Economic and Social Council

Economic Commission for Europe

Conference of European Statisticians

Group of Experts on Population and Housing Censuses

Nineteenth Meeting
Geneva, 4–6 October 2017
Item 2 of the provisional agenda
Innovations in census methodology and use of new data sources

The implementation of new data sources to increase the coverage of data on housing stocks in Poland

Note by Statistical Office of Poland[1]

Summary

       The aim of this document is to present the current state of works in the area of statistics on housing stocks as well as the actions taken up in order to ensure that data requirements on the 2020 Census of Population and Housing and following censuses will be met.

       As introductory the process of establishing the Building Base (BB), its data sources and methods applied for integrating data from different data sources are presented. The BB was established as one of data sources for housing topics in the 2011 Census of Population and Housing (NSP2011). Then the works undertaken in order to improve the usability of the BB as a data source on housing stocks for future censuses are described. As a result the Building and Dwelling Base (BDB) is going to be created. Finally, the concept of the Statistical Register of Population as a data source for future censuses as well as its correspondence with the BDB are mentioned.

---

[1] Prepared by Agnieszka Matulska-Bachura, Deputy Director of Trade and Service Department

Please recycle

Contents

I. Introduction

1. In Poland, the census of housing stocks constitutes the integral part of population and housing census. Census information enable to characterize the dwellings and buildings, in which dwellings are located as well as estimate the housing needs resulting from both actual lack of dwellings and low standards of existing housing stocks. The census usually covers the following core housing-related topics:

(a) The type of living quarter;

(b) The characteristics of dwellings including: occupancy status of dwellings, type of ownership, size of dwellings (the number of rooms as well as usable area of dwellings), technical and sanitary system (fittings);

(c) The characteristics of buildings with inhabited dwelling premises, including: the type of building, type of ownership, the number of dwellings in the building, and period of construction; and

(d) Information on tenure status.

2. In Poland, the method of conducting the census has been changed comparing to the 2002 National Population and Housing Census (NSP2002). The NSP2002 was the last one, which was carried out in traditional way. The census information was collected on the paper census questionnaire directly from respondents by census enumerators who visited all persons inhabiting the territory of Poland. Data from the paper questionnaires were registered in the IT system and then they were further processed and analyzed. Despite all its advantages that method of conducing census survey turned out to be very expensive and laborious. Moreover, the development of administrative data sources and IT technology applied for its hosting has been observed for many years. Therefore, considering all pros and cons it was decided to resign from the traditional mode census in favor of methods that combine different data sources. Such solution was also approved by the European Commission in the Regulation (EC) No 763/2008 of the European Parliament and of the Council, which provides specifically defined sources that might be used to collect statistical data for the purposes of the census.

3. As a result, in the 2011 National Population and Housing Census (NSP2011) the mixed method was applied. It combined data from regular statistical surveys and administrative data sources with data collected directly from population as part of sampling survey and so-called complete survey. Additionally, two complete surveys were conducted, covering persons staying in collective living quarters as well the homeless people. The Act on the NSP2011 stipulated that the information systems of public administration shall be used as widely as possible. Data not included in the administrative data sources or data ineligible in terms of the statistical data quality were collected from persons covered by the census. In this case, however, it is envisaged that modern techniques of data collection shall be used in order to eliminate paper forms.

4. As far as information on housing stocks are concerned in the NSP2011 the Building Base (BB) constituted the main source of information on buildings and dwellings. The base stores data on housing stocks that were possible to be compiled using administrative data or data collected within the statistical surveys specially dedicated to the housing topics. Information gathered in the BB were verified within the full-scale census survey conducted as a self-enumeration by Internet. However, not all information on housing topics was possible to be compiled basing on the BB. As a result some housing topics were also covered by sampling survey (20% of dwellings).

