United Nations ECE/CES/GE.41/2014/16

Economic and Social Council

Distr.: General
15 September 2014

Original: English

Economic Commission for Europe

Conference of European Statisticians

Group of Experts on Population and Housing Censuses organized jointly with Eurostat

Sixteenth Meeting
Geneva, 23-26 September 2014
Item 3 of the provisional agenda
Census methodology

The project for developing the methodology of register-based censuses in Estonia

Note by Statistics Estonia

Summary

Statistics Estonia has started preparations for the Register-based Population and Housing Census (REGREL), to be conducted in the next census round of 2020–2021. The first stage of preparations entailed running a project for developing the methodology of register-based censuses.

This paper gives an overview of the results and conclusions of the methodology project. The project was conducted in 2013, so some of the resulting suggestions have already been acted on (e.g. the Employment Register was established on 1 July 2014 with the amendment to the Taxation Act).

I. Introduction

1. Prior to the last census, in 2010, preparations were already started for the next population and housing census. An aim was set that the next census, held in 2020, will be register-based in Estonia.

2. The first stage of preparations entailed an extensive analysis, which was started in the autumn of 2010 and finished in September 2013.

3. In the project for developing the methodology of REGREL, 80% of which was funded by the European Social Fund, Statistics Estonia partnered with the Estonian Institute for Population Studies of Tallinn University, and the consulting firm Ernst&Young Baltic AS. The analysis stage involved a couple of dozen researchers, and experts both from Tallinn University and Tartu University, as well as lawyers and analysts from Statistics Estonia. A significant part was played by the representatives of administrative registers.

4. The analysis was done in two stages:

(a) meta-analysis of the obligatory characteristics of the population and housing census;

(b) detailed analysis of the characteristics that need to be analysed in terms of data quality, according to the results of the meta-analysis.

5. In addition to the meta- and detailed analysis, other analyses that support a register-based census were carried out:

(a) legal analysis;

(b) compilation of methodological guidelines for creating a glossary;

(c) analysis of international experience.

6. A total of 20 databases and the data collected in them were analysed.

7. The aim of the methodology project was to develop the methodology of REGREL with regard to the characteristics of individuals and dwellings, and to analyse the quality and coeffect of the registers that contain these characteristics. However, the project covered the overall aim of the population and housing census only partly, focusing on the obligatory topics presented in the Regulation No 763/2008 of the European Parliament and of the Council and in the Commission Regulation No 1201/2009 (Table 1). The actual censuses of the countries (incl. Estonia) are often much more extensive than the list of obligatory topics.

Table 1. Census characteristics analysed in the REGREL methodology project

Group of census characteristics	Characteristic
Census population or total population	Place of usual residence (total population)
Geographical characteristics	Location of place of work (locality)
Demographic characteristics	Sex, age, legal marital status
Economic characteristics	Current activity status, occupation, economic sector, status in employment
Educational characteristics	Level of education
Migration characteristics	Country/place of birth, country of citizenship, residence abroad and year of arrival in the country, previous place of usual residence and date of arrival in the current place of residence or the place of permanent residence one year before the census

Group of census characteristics	Characteristic
Household and family characteristics	Relationships between household members (household status, family status, type of family nucleus, size of family nucleus, type of private household, status of private household), tenure status of household
Housing characteristics	Housing arrangements, type of living quarters, occupancy status of conventional dwellings, type of ownership, number of occupants, usable floor space and/or number of rooms (density standard), water supply system, toilet facilities, bathing facilities, type of heating, type of building, period of construction

Note: the characteristics in the parentheses are not directly asked/recorded in the census but are derived on the basis of other characteristics.

Source: United Nations

II. Results of the methodology project

8. One of the most significant results of the methodology project was the network of main databases for REGREL. It is definitely not certain that there will be no changes in this network before 2020, but Table 2 lists the databases with which serious work has already begun and is continued.

