



Economic Commission for Europe

Conference of European Statisticians

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

Geneva, 26–28 September 2018

Item 2 of the provisional agenda

Methodology, new data sources including big data**Administrative data and sample surveys' data usage for
determination of the educational attainment of population in
the register-based Population and Housing Census in Latvia****Note by the Central Statistical Bureau of Latvia****Summary*

Population and Housing Census 2021 (hereinafter referred to as Census) will be carried out in Latvia on base of information obtained from administrative data sources, as well as regular statistical sample surveys, if necessary.

The Central Statistical Bureau of Latvia (hereinafter referred to as CSB) continues research and analysis of the availability of administrative data to develop administrative data processing methodologies in the field of educational attainment of population of Latvia.

This document will describe main results of the feasibility study of the usage of administrative data and information from statistical sample surveys to obtain data for the Population and Housing Census (hereinafter referred to as Census) section on the educational attainment of population. This Census topic is studied, and analysis of possible data sources is made. Furthermore, information on the highest successfully completed level of education of Latvian population from the Labour Force Survey (hereinafter referred to as LFS) is compared with corresponding data from administrative data sources. Evaluation of main problems on detecting the educational attainment from administrative data sources and proposals for imputation are made.

* Prepared by Peteris Vegis and Sigita Meldere.



Results of initial assessment of the availability and quality of Latvian administrative data in databases of the Ministry of Education and Science (hereinafter referred to as MES), the Higher Education Establishments (hereinafter referred to as HEE), the Ministry of Finance (hereinafter referred to as MF) and other institutions in comparison with the data from the LFS and Census 2011 and imputation made for missing data are described in this document.

I. Introduction

1. The 2021 Census in Latvia will be carried out based on information obtained from administrative data sources, as well as regular statistical sample surveys, if necessary. Work is started with the aim to study the availability of administrative data by assessing the quality of incoming data, as well as to develop administrative data processing methodologies and criteria for assessing the quality of data processing. Feasibility study was organized in the CSB on the use of administrative data and information from statistical sample surveys to obtain data for the Census Programme part on the educational attainment of population.

2. Data on persons' highest educational level is the core topic of Census and represents an overall level of development of the society.

3. The Eurostat Grant Project was related to the topic of the Census entitled "Educational attainment", which refers to the highest education level that is successfully completed by the person in accordance with the International Standard Classification of Education (hereinafter referred to as ISCED) 2011-A level.

4. The objective of this project was to make initial evaluation of the data availability and quality on educational attainment for every person analysing the possibilities to use the data:

(a) Of the MES, the National Education Information System (hereinafter referred to as NEIS), data from HEE and other educational institutions;

(b) Of other administrative data sources, for example Ministry of Health (hereinafter referred to as MH), MF etc.;

(c) From the regular statistical sample surveys performed by the CSB, for example, LFS and European Union Statistics on Income and Living Conditions (hereinafter referred to as EU-SILC);

(d) From the results of 2011 Census. This is a spacious data source on educational level of Latvian population for that time. One of disadvantages of this database is difference on coding of higher education data, because for 2021 Census there is need to code educational levels according to ISCED 2011-A levels: short-cycle tertiary education (ISCED 5 level), bachelor's degree (ISCED 6) and master's degree (ISCED 7), but in 2011 Census data on higher education were not differentiated and it is still a challenge for CSB. In data of educational level where higher education is not distributed on precise levels are labelled by the CSB as ISCED 67;

(e) The linkage of educational attainment with occupations/occupational groups.

5. In the field of education MES had created NEIS in 2010. In accordance with regulatory enactments this system contains information from the Register of Educational Institutions, the Register of Teachers, the Register of Educational Programs, the unified data base on records of children who have gained the compulsory education age and the Academic Staff Register for Educational Institutions, licensed and accredited educational

programs, students, pedagogues, academic staff of universities and colleges, documents and educational state statistics. Inclusion of data about the graduates of general education institutions was started since 2011, but about the graduates of vocational education institutions – since 2013.

