

# Combining data from registers, surveys and 2011 Population and Housing Census to prepare database for 2021 register-based Population and Housing Census in Latvia

Pēteris Veģis

Central Statistical Bureau of Latvia, e-mail: peteris.vegis@csb.gov.lv

## Abstract

In 2012 Cabinet of Ministers of Latvia made decision that 2021 Census will be based on administrative data or statistical sample survey data, if necessary. CSB of Latvia together with other interested institutions prepared a plan of activities for preparation and conduction of register-based Census in 2021 that was adopted by the Government in June 2015. In 2015 research was started on possibilities to use administrative data sources for Census needs and in parallel another activity to build a Social Statistics Data Warehouse, where all available administrative data will be stored, was started.

Paper will introduce with main results of research done mainly on indicators characterizing the economic activity and educational attainment of population and some future plans, including ongoing research on housing, household and family indicators.

*Keywords:* BNU2018, census, administrative data, register-based.

## 1 Introduction

Use of administrative data for the Population and Housing Census (hereinafter Census) needs was started in 2000 Census, when data from Population Register (hereinafter PR) was available. Address of usual residence was taken from PR and data comparisons were made with PR data.

Wider use of administrative data was organized in preparation, conduction and evaluation of results of 2011 Census. Three basic registers - PR, State Address Register and Real Estate State Cadastre Information System were used. Due to various reasons, e.g., non-response, unmet respondents or interviewer mistakes etc., during the Census it was not possible to obtain information on all persons registered with the PR. Therefore, to find out if the persons not surveyed can (cannot) be considered as the resident population of the Republic of Latvia, on 1 March 2011 the information of the State Revenue Service (hereinafter SRS), the State Employment Agency (hereinafter referred to as SEA), the State Social Insurance Agency (hereinafter SSIA), the National Health Service, the Ministry of Education and Science (hereinafter MES) and local government was used for estimates. This method was further developed for annual population estimates.

The preliminary research of administrative data sources for census needs was started with the Eurostat VIP.ADMIN grant project in 2015 when data on economic activity of population was checked. The next grant project was on educational attainment of population, but the third ongoing grant project is on household, family and housing data.

## **2 Economic activity of population**

The following Census core topics in accordance with the Recommendations of the Conference of European Statisticians and Regulation (EC) No 763/2008 were studied and analysed - current activity status, main job, occupation, industry (branch of economic activity) and status in employment. Annual data are obtained since 01.01.2015.

Assessment of the availability and quality of administrative data in the databases of the SRS, SSIA, SEA, National Education Information System (hereinafter NEIS) of MES, as well as data received from Higher Education Establishments (hereinafter HEE) in comparison with the data from the Labour Force Survey (hereinafter LFS) and results of the 2011 Census was performed. In addition, some research on Farm Structure Survey data and data from the Rural Support Service and the Agricultural Data Centre was done.

### **2.1 Determination of the current activity status**

Data necessary for the research of the economically active and not active population were available from several administrative data sources - information on taxpayers collected by the SRS, information on farm owners/users from the CSB Statistical Farm Register (hereinafter SFR), data from the SEA unemployed database, etc.

The research that was carried out formed a basis for the conclusion that the information included in administrative data sources can be used to acquire information on the economic activity status of persons necessary for the Programme of the 2021 Census. Work on the improvement of methodology that was started in 2015 is continued annually.

### **2.2 Determination of the economically active population**

The CSB has access to administrative data sources – SRS data on tax payers and SEA data on unemployed persons – based on an interdepartmental agreement for acquiring information on the economical characteristics of the population foreseen in the Programme of the Census. Part of the required information (about farm owners/users) may be also found in the SFR of the CSB.

Some differences were identified between the criteria used by the SRS and the employment definition of the ILO. Information on working hours is not available for all employed persons since they are not recorded for certain employees. The database also includes persons under working age. It is not possible to determine whether a person has worked during the reference week for a part of employed persons as data for them are available only on quarterly or annual earnings. The database includes information on persons who have already ceased their working relationships during the reference period, as the SRS data reflect all tax payments, but they may also be made after the termination of employment relations. Data on sickness benefits are collected with a long-time lag.

As regards **determination of employment** based on administrative data only data related to remuneration for work are selected from SRS tables that also contain information on types of income other than wages and salaries. A person is employed if at least one of the tables reflects data on the respective person's hours worked, income, as well as a periodical leave, with maintaining one's job. Information about employment abroad is obtained from tables on income obtained abroad by a natural person (resident) and income obtained abroad by a natural person (seafarer) who is employed on a ship used for international transport.

The analysis carried out on **determination of main job** during the initial project shows that approximately 80% of Latvian employees earn their income only from one job, while 20% are employed in several jobs. Main part of these 20% have their record of work in one of jobs, if not than usual methodology was used - the hours worked, or the average monthly salary were compared between them. It was decided that November would be better reference time as December that is not typical in relation to economic activity because of Christmas and new Year.

