definitive results of 2011 Population and Housing Census were obtained, several special data processing were done, as answers to requests made by a various type of users: ministries and government institutions, international organizations, researchers and students, City halls, firms or private persons and so on.

Hypercube web application

The 60 quantitative hypercubes and 21 compulsory qualitative hypercubes were prepared as Regulation no. 519/2010 asked for.

Assurance of statistical confidentiality including statistical disclosure control

The confidentiality was a big issue for the 2011 Population and Housing Census. The issue was present during several main activities of the census and regarded all categories of census staff.

The main activities where the issue of confidentiality was involved were:

- *Census preparation*

In the step of preparing the Census legislation, special articles regarding the collection and processing of individual number code in Census, ethnic group, religion and mother tongue were included, in respect of Law no. 677/2001 regarding people's protection regarding processing the data with personal character and free movement of these kind of data (article 12 from O.U.G. no. 36/2007 modified and updated).

Special specifications regarding the obligation to respect the data confidentiality were included in Staff's handbook and in supported material for staff's training. Trainers were obliged to describe in a special section of the staff, training the aspects regarding the confidentiality issues.

Each person from the census staff signed, in the same time with the work contract's signature, a special confidentiality agreement about conformation of confidentiality rules in any phase of the census: data collection, data processing or data dissemination.

- *Data collection*

During the data collection period, some measures for protection of data fulfilled in questionnaires were applied. For example, the enumerators had the obligation to deliver the fulfilled questionnaires daily to his/her supervisor.

It was forbidden to enumerators to speak about information about persons known in the context of census data collection with other persons besides supervisors who were also census's staff.

On each type of questionnaire, the following was printed:

DATA CONFIDENTIALITY

All information on this form is confidential and shall be used exclusively for statistical purposes.

- *Data entry and data processing*

The same restrictions regarding disclosure of individual information applied in the data entry and data processing phase.

A special chapter is related to the protection of data from the informatics point of view. The identification variables were kept separately, in a file with strictly restricted access.

The data entry was made by the Counties' Technical Secretariates, using an software system specifically designed for this. The system includes a client, desktop application and a server application.

The desktop application has been developed using .NET 4.0 technologies (Windows Presentation Foundation) and contained the specific user interfaces for the users to input the data collected on paper questionnaire, validate it against the predefined correlation rules and made the data correction (exclusive record and item imputation).

One particular feature of the desktop application was the ability to work offline (meaning, without connection to the server application). In this case, users can input and validate data locally, and when the connection with the server is restored to actually save them into the central database.

The server application was also developed using .NET 4.0 technologies (WebServices REST), and the main functions were: saving the data to the central DB, batch validation of the data (sets of questionnaire, re-validated), synchronise central DB with off-line work of different users. The central DB is an Oracle 11g Database Server.

- *Data dissemination*

Rules against data disclosure were applied in the dissemination phase. Generally, the printed tables used a convenient level of data disaggregation which did not allow statistical disclosure.

Confidentiality policy related to census data was the same as the confidentiality policy applied for any other statistical research by National Institute of Statistics.

We used cell suppression as main method of protection against data disclosure. The method was applied only for persons as statistical units, for all individual variables. Cells with less than 3 observations were protected at all level of disaggregation.

Another way of protection against individual data disclosure was the degree of disaggregation. For example, in tables presenting ethnic groups, religions or mother tongue, the categories with relatively small number of observations (less than 20,000 persons) were grouped in the "Other ethnic/ religion/ mother tongues" category.

3.5. Measures to ensure cost effectiveness

The census budget was established by Government Ordinance no. 1502/2009 as a total of 192,624 thousand lei. The budget was modified in august 2012 by Government Ordinance no. 847/2012, due to the need for supplementing the data entry period (data entry staff payment).

Being established by a Law, it was not possible to exceed the budget.