6. Up to 2018 data about graduates of the higher education institutions is not available within the NEIS in Latvia. A frame for the Register of Students and Graduates was created in 2015 with financing of the European Regional Development Fund, but the delay in the adoption of regulatory enactments is observed. It is planned that data about graduates of the HEE would be available starting from 2017 graduates. CSB of Latvia made bilateral agreements with HEE to obtain missing data about graduates of HEE from 2011 to 2016.

7. Despite this fact data about students who have graduated the education institutions abroad partly still will be missing as CSB have information just on those Latvian students whose foreign diplomas are recognized by Academic Information Centre of Latvia, who are receiving study credits and registered by Study and Science Administration and those who indicated it in the 2015 Population Microcensus. Information on educational level of inhabitants of Latvia is fragmented. Therefore, research on data availability and completeness was made and several alternative administrative data sources were checked and founded to get more complete data about educational attainment of population of Latvia.

8. The Cabinet of Ministers of the Republic of Latvia has developed regulations on the Occupation Classifier, the basic tasks and basic qualification requirements that are relevant to the occupation. According to them, occupation standards have been developed, which also include requirements for a certain minimum level of education. In order to obtain data on the education of certain occupations, the CSB contacted the Latvian Council of Sworn Notaries, the Latvian Council of Sworn Advocates and the Latvian Council of Sworn Bailiffs and reached an agreement on exchange of personal data. In addition, there is an agreement in preparation on the exchange of information with the Health Inspectorate, which maintains a Register of Medical Persons and Medical Support Persons. It contains information on the level of education obtained by these persons.

9. Data from the different sources were linked and attributed to persons using their unique personal ID code.

II. Determination of the highest level of successfully completed educational attainment according to ISCED-A (2011) classification on 1 January 2017

10. During the process of finding of the best solution for determination of educational attainment of inhabitants of Latvia on 1 January 2017 the following guidelines were considered:

- (a) The information obtained is close to the situation as of 1 January 2017;
- (b) The information in data files is linked by using the unique personal code of each person;
- (c) If a person has information about the successfully completed educational attainment in several data files, then the highest possible level of education is assigned;
- (d) Data sources are assigned according to the information they contain – the highest level of completed educational attainment obtained in administrative data and social surveys of CSB. If there are no newest information on person's educational level in these sources, data from 2011 Census are kept;

(e) If at the same time it was determined that a person has obtained “higher education” (without the breakdown into bachelors and masters), which was equalled to ISCED 67 (coding used by the CSB), as well as the level of education according to the ISCED-A (2011) was the level 5 – short-cycle tertiary education, then the selected level of education was – ISCED 5;

(f) Taking into account changes on educational system in 1996 and according to the opinion of the MES, it was assumed that the higher education (applicable for 5 years of studies) acquired before the 1 July 1991 is recognized as eligible for master's degree. Those persons aged 23 and over on the 1 July 1991 and who declared that they have a higher education in 2011 Census were granted ISCED 7 – a master’s or equivalent degree;

(g) For the 2021 Census purposes the obtained education indicators were compared to the level of education qualification ISCED-A (2011) distinguishing from ISCED 0 to ISCED 8 level.

III. Description of the processing of individual data sources

11. Data on educational attainment of residents of Latvia were collected gradually from different data sources including administrative data sources and regular statistical sample surveys. Data were linked together by personal ID code of each person. The number of persons with particular education level from studied data sources is included in Annex.

A. The 2011 Census database

12. The 2011 Census database is one of the most important and complete data source on person’s highest level of successfully completed educational attainment. The question about the level of education according to 2011 Census program was: “What is your highest level of education successfully completed?”. According to the Regulation (EC) No 763/2008 of the European Parliament and of the Council of 9 July 2008 and the 2011 Census Program, levels of education were classified by the ISCED 1997, whereas on 2021 Census the classification ISCED-A (2011) will be used. Almost all levels of education that are in 2011 Census database can be equated to ISCED-A (2011) classification as shown in Table 1.