II. Establishing the Building Base (BB) as the main data source for housing topics in the 2011 Population and Housing Census

5. As mentioned above the BB was created in order to collect in one database all accessible information on residential buildings as well as non-residential buildings with at least one dwelling (the objective scope of the BB). The base was established in 2010. Generally in the BB each building had its own record with the set of established fields that included information on: type of building, type of ownership, year/period of completion, number of dwellings, useful floor area of dwellings in a building, number of rooms, installations in a building: water supply system, sewage system, gas from network, type of central heating. At the preparatory stage there was also planned to obtain information on facilities for disabled people (outside and inside building) such as a lift or a porch. However, the data sources available for official statistics did not ensure the suitable coverage of data.

6. The first step was to identify the data sources on housing topics that were accessible to official statistics at being time. There were both statistical as well as administrative data sources. The following administrative registers were specified as the first:

(a) The *System of main data source for address identification of streets, real estates, buildings and dwellings* (NOBC) as the part of the National Official Register of the Territorial Division of the Country (TERYT); and

(b) The *Register of Lands and Buildings* (EGiB).

7. As far as the statistical surveys are concerned the following were identified:

(a) The *2002 National Population and Housing Census* (NSP2002);

(b) The *Report on residential buildings and dwellings in non-residential buildings (completed)* (B-07);

(c) The *Building Electronic File* questionnaire (EKB2010).

8. The reference population of buildings in the BB was established based on information included in the *System of main data source for address identification of streets, real estates, buildings and dwellings* (NOBC)[2]. The first list was dated as of 31.12.2009. The NOBC consists of reference tables separately for buildings and dwellings. These tables contain building and dwelling addresses, in configuration of statistical regions and census enumeration area. The new building is registered on the basis of information obtained from inspectors of constructional supervision within the Register of residential buildings and non-residential buildings as well as the buildings of collective accommodation (completed) for the needs of TERYT and B-07 Report on residential buildings and dwellings in non-residential buildings (completed). When the building is registered it is provided with the following unique identifiers: ID_NOBC – technical identifier of address and building, ID_BUD_NOBC – technical identifier of building and ID_ ADR_NOBC – technical identifier of address.

---

[2] NOBC – the part of the National Official Register of the Territorial Division of the Country (TERYT) which functions on ground of The Law of June 29th, 1995 on official statistics (Journal of Laws from 2016, item 1068 with amendments) and the Regulation of the Council of Ministers of 15 December 1998 on detailed principles of running, using and making available the territorial register and related to it obligations of the state administration agencies and the self-government entities (Journal of Laws from 1998 No 157, item 1031 with amendments). TERYT contains systems of: identifiers and names of units of territorial division, identifiers and names of localities, statistical regions and census enumeration areas, address identification of streets, real estates, buildings and dwellings.

9. *The Register of Lands and Buildings* (EGiB) was the next one regarded as a potential data source. The EGiB is one of the reference registers covering data on real estates. It is commonly used by public administration, within these as a data source characterizing the real estates for the purpose of deeds. Therefore, there are information on ownership as well as the cadastral value of real estate. However, after analysing the coverage of buildings it turned out that in some areas of country the EGiB worked very well while in others there were a lot of shortage in datasets. As a result, the EGiB was not taken as a data source in the BB in 2010.

10. *The Report on residential buildings and dwellings in non-residential buildings (completed)* (B-07) constitutes the supplement to the *Register of residential buildings and non-residential buildings as well as the buildings of collective accommodation (completed)* run for the needs of TERYT (*the National Official Register of the Territorial Division of the Country*). The Report has been prepared quarterly by inspectors of building control since 2000. It collects information for newly constructed residential buildings and modernized buildings as well as dwellings in non-residential buildings.

11. As the NSP2002 was conducted with the usage of traditional method all census information was compiled on the base of data collected directly from respondents and reflect the actual state. As the housing stocks are not so much prone to changes in time as population the NSP2002 can be used to impute information on many variable describing buildings and dwellings. However, due to the timespan between the NSP2002 and the NSP2011 that source was applied as a last one.