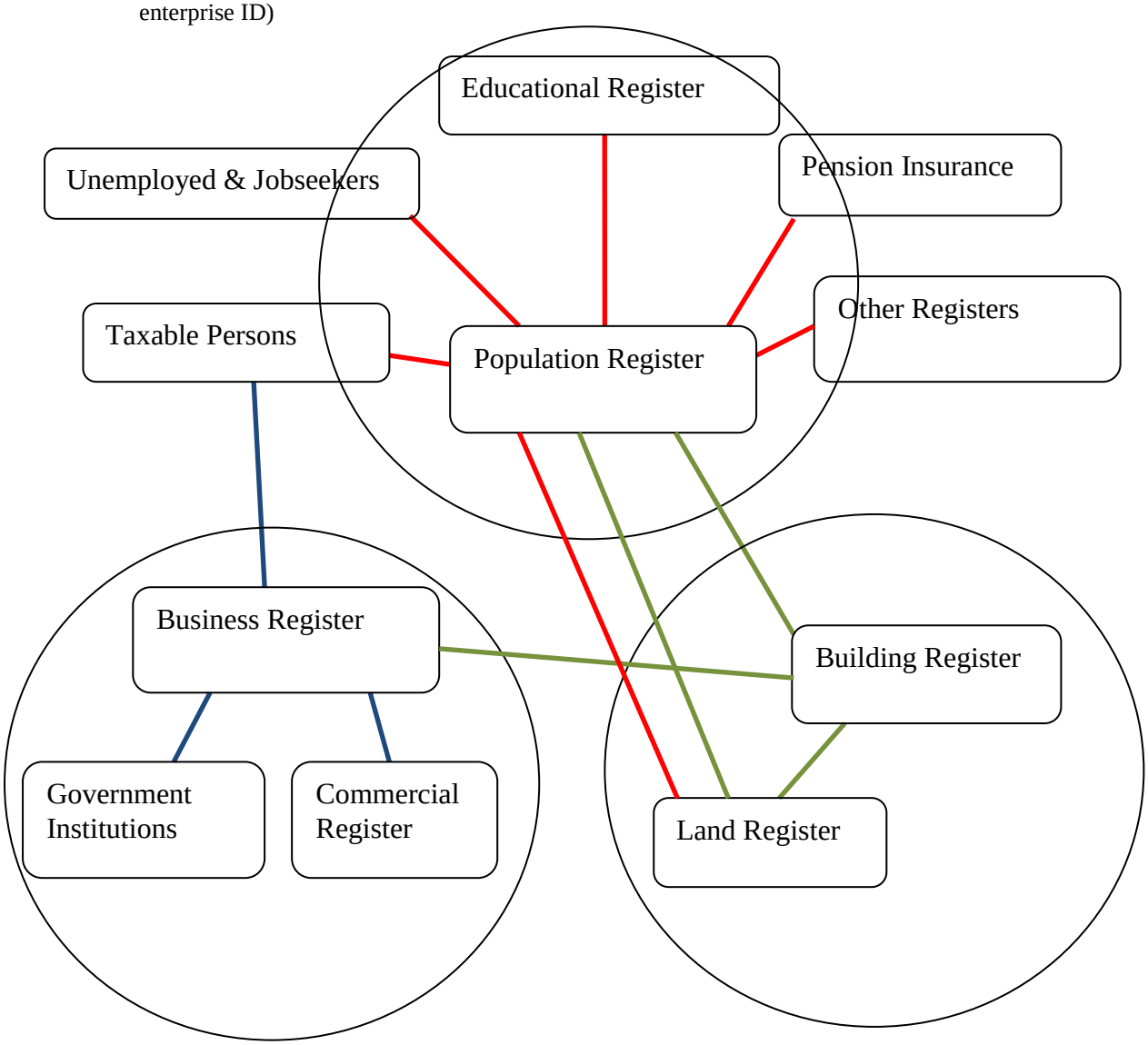
9. Data from different registers can be linked by using three forms of unique identification: personal ID, enterprise ID and address object ID (Figure 1). The newest form of unique identification is the address object ID. In the past few years, under the name of REGREL, the holders of several databases (including the Population Register, the Estonian Register of Buildings etc) have taken steps to link their databases to the Address Data System. Implementing an address standard in all registers would ensure that databases can be linked without any problems. By now, 95% of the addresses used in the Population Register are in line with the address standard, and can quite easily be linked to other registers.

