

**Economic and Social Council**Distr.: General
16 July 2015

Original: English

Economic Commission for Europe

Conference of European Statisticians

Group of Experts on Population and Housing Censuses**Seventeenth Meeting**

Geneva, 30 September to 2 October 2015

Item 3 of the provisional agenda

Innovations planned for 2020 census round, and results of tests**Innovations preparing for the Population and Housing
Census 2021 in Latvia****Note by the Central Statistical Bureau of Latvia¹***Summary*

After the previous Population and Housing Census (PHC) 2011 in Latvia there was a governmental decision that the next PHC in 2021 should be organized with all the information derived from administrative registers and other databases or regular sample surveys of the Central Statistical Bureau of Latvia (CSB), moving away from population surveys on the Internet or at their homes. An inter-ministerial (inter-institutional) working group under the guidance of the Ministry of Economics (ME) was established to work out timely proposals for the preparation for the next census to improve the organization, including particular actions and indicative funding.

The Action Plan adopted by the Cabinet of Ministers of Latvia in June 2015 includes activities of the CSB – creation of the Social Statistics Data Warehouse, working out and further improvement of a new method for estimating the annual number of population, the Micro Census 2015, evaluation of administrative data availability, quality and data consistency for the PHC needs, working out of a system of indicators as a part of general preparation for the PHC 2021 in Latvia. The new approaches worked out and planned activities are explained in the document.

I. Introduction

1. During the PHC2000 the CSB used the Population Register (PR) data. In 2011, in addition to the PR, the State Address Register (SAR) and Real Estate State Cadastre Information System (RESCIS) data were used. Besides, the Ministry of Education and Science (MES), the State Revenue Service (SRS), the State Social Insurance Agency (SSIA), the State Employment Agency (SEA), the National Health Service (NHS) and local governments provided administrative information on persons who were not counted in the PHC2011, but found as active in the institutions' registers, and, using mathematical statistical methods, the CSB developed criteria to determine whether the person is considered as a Latvian resident or not.

2. The Cabinet of Ministers adopting an informative report on the PHC results issued a Decision in February 2012 that the ME under which is the CSB of Latvia must establish and manage an inter-ministerial (inter-institutional) working group to develop proposals for the timely organization of activities for the next census improvement with the aim that the 2021 census will be register based, that is, all the necessary information will be obtained from administrative records and other databases, moving away from population surveys on the Internet or at their homes.

3. The Inter-ministerial working group prepared an Informative report that was adopted by the Cabinet of Ministers in April 2013 when the final decision was made that the PHC2021 will be carried out using administrative data sources, as well as data from regular sample surveys organized by the CSB.

4. The Action Plan, which will be implemented by the CSB and other involved institutions and local governments, was worked out by the inter-ministerial working group on the base of the Informative report and adopted by the Cabinet of Ministers in June 2015. Several activities are to be undertaken by other institutions - improvement of the SAR and the RESC IS by the State Land Service (SLS), improvement of the PR data base by the Office of Citizenship and Migration Affairs (OCMA), improvement of the Education Register (ER) as regards higher education institutions to be carried out by the MES, provision of availability of administrative data on occupation of employees by the SRS.

II. Activities to be taken by the CSB of Latvia

A. Creation and maintenance of the Data warehouse

5. To provide accurate annual population assessment, the CSB has initiated building of the information system "Social statistical data warehouse" as a specially designed multi-user system, set up with an aim to create a single administrative data warehouse in the CSB for the use in social statistics. It has launched the information system, which includes integrated and interconnected information from such data sources as the PR, the SAR, the SREC IS, the SRS database, the SSIA database, the SEA database, the HPC database, the MES database, local government databases, etc. as well as the CSB regular surveys.

6. The system uses three terms constituting the system base units:

- (a) data source - a set of tables containing structured information about a particular group of objects (persons or housing units);
- (b) table - file containing data arranged in rows and columns;

(c) data - textual or numerical information depicted in formalized manner, which the user or data processing devices may interpret, process or transfer. Classifiers that can be found in the tables will also be considered as data.

7. The system is designed in such a user-friendly way that employees without programming language knowledge can create data source metadata descriptions and data source periods, import and link multiple data sources, perform data checks and corrections, create data versions valid for a specified period of time, perform data selection export (using metadata-based field descriptions) etc. in a graphic user interface structured on a base of process steps.

8. A user rights management module is developed within the system that provides the functionality to control / restrict access to data and actions with different datasets for different users.

9. Further activities before the PHC2021 include replenishment of the existing data warehouse feature functionality - the definition of complex links between tables within a single source or between a number of sources; an automatic data repair option - supplementing the data version of rectification functionality with automatic correction of the data.

10. Further activities also include new functions for the development of data warehouse - aggregated data processing capacity; developed reporting functionality; possibility to use common classifiers; unique identifier embedded in the system; multi-period analysis of data; availability of various qualitative and quantitative protocols and reports.