12. After preliminary analysis of data coverage in the above specified sources it turned out that for one- or two-dwelling buildings information on building can be easily imputed for dwellings while in case of multi-dwelling buildings such solution requires undertaking much more efforts. Therefore, the additional data source on residential multi-dwelling buildings was needed. As a result, the Building Electronic File (EKB2010) was established as a tool dedicated for collecting information for residential multi-dwelling buildings. The obligation to report necessary data was imposed on the entities, which rule, manage or administrate housing stocks located in the buildings owned by employers, entities of local governments, housing cooperatives, social construction associations, the State Treasury, housing communities and others. The survey was conducted by means of electronic reporting via the CSO's website (about 80% of reports), application on CDs (less than 2%) and paper questionnaire (about 19%). In the 2010 edition of survey almost 180 000 units were covered by survey. As a result in the 2010 edition of survey information on over 270 000 buildings were collected.

13. In order to integrate data coming from different data sources the NOBC identifiers were used as the reference identifiers. As mentioned earlier, when new building is reported in the B-07 questionnaire in the NOBC a new record is created and automatically the NOBC identifiers are assigned to the given building. In the EKB2010 the IT system of survey automatically sets up the NOBC identifiers regarding the address of building. Another method was applied for incorporating into the BB data coming from the NSP2002. The special integration key had to be created. It consisted of information on address as well as the reference number of questionnaire in the NSP2002.
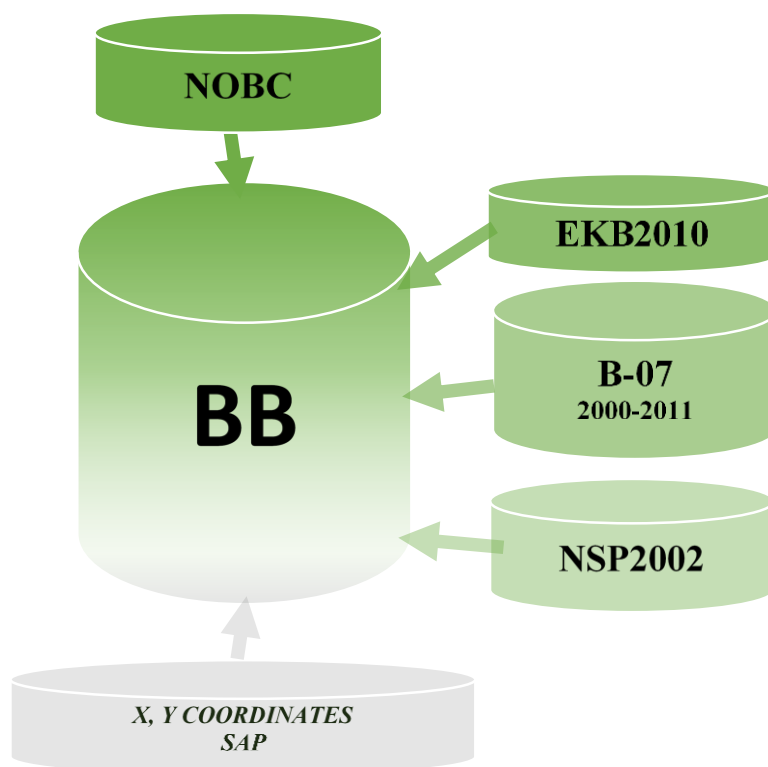
14.   For some topics information can be found in more than one data source specified above:

Table 1
Data Sources

Topics	NOBC	EKB2010	B-07	NSP2002
Type of building	x	x		x
Type of construction (individual, community, collective, for rent or sale, etc.)			x	
Year of completion	x	x	x	x
Number of floors in a building			x	
Number of dwellings in a building	x	x	x	x
Useful floor area of dwellings			x	x
Cubature of building			x	
Type of ownership	x	x	x	x
Water-supply system in building		x		x
Sewage system in building		x		x
Type of central heating		x		x
Gas from network		x		x
Modernization works since 2001		x		
Scope of construction works			x	
Lifts and porch for disabled people		x		
Applied technology construction			x	

15.   As some information is accessible in more than one data source the method for selecting the most suitable data had to be established. The buildings and dwellings in the NOBC constituted the reference population for the BB. The EKB2010 was regarded as the most reliable and updated to the reference date of census data source. The B-07 was the second most suitable data source, followed by the NSP2002.