Table 2. Administrative databases and registers for REGREL, 2013[a]

	RR	EHIS	EMTA	EHR	KR	ARIREG	RKOARR	STAR	KIRST	KVKR	PKR	EMPIS	VANGIS	KOPIS
Place of usual residence	X							X					X	
Previous place of usual residence	X							X					X	
Sex	X													
Age	X													
Legal marital status	X													
Country and place of birth	X													
Citizenship	X													
Year of entry in the country	X													
Relationships between household members	X													
Current activity status	X	X	X						X	X	X	X	X	X
Location of place of work			X[b]											
Occupation			X[b]											
Economic activity			X[b]			X	X							
Status in employment			X			X								
Level of education	X	X												
Occupancy status of conventional dwellings/type of ownership	X			X	X	X								
Data on housing	X			X	X	X								

[a] RR – Population Register; EHIS – Estonian Education Information System; EMTA – Register of Taxable Persons; EHR – State Register of Construction Works; KR – Land Register; ARIREG – Commercial Register; RKOARR – State Register of State and Local Government Agencies; STAR – Register of Social Services and Benefits; KIRST – Health Insurance Database; KVKR – National Defence Obligation Register; PKR – State Pension Insurance Register; EMPIS (Estonian Unemployment Insurance Fund) – Register of persons registered as unemployed or job-seekers, and of provision of labour market services; VANGIS – Register of Prisoners; KPR (KOPIS) – Register of Mandatory Funded Pension..

Figure 1. Registers and connections (red line – personal ID, green line – address object ID, blue line – enterprise ID)



10. The aggregated results of the meta- and detailed analysis of the methodology project are presented in Table 3. It is important to notice that these assessments were given in 2013. The situation is changing every year: the quality of administrative data is improving, changes have been made in registers, etc.

11. All these assessments rely on the assumption that only data from administrative sources are used to form census characteristics.

Table 3. Assessment: core characteristics

Compliance to census standard	Non-derived characteristics
A. Complete compliance	Sex, age, country of birth, country of citizenship
B. Partial compliance	Current activity status, status in employment, type of living quarters, floor space/number of rooms
C. Limited compliance	Educational attainment, legal marital status, place of usual residence, previous place of usual residence one year prior to the census, ever resided abroad and year of arrival in the country, relationships between household members, tenure status of households, 10 housing characteristics
D. Not available	Occupation, location of place of work (local activity unit), industry (local activity unit)

III. Roadmap to a register-based census

12. What follows is a shortened overview of the final conclusions of the REGREL methodology project, which provide a general roadmap to achieving a register-based population and housing census by 2020.

13. The conclusion of the methodology project was that it is possible to achieve a register-based census by 2020, provided the problems that were pointed out will be solved.

14. Here is an overview of the main steps that should be taken, in the opinion of the participants of the methodology project, to achieve a register-based census.

A. Changes in mentality

15. Developing a register-based census and, more generally, register-based statistics, requires, first and foremost, changes in mentality.

16. Firstly, it is necessary that registers, and the offices in charge of them, to take the needs of statistics into account, and not to consider them to be of lower importance in comparison with their administrative duties. It is important to acknowledge that with several concrete problems discovered during the REGREL project, distinguishing or contrasting administrative or statistical duties is basically unfounded because the basic census requirements – the universal inclusion of data objects, and the data being actualised and trustworthy – apply to registers as well as censuses. Thus, these needs are often not limited to the specific needs of censuses or statistics.

17. For Statistics Estonia, register-based statistics will entail a much more serious obligation to instruct registers on collecting statistical data, and to co-ordinate their activities, especially in terms of activities and methods that guarantee data quality.

18. Secondly, what needs a change in mentality are the administrative load and costs related to the register-based census. Although in the case of other equal conditions, retrieving census or statistical data from registers is cheaper and less of a burden on respondents than the separate collection of the same data for statistics, adjusting the current register system to the census requirements is rather labour-intensive and not at all cheap.

When developing a register-based statistical system, it should be acknowledged that general saving is unattainable without increasing investments and contributions to specific registers.