Table 1

Equation of levels of education in 2011 Census to forthcoming 2021 Census

<i>Level of education in 2011 Census database</i>	<i>Level of education in 2011 Census questionnaire</i>	<i>Equivalent ISCED-A (2011) for 2021 Census</i>
1	Doctoral degree	8
2	Higher education	5, 6, 7
3	Professional secondary education after secondary school	
4	Vocational education or professional secondary education after basic education	3
5	General secondary education after basic education or vocational education	3
6	Basic education or professional basic education	2
7	Primary education	1
8	Without formal education or less than primary education	0
9	Cannot read or write	0

13. At the time of project, the problem on using 2011 Census data on higher education was identified (see point 4 d). In the 2011 Census 404,3 thousand (22,7%) of population aged 15 and over indicated that they have a higher education. Those persons who in 2011 Census declared that they have levels of education “without formal education or less than primary education” or “cannot read or write” according to the new classification will be classified as ISCED 0 level.

B. Database of MF for 2014–2016

14. One of the data sources for 2021 Census, to determine level of education for a person, will be the database of MF on the employees of the direct subordinate institutions of the State of Latvia. MF data are given annually on personal level. Education of person in the MF database is a compulsory indicator for all officials, but some institutions have added educational indicator to their other employees, thus giving a possibility to obtain information on the education level for a wider range of people. A mapping of educational levels and degrees of MF database for 2021 Census needs is done and specific method of transcoding is developed. The highest or certain (in case of ISCED 67) level of education is assigned to person in the process of transcribing education levels according to ISCED-A (2011).

C. Data of MES

15. MES has a *database of all teachers* since 2012 that are practicing in Latvia and there are data on persons' education that can be used for 2021 Census. 13 text notes were transcribed to ISCED-A (2011) level. The highest or certain (in case of ISCED 67) level of education is assigned to person in the process of transcribing education levels according to ISCED-A (2011).

16. MES has a *database of all academic staff* who works at HEE since 2012. After the consultation with MES, a decision has been considered that a professor, assistant professor, associate professor must have a doctoral degree and a lecturer, researcher, leading researcher, and assistant needs at least a master's degree. The highest level of education is assigned to the person in the process of transcribing national education levels according to ISCED-A (2011).

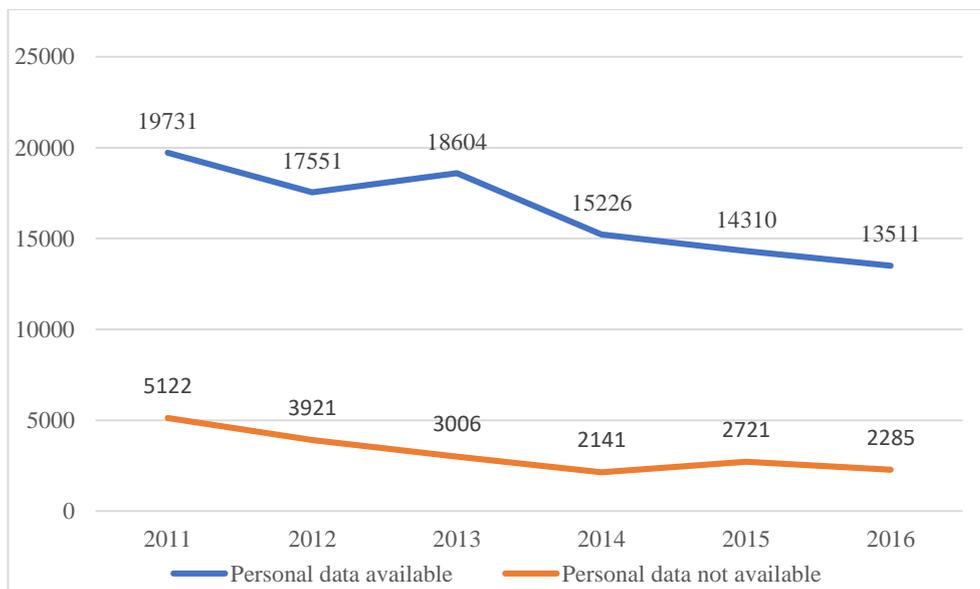
17. MES has a *database on graduates* since 2011 and CSB can assign the highest level of education to the person in the process of transcribing education levels according to ISCED-A (2011).

18. MES *database* contains information on *school pupils*, including their previously obtained level of education or the educational program code that they were obtaining at that moment, thereby this information was used to assign the ISCED-A (2011) level of education to a person.