Chart 1
Data sources in the Building Base for the needs of the NSP2011

16.   After completing the NSP2011 the BB was supplemented with information on the x, y coordinates (grid reference). This information come from the spatial address databases run in the CSO of Poland. One of data source applied in this base just after the census the results of the pre-census survey were. Within the survey the census enumerators verified all address points. They confirmed or modified the address points included in the prepared list or removed the given point if there was no building there. Moreover, they were obliged to mark the building at the digit map displayed on their hand-held electronic appliances. Assigning the grid reference to the building enables to present the results at the map.

17.   Within the NSP2011 the BB fulfilled many functions. It was used as a data source for information on: one- and two-dwelling buildings that were not surveyed within the sampling survey, multi-dwelling buildings as well as on the selected features and parameters of dwellings not surveyed within a sampling survey - by imputation. Moreover, the BB constituted the data source for downloading the pre-filled in questionnaires of census, which were verified by respondents within the self-enumeration process.

18.   The experiences gained within the works related to creating up the BB and compiling the census results on housing topics indicated the areas that needed to be further developed. The most challengeable task was to join individual persons with dwellings they inhabit. The difficulties mainly resulted from the lack of coherence between the address of residence and address of building. The additional works were needed to be undertaken in the NOBC.

19.   Moreover, some other constraints related with the data coverage were identified. First of all, after integrating information included in the BB with data collected for some variables directly within the census surveys the lack of data was recorded. In case of water supply or sewage system there was no gaps while for the year of building completion the percentage of non-settled values accounted for almost 7%, for the useful floor area or the

number of rooms or ownership – above 5%. As a result, the searching new data sources was strongly recommended.

III. Development of the Building Base into the Building and Dwelling Base

20.   After finishing the works related to the NSP2011, the experiences gained within it were summed up and the set of recommendations for development of the BB was elaborated. These recommendations constituted the basis for planning the timetable of necessary works that should be carried out in order to ensure that the BB will be able to meet the information needs of both the nearest the 2021 Population and Housing Census (NSP2021) as well as the following censuses.

21.   It should be regarded that nowadays the BB is the product of regular survey, which was introduced into the official statistics in 2014. As a result it is updated yearly with information available for official statistics. The reference population of buildings is drawn from the NOBC as of 31 December of each reference year. In order to get quality information on buildings data collected within the *Report on residential buildings and dwellings in non-residential buildings (completed)* (B-07) are applied. Then the x, y coordinates are updated.

22.   Apart from regular updating the BB it was supplemented with two one-off downloads of data, which came from: the NSP2011 as well as from the second edition of the Building Electronic File (EKB2014), which was conducted in 2014.

23.   As far as the EKB2014 is concerned, the buildings covered by survey were selected purposively taking into account the lack of information identified in the BB. As a result data for almost 43 000 of buildings were collected from about 4 900 entities which rule, manage or administrate housing stocks located in the buildings owned by employers, entities of local governments, housing cooperatives, social construction associations, the State Treasury, housing communities and others. Similarly to the first data collection (in 2010) the subjective scope comprised: type of building, year of completion, type of ownership, installations in a building (water-supply system, sewage system, type of central heating, gas from network), number of dwellings in a building. In order to get the most updated data just before national census the third edition of survey is foreseen for the 2019 reference year.

