Economic and Social Council

Economic Commission for Europe

Conference of European Statisticians

Group of Experts on Population and Housing Censuses

Eighteenth Meeting
Geneva, 28 - 30 September 2016
Item 7 of the provisional agenda
Possible uses of new data sources (e.g. "Big Data") for censuses

New data sources and new technologies in traditional methodology of population censuses: plans for the Russian National Population Census of the 2020 round

Note by the Federal State Statistics Service (Rosstat)

Summary

The outcomes of the most recent Russian National Population Census of 2010 necessitate the need to modernize some approaches to population censuses; without such modernization its main objective of having a full and objective picture of modern society cannot be achieved. Censuses should utilize modern data collection methods and reduce respondent burden.

In addition to traditional paper questionnaires, the 2020 census will use electronic questionnaires at tablets and Internet (online) census. All existing administrative population databases have been reviewed; such review resulted in identifying the Uniform System for User Identification and Authentication suitable for the census. This system will be used as an entry point to the portal of Internet census. Furthermore, consideration is given to using geoanalytics Big Data - aggregated data of mobile phone operators – for monitoring census cover

I. Introduction

1.   The outcomes of the most recent Russian National Census of 2010 necessitate the need to modernize some approaches to population censuses; without such modernization its main objective of having a full and objective picture of modern society cannot be achieved.

2.   Censuses should utilize modern data collection methods and reduce respondent burden. This is due to the development of information technologies, wide use of the Internet in everyday life, the wish expressed by respondents already in 2010 to fill in the questionnaire by themselves whenever convenient to them, great difficulties with mobilizing temporary enumerators, especially in cities.

II. Background

3.   The 2020 census will utilize both paper and electronic questionnaires (with identical questions) and three ways to collect the data about the population:

- • - self-administered online completion of electronic questionnaire;

- • - interviewing and completion of electronic questionnaire at tablets with pre-installed specialized software by enumerators;

- • - interviewing and traditional completion of paper questionnaire by enumerators.

4.   During the census the data will be also collected in special premises, in paper or electronic format.

5.   During the 2020 Census it is proposed to introduce a public service of participating in the census via completion of electronic questionnaire by respondents at the government census portal. The possibility to participate in online census will be provided to all people. The experience of censuses in other countries demonstrates that 25%-60% of coverage by Internet census can be expected.

6.   At the close of the Internet census, enumerators will visit all premises of enumeration districts, record the confirmation that individuals participated in the online census and interview those who did not participate in the online census (using tablets or paper questionnaire).

7.   Tablets with electronic questionnaires can be used by enumerators in large cities.

8.   In remote areas and in small rural settlements where it is unpractical to use tablets enumerators will use paper questionnaires. Some areas that were considered to be remote in 2010 can pass into the category of normal areas due to the online census (e.g. in areas with good Internet connection but difficult for physical access by enumerators).

9.   In their work enumerators will use modern technological solutions (GPS module for tablets, other electronic devices for transferring information on enumerator's position location) for quality control purposes.

10.   To monitor the coverage of all housings where people live (or can live) the census will use electronic repository of respondent addresses that was created during the latest 2010 population census.

11.   Operational census plans will be prepared based on this repository and online census coverage will be monitored. Such monitoring will include checking whether a respondent has participated in the census by any mode – online, at a station or at home by being interviewed by an enumerator.

III. New data sources

12. The Russtat reviewed the possibility to use existing administrative population data for checking the census data for consistency. Administrative databases have been established for agency-level registration of population in the Compulsory Health Insurance Fund, Pension Fund, Federal Tax Service, Passport and Visa service. All these sources have their methodological specifics and some limitations or redundancies of data.

13. Thus, the Federal Tax Service registers individuals according to the places of residence as well as according to the locations of owned real estate, vehicles and other property. Therefore an individual can be at the same registered in tax authorities of different areas.

14. In the databases of the Pension Fund an individual account for an insured person is opened only once. If an insured person changes his/her place of residence, his/her individual account continues being registered at the place where it was opened and it is not transferred to a new location. The data of an insured person's account are updated at the place where the account was opened rather than where a person lives.

15. In the Compulsory Health Insurance Fund insured persons are registered irrespective of changes in the place of residence and registration at the place of temporary residence. In the data maintained by passport and visa services individuals who have moved to other locations are not to be deregistered as per the existing legislation.

16. Thus, due to the specifics of the available agency-level population data, such data cannot be used as an alternative or additional source for statistical purposes and for population census.

17. Today Russia has the Unified Portal of Public Services which is a way to access e-government services provided by various government agencies for Russian citizens registered in the Unified System of Identification and Authentication (USIA). Today, the USIA contains registrations of about 25 million people (almost 17% of the country's population). By 2020, the number of registered users is expected to increase, both due to the unification with municipal and agency-level personal identification systems and extension of public services.

18. The USIA contains personal data of 'authenticated' users who provided such data during initial registration and verified in agency-level databases of the Pension Fund and Federal Migration Service. Some of such personal data match the topics collected in the population census: sex, date and place of birth, nationality, place of residence, data about children (sex and date of birth) and contact details. Organizing online census by adding this service to the Unified Portal of Public Services will allow a respondent when completing census questionnaire to use his/her username and password and other data of his/her user account, i.e. personal data available in the USIA. We should take into account the administrative time lag between personal data update at the portal and also ensure that the principles of self-identification are complied with. Therefore there will be a possibility for a respondent to change the data in the questionnaire as compared to those available in the USIA.

19. Another area for development and innovations in the 2020 population census is about using Big Data for improving coverage and quality of census outcomes. Big Data are the source of new data which official statistics cannot ignore.

20. The use of Big Data means working with wealth of information, often heterogeneous and not structured, such data are received in online environment and from various (in most cases private) sources.

21.   As a first step, we will analyse the available in Russia sources of non-statistical data (Big Data) about the population size and/or structure and addresses of housings where people live, or may live, in each region in Russia and try to assess accessibility and evaluate costs for using such data for census purposes.

22.   One of such non-statistical sources uses geoanalytics which is a modern approach for obtaining aggregated datasets from mobile operators using the technology capabilities of Big Data, such data are to be further reviewed and statistically analysed. Geoanalytics is based on the method consisting in analysing technology events of mobile network operators and building mathematic predictive models to obtain data about the population, without the use and processing of personal data of subscribers. Using such method allows analysing the size and trends in the movement of people over time and in big areas.

23.   The review of the size and trends of movements of people based on geoanalytics is an efficient and compelling source of data is proven by the fact that many government and municipal organizations already use such data in their activities, particularly, for urban development.

24.   As part of the preparations to the 2020 Census, the Russtat intends to consider using the following indicators of geoanalytics big data for monitoring census coverage: *population* living in an area in the reporting month; and *average number of people* who were in an area at home (at work, summer house) on the night of census.

25.   The Russtat considers Big Data as a potential element of statistical information system and as a potential source of data for various statistical topics, including population. The use of Big Data is in line with the ongoing programme to modernize Russian statistics for reducing respondent burden and for cost saving. Integration of the sources of big data into the information system requires first of all addressing the issue of legislative framework for accessing big data and using the data for building official statistics.