B. A new method for estimating the annual number of population

11. Analysis of the PHC2011 data resulted in acquiring information that helped the CSB to specify the number and composition of the population in the country. On 1 March 2011, the number of Latvia's population accounted for 2 073.4 thousand. As compared to the information published prior (according to the PR), the population number was by 155 thousands or 7% smaller. The data confirmed that a part of the country's population does not fulfil requirements of the Population Register Law and information included in the register on population migration is incomplete.

12. It was ascertained that it is necessary to find ways to estimate more precisely the number of population. The possible ways:

- (a) to improve the data quality on usual residence in the PR (the advisable way, however at the moment it can't be fulfilled);
- (b) to evaluate methods applied by other countries (the information from other countries were collected, but it was not possible to adopt methods used by other countries);
- (c) to work out a new method (what was done).

13. In 2012, the CSB worked out a new method for estimating the annual number of population (the method was developed by the Mathematical Support Division – Mārtiņš Liberts, Jeļena Vaļkovska)¹. The method is based on the PR and on individual data from other administrative registers (all data sources have ID codes which are used to merge data).

¹ Central Statistical Bureau of Latvia (10.02.2014.) *Method used to produce population statistics*. Access at: http://www.csb.gov.lv/sites/default/files/dati/demstat_metodologija_ver21_eng_0.pdf

14. To evaluate the residence status of each individual the Logistic regression model has been developed. With the help of administrative register data, on each person registered within the PR there are more than 200 characteristics variables developed. The aim of the model is to predict the probability (the number falling within the interval from zero to one) of being a resident for each individual, to estimate the status of the actual place of residence at the beginning of the year for every resident registered (on the individual level) in Latvia. The necessary probability to be included in the resident population differs depending on the age and gender.

15. The model has been developed using data from the 2011 Census on the actual place of residence and data from administrative data sources on 2010, 1 January 2011 or 1 March 2011. As the result, the database on the individual level on 1 January for the certain year is set up. The database allows preparing the necessary data tables for national and international data users.

16. Apart from the PR data, population statistics is produced by using also individual data from other administrative registers available to the CSB and meeting the corresponding time period (starting from 2010). The CSB has an access to data in the following administrative registers – the SRS, the SSIA, the MES, the SEA, the NHS, the Agricultural Data Centre (ADC), the Rural Support Service (LRSS), the Road Traffic Safety Directorate (RTSD), main higher education institutions databases, the Social Security Administration Information System (SSAIS) on persons that have received social benefits from local governments (since 2012) and the Latvian Prison Administration (since 2015).

17. To evaluate the quality of estimation household survey data are used (data on the individual level from the Labour Force Survey (LFS), the EU-SILC survey (EU Statistics on Income and Living Conditions), the European Health and Social Integration Survey (EHSIS)); data on social benefits paid from local governments. The evaluation of the results of the method will be one of the 2015 Micro Census's tasks.

18. The method has been presented and positive comments received from Eurostat, OECD experts, the Latvian Statistical Association, experts from the Bank of Latvia, demographers. The main conclusion was – if the information on usual place of residence in the Population Register has quality problems, the developed method is the best solution. Detailed information about the methodology is available on the CSB web page.

C. The Micro Census 2015

19. The main aim of the Micro Census 2015 is to evaluate the quality of population estimates made by the CSB using the new method. It will help to improve this method afterwards.

20. The Census date is 1 September 2015, but information will be asked also about usual residents of the housing unit on 1 January 2015 and expected situation on 1st January 2016. A two-stage probabilistic sampling design is used where administrative territories are primary sampling units and census enumeration areas are secondary sampling units. The size of a sample is 15'000 housing units.

21. In addition questions about educational attainment are asked to obtain information about the level of higher education obtained and the country of studies. It will help to work out estimation methods in cases when particular data will not be available in the Education Register.

D. Evaluation of administrative data availability, quality and data consistency

22. Preparation for the PHC2021 in Latvia will include several activities on the evaluation of data in administrative data sources to be used:

- (a) identification of administrative data that describe the economic characteristics of the population, evaluation of the data quality and connecting with the PR data. Working out of estimation/imputation methodology for missing data;
- (b) identification of administrative data characterizing the demography as well as households, evaluation of the data quality. Development of the necessary algorithm for family matrix and its application through the PR data;
- (c) matching of data from administrative sources characterising demographic information of population with administrative data characterising the place of usual residence and the housing unit. Working out of estimation/imputation methodology for missing data;
- (d) the pilot census in 2020 (data from 40 administrative data sources and regular sample surveys will be used), analysis of the results, introduction of the necessary clarifications and amendments to the organizational and methodological documents.

E. Working out of the system of indicators

23. This system is based on Regulation (EC) No 763/2008 of the European Parliament and of the Council of 9 July 2008 on population and housing censuses, which determines the set of core topics².

24. This activity will include working out of data estimation methodology for indicators and population groups, for which there are no complete data available in administrative data sources (educational attainment of immigrants, defining of kinship and cohabiting partners, historical data on employment, the place of work, etc.).