24.   Apart from the works over the BB described above, the additional activities were undertaken in order to enlarge the scope of the base with information on dwellings. As a result, the Building and Dwelling Base (BDB) is planned to be created in the end of 2018. The first reference for joint base is 31 December 2017. In 2016, the exercise, which consisted in assigning records on dwellings to the buildings records in the BB was done. The reference date was 31 December 2013. Moreover, the coverage of information characterizing the dwellings was reviewed.

25.   During the current works related with updating the BB as well as experimental joining the BB with dwellings records, some areas that require additional efforts were recognized. First of all, in order to improve the coherence between the address data recorded for population and housing stocks the methods for assignment of technical identifiers for buildings and dwellings in the NOBC are recommended to be modified. Nowadays, these works are in progress. Moreover, there are still some variables and/or some areas of country, for which data are incomplete or outdated. In order to fill in this information gap new data sources have been searched for. The potential data sources selected for implementation into the BB can be divided into three groups: statistical data sources, administrative data sources and others.

26.   In the official statistics, there are surveys of which purpose is to obtain data necessary for monitoring different social and economic phenomenon. As a result they often collect data on housing stocks. The example of such survey is the *Households Budget Survey*. It collects information about the living conditions of households, within this about housing conditions. Each year it is possible to obtain information on about 37 000 of dwellings.

27.   Moreover, it is planned to use administrative data stored in the *Registers of Real Estate Prices and Values* (RCiWN). The RCiWN is the main data source in the *Real estate sales surveys* and *Price on residential premises survey*. In the RCWiN transactions on the sale of real estate are recorded. Apart from the basic information about transaction, for example kind of premises, its size or value, description of real estate is included.
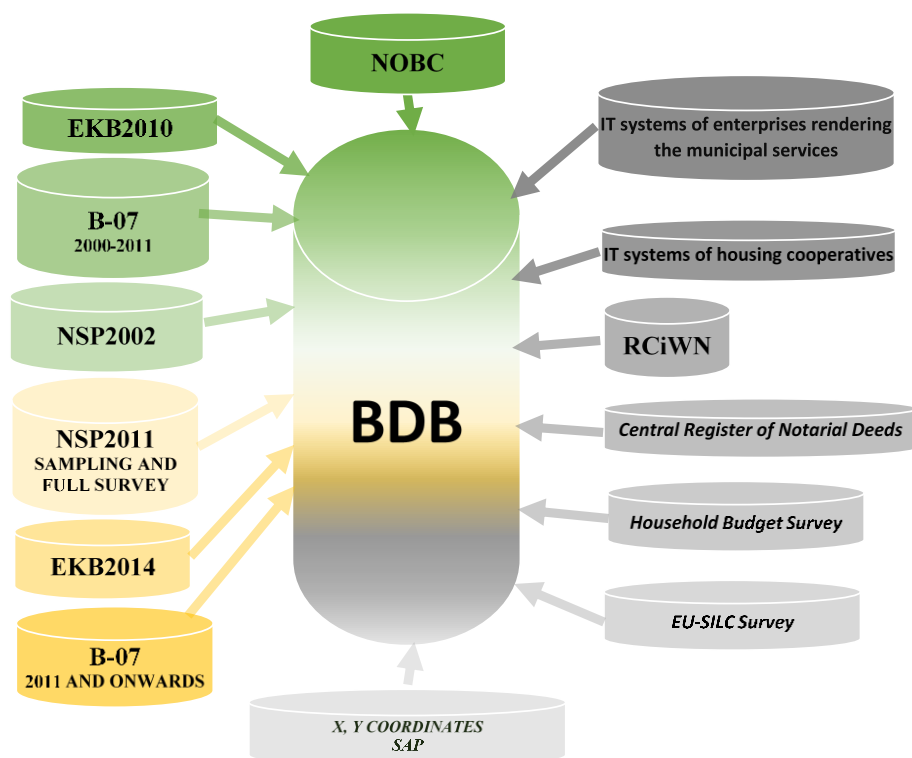
28.   As far as the administrative data sources are concerned, the Central Register of Notarial Deeds is going to be used. In the notarial deed data characterizing the premises are included, but not only. The great advantage of that data source is that it specify the owner of real estate. As a result, it is possible to establish how many dwellings are inhabited by the owners and how many people possess more than one dwelling (so called second dwellings).