D. Data files on graduates of HEE for 2010–2016

19. As there is no register of graduates of HEE in Latvia, CSB has collected a personal data from several HEE (from 79.4% of total number of graduates of HEE in 2011 to 85.5% in 2016) after the 2011 Census on the base of bilateral agreements. As it is shown in the Figure I the total number of HEE graduates decreased in Latvia from 24.9 thousand to 15.8 thousand, but proportion of graduates without personal information decreased, too (from 20.6% of total number of graduates of HEE in 2011 to 14.5% in 2016). Nevertheless, to get more complete data set other HEE were contacted and asked to provide annual personal data about graduates on base of bilateral agreement.

Figure I
Number of graduates of the HEE by availability of personal data



20. At the process of alignment of data from HEE, the national classification of educational programs has been transcoded to the corresponding ISCED-A (2011) level.

E. Data files on students of HEE for 2010–2016

21. The data contains information on students of HEE and this information was used to assign the ISCED-A (2011) level of education to a person. Considering that students of HEE should have obtained at least general secondary education, the ISCED-A (2011) level 3 was assigned to persons in the data file.

22. In addition, information on graduates of colleges for 2010 was obtained when data were prepared for the 2011 Census – ISCED 5 was assigned to all persons if there was no information on higher levels of education.

F. Data from Statistical Sample Surveys of the CSB

23. Data on educational level of persons who were surveyed in *LFS* are used for detecting an education level for 2021 Census needs. Principles for the alignment of education level indicated in the *LFS* in accordance with ISCED-A (2011) were worked out.

24. The alignment of education level indicated in *the EU_SILC* in accordance with ISCED-A (2011) has been prepared within the project and used for detecting of education attainment for 2021 Census needs. The highest level of education is assigned to person in the process of transcribing education levels according to ISCED-A (2011).

25. Information on level of persons' education can be taken from the database of annual *survey on the use of information and communication technologies in households* that was held by CSB in 2016. The survey contained the question "What is your highest successfully obtained level of education?". The highest level of education is assigned to person in the process of transcribing education levels according to ISCED-A (2011) is included in specific method of classification of educational level.

26. In 2015 CSB has performed a *Population Microcensus* with the aim to determine how precise the official statistics reflects the number of population in Latvia, in cities under state jurisdiction and counties. One of question in this program was an indicator on education – “What is your highest successfully obtained level of higher education?”. The level of education obtained in Population Microcensus was transcribed into ISCED-A (2011) classification: doctoral degree corresponds to ISCED-A (2011) level 8, master’ s degree or comparable education – corresponds to ISCED-A (2011) level 7, bachelor’ s degree or equivalent education – to ISCED-A (2011) level 6 and the first level of professional higher education corresponds to ISCED-A (2011) level 5.

27. Information on levels of persons’ education can be taken from the database of *Adult Education Survey* that was held by CSB in 2015 and that is organized once in 5 years. This survey contained the question “What is your highest successfully obtained level of education?”.

G. Information from Study and Science Administration database for 2010–2016

28. This database contains information on HEE students who are residents of Latvia and who have taken a student credit (a state guaranteed loan that is intended to cover the student’s subsistence expenses) and a study credit (that is intended for tuition fees) and who are studying abroad. When the student is graduating HEE abroad, he or she must inform the Study and Science Administration on the degree of education obtained.

H. Latvian Academy of Sciences data on Doctors of Sciences

29. This database contains information on Doctors of Sciences. The ISCED 8 was assigned to all persons in data file.

I. Academic Information Centre data on persons who obtained doctoral degree abroad for 2012–2016

30. The education level to all persons in data file was assigned ISCED 8.

J. Persons with particular occupations

31. Data about advocates and their assistants for 2016 received from *Latvian Council of Sworn Advocates*. Data about notaries and their assistants for 2016 received from *Latvian Council of Sworn Notaries*. Data about bailiffs and their assistants for 2016 received from *Latvian Council of Sworn Bailiffs*. The education level to all persons in data file was assigned ISCED 7 or ISCED 8 according to their occupation.