III. Actions to be taken by other institutions

A. Improvement of the Education Register

25. The ER contains information about school graduates. It was created in 2008. Data on vocational school graduates is available since 2011. Higher education institutions in Latvia are not included in the ER yet.

26. To provide the education sector with complete and high-quality information, as well as to provide information on the highest educational attainment of the population for the PHC2021 the MES plans development with financing of the ERAF funds in 2015.

² <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2008:218:0014:0020:EN:PDF>

B. Availability of administrative data on occupation of employees

27. The SRS started data collection on occupation, hours worked and salaries in 2014. The CSB renewed an inter-institutional agreement with the SRS on data exchange in March 2015. Still data completeness, data quality and compatibility checks should be done by the CSB.

C. Improvement of the Population Register

28. Software development for the PR with an aim to provide quality information needed for the 2021 Population Census provided by the program for obtaining information and the annual population assessments is included in the Action Plan for the PHC2021.

29. It includes advanced address data quality and availability for the statistical needs, improved existing software to reduce the workload of manual operators for processing the change of address, improved interfaces for data quality checks, a regularly powered software to find and add the SAR code for addresses included in the PR, complement the external interfaces and data discharge for statistical needs of information on the address element (county, parish, city, village, street) of the SAR code, as well as the place of residence for automatic cancellation if the person has lost the legal status in Latvia.

D. Improvement of the State Address Register

30. There are no codes for a part of addressing objects in the SAR (approximately 8% of all housing units). In some cases there are mistakes in geocodes of buildings.

31. The main activity is clarification of information and prevention of errors in the SAR ensuring an access code granted to all address objects, including apartments in apartment buildings and buildings mainly in Latgale region where are municipalities with predominantly village-type population. The SLS should organize this work in close co-operation with local governments.

E. Improvement of the Real Estate State Cadastre Information System

32. Specification of the classification of utilities (engineering equipment) should be done on the base of legislative acts worked out by the Ministry of Economics, coordination between stakeholders should be organized. The introduction of the new classification in the related IS and data conversion and the incorporation of changes in the Real Estate State Cadastre Information System (RESCIS) should also be addressed.

33. Identification of acts from the state and local government archives concerning the acceptance of buildings into operation and obtaining of information on the date when the building and utilities in it were completed and registration of information in the RESC IS.

34. Registration of existing, but unregistered in the RESC IS buildings by the SLS in co-operation with local governments.

35. Working out and implementation of the buildings' declaration system for persons (owners of buildings and structures) to provide data to the SLS for updating of information as regards utilities (engineering equipment) in the RESC IS and E-service development and implementation for the electronic transmission of data.

36. Data collection from holders of large utilities to carry out the analysis and obtain data on utilities for buildings registered in the RESC IS.

IV. Conclusions

37. Preparation to the PHC2021 that will be a register based Census in Latvia started in 2012. The Action Plan for preparation and conducting of the PHC2021 is approved by the Cabinet of Ministers of Latvia. The Cabinet approved financing for the CSB as a long term commitment up to 2023.

38. Development of the data warehouse in the CSB is outsourced and based on the contract that is signed for particular steps.

39. In modern society, when it takes only a day or even an hour to change a place of residence, it is important, that people understand the importance to inform the authorities (the Population Register) about changes in their data. In almost every policy area - economic, social or environmental, there is a requirement for high quality demographic statistics to help to formulate operational objectives and to evaluate progress both on the national and the international level. At the end of 2013 the Regulation on European Demographic Statistics entered into force. However, population statistics still requires harmonization - in the Regulation, only Article 4 demands that data on the number of population had to be prepared according to the usual residence population definition. In case of Latvia, it is obvious that there is a difference, if data are produced according to the Population Register, or if the estimation is made to meet the usual residence definition. That is a reason why official statistical data on the number of population are only produced using the developed method and these data are used by national and international data users.

40. As the PHC 2021 in Latvia will be register based Census the method worked out by the CSB of Latvia will be used to evaluate the residence status of each individual.

41. Additional funding for the MES, the CMA, the SLS and the EM will promote acquisition of complete and high quality information for the PHC2021, but at the same time actions will provide quality improvements to the core function of these institutions. At the same time financing for other institutions (the SLS and the OCMA) has been postponed up to now. Nevertheless, it is crucial to get it as otherwise there could be problems to implement the Action Plan.

42. These activities will not only provide an opportunity to make the 2021 census as a register based census, but will also contribute to the core functions of the involved institutions:

- (a) the fight against the shadow economy (the SRS activities);
- (b) the higher education system arrangement (the MES activities);
- (c) a housing self - declaration system alignment (the SLS activities);
- (d) SREC IS data availability and quality improvement (the SLS activities);
- (e) SAR data availability and quality improvement (the SLS activities);
- (f) Improvement of regular social statistics survey organization in the CSB including a wider use of the administrative data warehouse, to reduce the respondent burden and improve the data quality (activities of the CSB).