The **occupational code of employed persons** has been included in the SRS database since 1 July 2013. Employers must submit information regarding the occupation of employees and the number of hours worked to the SRS. Overall, information on the occupation of employees corresponds with the ILO methodology and it covers about 83% of all employed persons. After comparing occupation codes at two-digit level, it can be concluded that administrative data often do not contain information on those occupations that are present in the sectors with the highest rates of shadow economy in Latvia (construction, trade, taxi services, etc.). Moreover, additional research should be carried out on those employed in agriculture, considering the seasonal nature of agricultural work. The compliance rate of occupation codes in LFS and SRS data at the two-digit level amounts to 61.5%, which is valued as satisfactory. The correspondence at the level of one-digit codes was higher - 68.9%. Development of methodology for obtaining the missing information on occupation of persons employed from regular statistical surveys performed by the CSB and imputation is done in 2017. Work is continued also considering the possibility to use data of various professional associations, etc.

A 4-digit code (NACE Rev. 2) of **Industry (branch of economic activity)** is included in the CSB Statistical Business Register (hereinafter SBR). Industry (branch of economic activity) in the main job for employed persons was determined by combining the SRS data with the SBR data. One of the problems – municipalities in the SRS reports use only one general public administration activity code (NACE 8411), but various businesses are under the supervision thereof. To solve the problem, a methodology for adding appropriate NACE code for the persons employed by enterprises of local governments was developed. During the study, the information on the industries (at section level) obtained from the administrative data was compared with the LFS data (employed persons living in private households). It was possible to use the Section code from SBR or SRS data for 99.4% of employees that had the occupation code. The compliance rate between estimates and LFS accounted for 72.8% that is satisfactory.

It is possible to determine **the status in employment** for those persons who are employed and who have a specific main job. In Latvia, the data are available from two sources: information on taxpayers gathered by the SRS or SFR on owners or users of farms. It is necessary to continue studying the issues related to the employment status by providing special attention to the improvement of the methodology for determining the employers and self-employed status.

SEA collects information on registered **unemployed persons**, job seekers and persons having other status. The CSB and the SEA have concluded inter-departmental agreement on the data receipt. Unemployed persons (99.8%) and job seekers (0.2%) registered with the SEA met the ILO definition of unemployed, while persons having other status did not. Some differences are found with ILO definitions. Only persons who have registered with the SEA are included in the data base. A maximum age limit is defined for the registered unemployed persons. Work on imputation of unemployed persons is started as there is a difference in comparison with the LFS based estimates.

### **2.3 Determination of economically not active population**

The CSB has access to SSIA data on persons receiving state pensions. A comparison with the LFS data and data from 2011 Census suggests that recipients of pensions from administrative data sources are set within the limits of the confidence interval. Information on capital income receivers was obtained from the SRS. Individual data provided by the MES and HEE are used to obtain data about students.

## **3 Educational attainment**

The main source of data is NEIS of MES and HEE as up to 2017 there was no higher education register. In addition, other data sources of professional data bases (Register of Medical Persons and Medical Support Persons, Register of Sailors, data about education of lawyers, bailiffs, notaries and their assistants etc.). Data about higher education obtained abroad is available just partly. Therefore, additional question about it was included in 2015 Population Micro Census and in 2018 External Migration Panel Survey.

Results from collecting data about the highest completed level of education from different administrative data sources (NEIS of MES, HEE, etc.) show that only 1,4% of all population aged 15 and over have no information about the educational attainment. Whereas, for 64,2% of all population aged 15 and over the data source of educational attainment is the 2011 Census database.

The classification of levels of education for the 2021 Census has changed comparing with the classification used in 2011 Census and the result can be observed in data – 7,3% of persons aged 15 and over have obtained higher education, but the new classification and Commission Implementing Regulation (EU) 2017/543 requires more detailed levels: short-cycle tertiary education or bachelor's or equivalent level or master's or equivalent level.

After the data correction and imputation, comparison of results with other data from CSB surveys, analysis of the quality of the data sources will be continued, possibility to define priorities for the data sources will be studied to improve the quality of educational attainment indicator annually.

The accumulative education database on highest education level according to ISCED-A (2011) classification was made and data is used as preprint in regular statistical sample surveys. Work will be continued on maintaining and annual updating of this database that contains information on persons' education from all available data sources and on identifying other possible additional data sources from which information about the level of education of the population could be obtained.

## **4 Conclusions**

Research done is a good base for successful register-based Census 2021 as regards topics of economic activity and educational attainment. Nevertheless, work should be continued to improve data quality and data estimation and imputation methods. Identifying of new data sources, analysing their quality and following of legislative changes in existing ones also should be continued. Research is now continued with checking of administrative data availability on household, family and housing indicators.

New circumstances that declared address of persons will be used should be considered. Existing annual population estimates methodology will be changed. The aim is to develop a methodology based on data from administrative registers but does not directly use the 2011 Census data, which ages each year and becomes less informative about the current situation. As for 2021 Census declared usual residence will be used some problems are identified. Children under the age of 15 without adults are declared to be living in the place of residence. The proportion of single parents, especially lone fathers, increases in the family nucleus. In addition, household concept will be changed from housekeeping concept to household-dwelling concept. Influence of this change should be evaluated, too.

As all annual data is stored in the Social Statistics Data Warehouse it is necessary to improve functionality of it.

## **References**

Conference of European Statisticians *Recommendations for the 2020 Censuses of population and Housing*, United Nations, New York and Geneva, 2015

*Regulation (EC) No 763/2008 of the European Parliament and of the Council of 9 July 2008 on population and housing censuses*, Official Journal of the European Union, 13.08.2008

*Commission Implementing Regulation (EU) 2017/543 of 22 March 2017 laying down rules for the application of 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*, Official Journal of the European Union, 23.03.2017