32. Data received from *Register on Medical Persons and Medical Support Persons* collected by Health Inspectorate of MH for 2016 contains information on education of doctors, nurses, dentists etc. medical persons corresponding to the ISCED-A (2011) classification.

IV. Accumulative education database

33. CSB has started to build an accumulative database that contains information on persons’ education from all possible data sources. The basis for this database are data from

Population Register– resident population since 2011 for whom the data on education are linked using unique personal ID code and one highest completed level of education for every person is set. For the first time, the level of education was calculated for 1 January 2016, in the future, with new data sources added, the highest completed level of education for every person will be calculated every year. Data from this database are used for prefilling of the regular statistical sample surveys (LFS, EU-SILC).

34. The work on accumulative database goes on and for the 1 January 2018 new administrative data sources will be available – more data from colleges, Latvian Maritime Academy etc. Starting with 1 January 2018 not only resident population will be added to the accumulative database but also national identification number of a person will be added to this database if it will be present in administrative data sources on level of education of a person.

35. Collected data were stored in the Social Statistics Data Warehouse (SSDW). It is planned that all administrative data on persons` level of education in future will be stored only in SSDW and data will be made accessible to those involved in data analysis on education. Automatic or semi-automatic transcribing to ISCED-A (2011) classification will be implemented within SSDW for those administrative data files on education where possible. Data on highest determined level of education for a person annually will be stored in SSDW and this information will be used for prefilling of the regular statistical sample surveys (LFS, EU-SILC etc.).

V. Education imputation on 01.01.2017

36. The level of education was imputed for persons whose level of education was not known or for whom a sign of the education level being imputed in the 2011 Census data was added. Such persons were 50 697 or 3.08% of all persons` data as of 01.01.2017.

37. For those whose education was known as "higher education", the level of education has been classified according to the ISCED-A (2011) classification. The following persons were 119,468 or 7.26% of all persons` data.

38. A correction in the level of education was made according to the occupation. Such persons were 4,639 or 0.28% of all persons` data.

A. Imputation methodology

39. *Imputation conditions* prepared by the Census team related to the imputation of level of education:

(a) The conditions that prohibit employees with certain occupations from lowering the level of education they may have;

(b) Conditions that prohibit the imputation of inappropriate educational attainment for persons of a certain age.

40. The first type of conditions that were attributable to the occupation were used in the first phase of imputation. The second type of conditions were applied throughout the process of imputation, so that levels of education that are not appropriate for a certain age would not be granted. The donors did not have the "Doctoral or equivalent" level of education.

41. The *imputation* was carried out using the k-nearest neighbours` method [1]. For each recipient of the imputation process, set of donors was made from which five (k=5) nearest neighbours were selected. Neighbours were chosen calculating the Euclidean distance

between each recipient and donors. The list of the variables used for distance calculation is available at point 49.

42. The process was divided into three stages:

- (a) Imputation of education for employees in the occupation, which is included in the imputation conditions;
- (b) Imputation of education for employees with an occupation, which is not included in the conditions of imputation;
- (c) Imputation of education for non-employed persons.

43. At each stage, the imputation was performed for persons for whom:

- (a) The level of education is unknown;
- (b) The level of education was imputed in the 2011 Census data;
- (c) A sign of "Higher education" given, which needs to be broken down to the ISCED- A (2011) classification. (ISCED 5 – short-cycle higher education, ISCED 6 – Bachelor or equivalent, or ISCED 7 – Master or equivalent).

44. During the first stage, six imputation groups were defined in accordance with the conditions of the imputation. Imputation was performed in each of these imputation condition groups. During the first stage, a correction in the level of education was made in accordance with the conditions of imputation.

Example:

If the code of the profession is "3111" (technician) then the level of education must be at least "Secondary education", but the person has a lower level of education – "Primary education", then an adjustment is made, and the person is given "Secondary education" in order to be educated in accordance with the conditions of imputation.

45. During the first stage, for those who were covered by the occupation, the level of education was chosen only from those donors who had the same occupation in accordance with the conditions of imputation.