29.   When the possible data sources were reviewed, it was recognized that there are other than administrative and statistical databases that include information on dwellings or buildings. Due to the technological development the enterprises that render municipal services (in the field of water supply, waste wastewater discharge, network gas supply or network heat supply) store information on their services receivers in the IT systems. In order to implement these data sources into the official statistics the CSO of Poland entered to the grant agreement with the European Commission "Improvement the quality of European censuses (2012 and after 2021)". The aim of works undertaken within this grant agreement is to assess the feasibility for regular collecting data on the addresses where the services are provided. The pilot study, which obtained a set of meta information on the datasets and IT systems operated by entities rendering the afore-mentioned public services, has already been completed. The results of pilot study prove that the majority of enterprises in question host the IT systems and it is possible to collect necessary data from them. The development of the methods for integrating data obtained from the afore-mentioned IT systems with the BB will be the next step.

30.   Similarly to the entities rendering the municipal services, the housing cooperatives also use the IT systems in order to manage the housing stocks they administrate or rule. In the nearest future, it is planned to carry out the pilot study, which will collect meta information on the IT systems of the cooperatives. As in Poland there are still relatively many dwellings located in the buildings owned or co-owned by the housing cooperatives, it is crucial to establish the regular data collection.

31.   The action undertaken so far as well as those planned in the following years aim at developing the BB into the BDB in order to create the data source for housing topics in the NSP2021. It is planned that finally the BDB will incorporate the following data sources:

Chart 2
Data sources in the Building and Dwelling Base for the needs of future censuses*

*The green elements represent the data sources used for creation of the BB for the needs of the NSP2011, yellow parts contain sources that have been applied so far for updating the BB while the new recognized sources are specified in the grey fields.
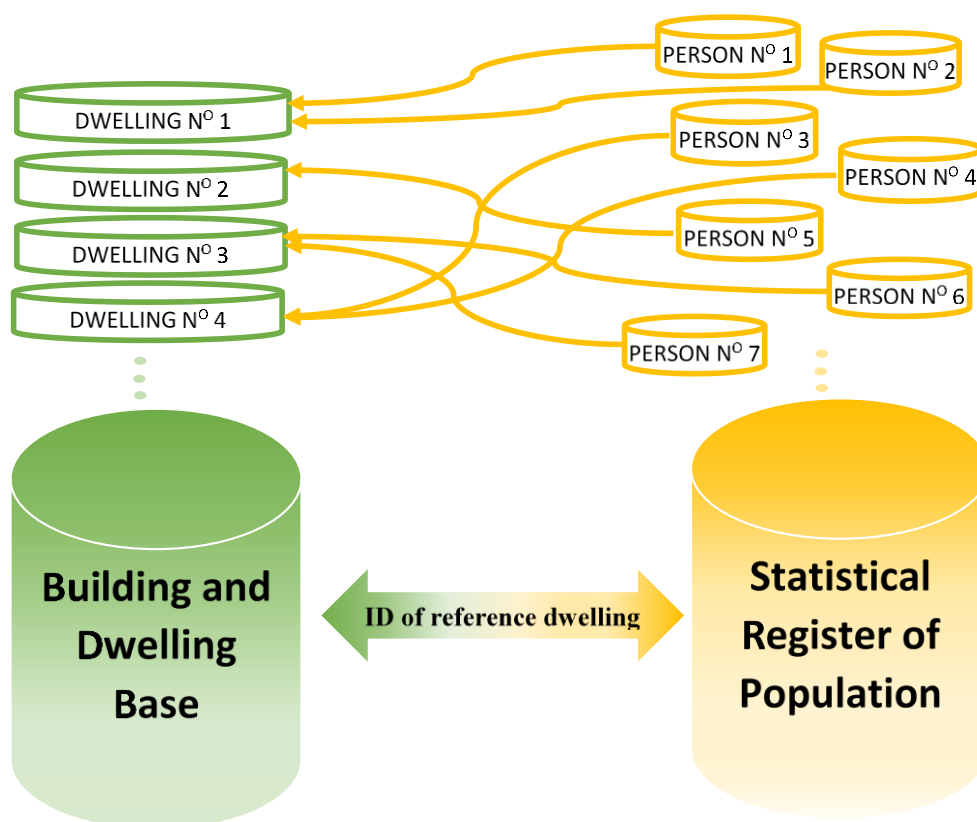
IV. The Building and Dwelling Base as a part of the Statistical Register of Population