Example:

Occupation with code "2240" (doctor's assistants, feldsher, military paramedics is selected). Imputation condition applies to the occupation, it determines that if occupation is with the code "2240", then level of education must be at least 5 (ISCED 4 – post-secondary education, which isn't higher education).

46. During the second stage, employees who had a certain occupation but were not included in the conditions for imputation, the level of education was chosen only from donors who had the same occupation. There was no correction in the level of education at this stage, as the occupations are not subject to conditions of imputation in this group.

Example:

Occupation with code "5223" (retailer) is selected. Imputation conditions doesn't apply to the occupation, which determines restrictions for imputation of level of education.

47. During the third stage, for all persons not employed, the level of education was selected from all donors, without dividing the individual groups of donors and recipients by occupation (no occupation). There was no correction in education at this stage.

B. Comparison of Census estimates with LFS estimates

48. A comparison was made with the LFS using survey results from Q1 to Q4 2016 (Figure II). Before comparing the data from the Census, the employees who:

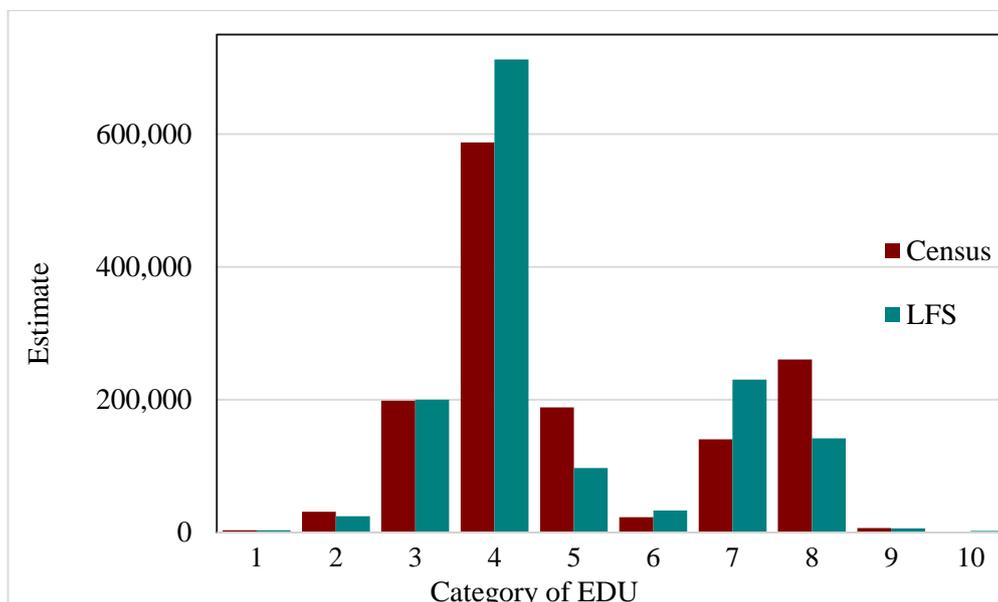
- (a) Are 75 years of age and older;
- (b) Living in collective households.

49. The variables used for distance calculation were gender, age, nationality, country of citizenship, type of citizenship, region of declared place of residence (derived from the territorial code).

50. For each of the domains, a 95% confidence interval was constructed from the LFS data and compared whether the Census data included in it.

Figure II

Comparison of level of education (EDU) in LFS and in census for 15–74 years old residents in private households



51. Comparison of 15–74 years old residents in private households’ distribution by the highest successfully obtained level of education in Census estimates (including imputation) and in LFS estimates on 01.01.2017 is given in Table 2.

Table 2

Distribution of 15–74 years old residents in private households by the highest successfully obtained level of education in Census estimates and in LFS estimates on 01.01.2017

<i>Educational attainment (highest completed level)</i>	<i>Census estimates</i>		<i>LFS estimates</i>	
	<i>Number of persons</i>	<i>%</i>	<i>Number of persons</i>	<i>%</i>
ISCED level 0 – Less than primary education	3,084	0.2	3,319	0.2
ISCED level 1 – Primary education	31,218	2.2	24,381	1.7
ISCED level 2 – Lower secondary education	198,316	13.8	199,701	13.8
ISCED level 3 – Upper secondary education	587,638	40.8	712,457	49.1
ISCED level 4 – Post-secondary non-tertiary education	188,177	13.1	97,007	6.6
ISCED level 5 – Short-cycle tertiary education	23,062	1.6	32,889	2.3
ISCED level 6 – Bachelor's or equivalent level	140,410	9.8	230,357	15.9
ISCED level 7 – Master's or equivalent level	260,246	18.1	141,766	9.8
ISCED level 8 – Doctoral or equivalent level	6,517	0.4	6,009	0.4
Is not indicated	NA	-	2,397	0.2
Total	1,438,668	100.0	1,450,283	100.0

C. Conclusions

52. The developed three-step method can also be used for data prepared by the Census team for data as of 01.01.2018.

53. There are statistically significant differences between LFS and Census estimates.

54. Reasons for differences are several:

(a) After the restoration of independence in 1991, a complex educational system was created in Latvia, which was partly changed year after year. People have problems to follow all changes, for example they are putting themselves in ISCED level 3, but from administrative data most probably the right level is ISCED 4 and similar problems about a higher education;

(b) People who graduated during soviet time have problems to apply the level of education achieved to the present more detailed classification of education. Therefore, decision was made after consultations with the MES to grant ISCED level 7 (a master's or equivalent degree) to persons who graduated 5 years' studies in HEE before 1991;

(c) As regards ISCED level 8 (doctoral or equivalent level), CSB of Latvia receives data from Latvian Academy of Sciences and in addition partly data used from the 2011 Census. Data quality is evaluated as good. Therefore, LFS figure can be recognized as underestimated;

(d) Sample surveys like LFS could indicate somehow wrong situation as people get partly confused about attribution of their education to the present education classification. Therefore, Census estimates in the accumulative education database could be recognized as more precise data source and they will be used to prefill data in statistical sample surveys, but still results will be evaluated for further methodological improvements as for Census as for surveys.

VI. Results

55. Within the project CSB have calculated a highest completed level of education for all resident population of Latvia on 1 January 2016 and on 1 January 2017 by educational attainment obtained from administrative data sources and statistical surveys of CSB. Provisional results of project are shown in Table 3 (before imputation) and Table 4 (after imputation).

Table 3
Educational attainment of resident population of Latvia on 1 January 2017

<i>Educational attainment (highest completed level)</i>	<i>Number of persons</i>	<i>%</i>
Total	1,950,116	
ISCED level 0 – Less than primary education	7,055	0.4
ISCED level 1 – Primary education	45,388	2.8
ISCED level 2 – Lower secondary education	263,563	16.0
ISCED level 3 – Upper secondary education	657,343	39.9
ISCED level 4 – Post-secondary non-tertiary education	219,920	13.4
ISCED level 5 – Short-cycle tertiary education	17,115	1.0
ISCED level 6 – Bachelor's or equivalent level	68,273	4.1
ISCED level 7 – Master's or equivalent level	217,236	13.2
ISCED level 8 – Doctoral or equivalent level	8,044	0.5
Not stated (of the persons aged 15 years or over)	23,124	1.4
ISCED 67 – Bachelor's/Master's or equivalent level	119,468	7.3
Resident population aged 15 and over	1,646,529	100.0
Not applicable (persons under 15 years of age)	303,587	

Table 4
Educational attainment of resident population of Latvia on 1 January 2017 after imputation

<i>Educational attainment (highest completed level) after imputation</i>	<i>Number of persons</i>	<i>%</i>
Total	1,950,116	
ISCED level 0 – Less than primary education	6,632	0.4
ISCED level 1 – Primary education	44,040	2.7
ISCED level 2 – Lower secondary education	258,078	15.7
ISCED level 3 – Upper secondary education	654,217	39.7
ISCED level 4 – Post-secondary non-tertiary education	214,380	13.0
ISCED level 5 – Short-cycle tertiary education	23,156	1.4
ISCED level 6 – Bachelor's or equivalent level	142,671	8.7
ISCED level 7 – Master's or equivalent level	295,311	17.9
ISCED level 8 – Doctoral or equivalent level	8,044	0.5
Not stated (of the persons aged 15 years or over)	-	-
ISCED 67 – Bachelor's/Master's or equivalent level	-	-
Resident population aged 15 and over	1,646,529	100.0
Not applicable (persons under 15 years of age)	303,587	