32.   Comparing to the previous censuses the upcoming census is going to be conducted with wide application of data as well as analytical possibilities of the newly created Statistical Register of Population (SRP). The conceptual works related with creating the SRP and its updating started up in 2016. The idea for establishing the SRP was driven by the necessity for adopting and implementing new methodology for compiling the population size and its structure for the purposes specified in the Regulation (EU) No 1260/2013 of the European Parliament and of the Council of 20 November 2013 on European demographic statistics. Moreover, in the long-term perspective the SRP is to be used as data source for the post 2021 censuses (annual update of the NSP2021 data).

33.   The SRP will gather information on population, which will come from administrative data sources and statistical surveys (BAEL, EU-SILC, etc.), within these data collected in the previous censuses (NSP2002, NSP2011) as well as from any other sources of information. In order to compile information on housing conditions of persons, families and households the SRP will be integrated with the BDB.

Chart 3
The integration of the Building and Dwelling Base with the Statistical Register of Population

34.   At the end of current year, it is planned to conduct the pilot exercise for integrating the SRP with the BDB. It will allow to place persons registered in the SRP into dwellings in the BDB. Getting the coherence between those two elements is one of the most important part of the works towards building the SRP. In the BDB each dwelling will be equipped with its NOBC identifier while in the SRP each person will have an address of residence.

35.   It is assumed that the SRP will be used in order to:

(a)   Store data from IT systems and data collected directly from respondents by means of different channels (CAxI);

(b)   Validate, control and correct stored data and their analysis;

(c)   Adjust data from the administrative data sources with statistical data;

(d)   Integrate data for compiling values of census variables;

(e)   Initially download the census questionnaire;

(f)   Compile and disseminate result of census;

(g)   Conduct analysis of data applying advanced statistical methods as well as data mining.

V. Conclusion

36.   Among the statisticians all over the world the awareness arises that the traditional way of conducting the population and housing census gradually disappears. More and more countries resign from collecting census information directly from inhabitants in favor of other methods of data collection, which do not require the direct contact with the respondents. Moreover, new statistical and mathematical methods are searched for, which could be applied in order to compile census data at the needed level of detail.

37.   The BB was created in order to provide information on housing stocks in the NSP2011. However, just after completing the NSP2011 and disseminating the results of census it turned out that the BB had to be developed. In 2014, the BB was introduced as a regular survey into the official statistics. Nowadays, in the dawn of next census the BB is seen as the main data source for housing stocks. However, in order to meet the information needs it has to be developed. Therefore, it is being enlarged by the mode for dwellings.

38.   Based on the experiences gained within the NSP2011, it can be stated that the integration of the BDB with the SRP is the most challengeable tasks that has to be carried out before next census. The method, which will allow to place persons in the dwellings they inhabit has to be elaborated and tested before the reference date of census.

39.   Regardless of the method of conducting the next census, undoubtedly, the BDB will constitute the crucial role in the statistics on housing topics. It is planned to be further modernized in order to provide information related not only to social issues, but also to market aspects of residential premises. It will be used as a data source for compiling the balance of housings stocks but also for monitoring the phenomenon that take place on the market of residential premises.