VII. Final conclusions and recommendations for future work

56. Results from collecting data about the highest completed level of education from different administrative data sources 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. This is one of the main sources for a determination of the highest completed level of education of resident population in Latvia, therefore an additional research to evaluate quality of these data is necessary, especially when the level of education is determined for younger persons. In co-operation with Mathematical support division of CSB for the missing data (persons aged 15 and over with unknown level of education) and for data that source is 2011 Census database, data correction and imputation have been done. By analysing the results, decision about data imputation necessity have been made. This work should be continued annually.

57. 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.

58. 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.

59. The accumulative education database on highest education level according to ISCED-A (2011) classification was updated with new information from administrative data sources like Register of Medical Persons and Medical Support Persons, CSB annual survey on the use of information and communication technologies in households, MES data on school pupils, Data files on students of some HEE. They contain information on the level of education obtained by these persons. Work will be continued on maintaining and updating of the accumulative database that contains information on persons' education from all available data sources.

60. Considering complexity of education system in Latvia and changes made in the system year by year it is recognized that respondents of the statistical sample surveys sometimes have problems to indicate exact level of education in accordance with the present classification. Therefore, Census estimates in the accumulative education database could be recognized as more precise data source to determine the level of education of the population as generalized sample survey data and they will be used to prefill data in statistical sample surveys, but still results will be evaluated for further methodological improvements as for Census as for surveys.

61. Use of data from accumulative education database to prefill education data for persons included in the samples for LFS and other regular statistical sample surveys will be continued. It is important to continue work on improving the quality of LFS and other regular statistical sample surveys' data, giving additional explanations to respondents about the differences between ISCED 4 and ISCED 5 as well between ISCED 7 and ISCED 8.

62. Collected data from administrative data sources and sample surveys is a good foundation for 2021 Census needs before the Register of Students and Graduates is introduced. As data from this Register will be available starting from 2018, it will become an important data source and no more data will be asked from each HEE, but still data checks should be done, and metadata should be checked to ensure data quality.

63. Additional information about higher education obtained abroad is expected from the External migration sample survey that was organized in 2018.
64. Work will be continued on identifying other possible additional data sources from which information about the level of education of the population could be obtained.
65. Results of this Grant project were presented and discussed during Twenty-first Baltic Census Seminar on 18th May 2018 in Vilnius, Lithuania.

Bibliography

Hasler, C. un Tillé, Y. (2015). Balanced k-Nearest Imputation. arXiv:1501.07622.
Pieejams: <https://arxiv.org/pdf/1501.07622.pdf>

Annex

Resident population aged 15 and over by the available information in data sources about their level of education on 1 January 2017*

<i>Data source</i>	<i>Number of residents</i>
The Population and Housing Census 2011 database	1,618,712
Ministry of Education and Science data on graduates for 2011–2016	176,152
Data files on graduates of HEE for 2010–2016	83,736
Ministry of Education and Science data on teachers' education for 2012–2016	49,946
Labour Force Survey data for 2014–2016	36,271
Ministry of Finance database for 2014–2016	17,451
EU_SILC survey database for 2014–2016	17,637
Adult Education Survey Data for 2015	5,717
2015 Population Microcensus database	5,505
Data of Latvian Academy of Sciences	5,362
Ministry of Education and Science data on the education of academic staff for 2012–2016	5,870
Latvian Council of Sworn Advocates data 2016	1,277
Study and Science Administration database for 2010–2016	363
Latvian Council of Sworn Notaries data 2016	171
Latvian Council of Sworn Bailiffs data 2016	126
Academic Information Centre data for 2015–2016	75
Survey on the use of information and communication technologies in households data for 2016	6,822
Ministry of Education and Science data on school pupils for 2012–2016	178,455
Data files on students of HEE for 2010–2016	161,242
Ministry of Health of the Republic of Latvia Health Inspectorate Register on Medical Persons and Medical Support Persons for 2016	27,332

* The data on educational attainment on individual level can overlap in several data sources.