19. Thirdly, the culture and behaviour of registers and registering must change. Registers and register-based statistics will not work without citizens who understand the need of submitting data and who perform their duties. Improving the culture of registers and registering takes intentional raising of civic awareness and regular correspondence with the public on the topic.

B. Adding missing characteristics to the registers

20. The obligatory personal topics that are not covered by the registers are: occupation, economic sector and the workplace location (two of the latter defined by the local kind-of-activity unit).

21. With regard to occupation, the first suggestion is to add it to the data set of the employee register planned by the Estonian Tax and Customs Board. With this solution, occupation would be recorded for all employment relationships, which are added to the register, together with other relevant basic information (e.g. the type, start and end of an employment relationship, full- or part-time work, etc).

22. Regarding the workplace location and economic sector, the problem is that these data are not collected in the registers with the level of detail of local kind-of-activity unit. The information on local kind-of-activity units is recorded in registers regarding only a small share of legal persons who are operating in specific economic activities. There is also the problem that registers usually contain only the legal address, which may not be in accordance with the actual site of an enterprise or an institution. It is impossible to associate employed persons with a specific local kind-of-activity unit. Associating individuals is only possible on the level of an enterprise or an institution.

23. In order to form census characteristics, only proper recording of local kind-of-activity units is not enough. In addition to that, it is necessary for all persons employed to be associated with at least one local kind-of-activity unit. Since person-specific information on employment is recorded in the registers of the Estonian Tax and Customs Board, it would be the most efficient to associate persons with local kind-of-activity units based on these registers.

C. Improving the accuracy of data on permanent residence

24. The analyses of the REGREL methodology project confirmed once again that the main hurdle on the way to a register-based census is the inaccuracy of residence data in the Population Register. In addition to the accuracy of the permanent residence, the accuracy of all household and family characteristics, and that of all dwelling characteristics, depend on residence data (if a person's registered place of residence is inaccurate, the data of a wrong dwelling will be associated with him or her). Thus, all in all, inaccurate residence data have a serious effect on more than a half of census characteristics.

25. A separate organisational problem that needs solving is how the residence of institutionalised population is recorded in registers. At the moment, the register-based recording of detained persons has been organised fairly well through the Register of Prisoners. A suitable solution must also be found for collecting data that reflects staying in social welfare institutions.

26. A small, but a very difficult group to record based on registers is the homeless. To enter them in registers, creating a register of shelters should be considered. It would collect

data from the entire country about all homeless persons who have used accommodation services in shelters.

27. Several recommendations were aimed at nationally improving the accuracy of the registered place of residence. The issue of undocumented emigration could be handled with the help of expanding data exchange with the population registers of foreign countries. In 2004, regular data exchange was introduced between the Estonian and Finnish population registers, which has significantly improved the accounts of changes in residence that happen between the two countries. It is advisable to extend a similar practice to other countries as well.

D. Entering archive data into registers

28. In the case of several register-based census topics (e.g. marital status, relationships between household members, residence abroad and year of entry into the country, level of education), the insufficient item response rate was caused by the register lacking data on earlier events.

29. Among the characteristics of a register-based census, the formation of which is currently limited, entering archive data into a register would be of help with regard to the level of education. Entering earlier leaving certificates into a register would enable creating a document-based education database as accurate as the Estonian Education Information System (EHIS) up to about 80% of the inhabitants of Estonia.

30. Among the dwelling-related characteristics of a register-based census, entering archive data into a register would be of help in the case of buildings that are missing from the Estonian Register of Buildings (EHR), and with regard to the building year of dwellings. If an additional analysis determines from which period the buildings are missing from the register that they belong to, and if the majority prove to be older buildings, it is worth digitalising these data according to the archive materials of the Building Register.

E. Recording relationships between household members in the Population Register

31. The analysis carried out during the REGREL methodology project showed that no register in Estonia collects comprehensive data on households and families.

32. At the same time, the Population Register contains most of the information (permanent residence, relations between spouses and between children and parents) that is necessary to create register-based characteristics of households and families. This fact significantly decreases the costliness of realising the recommended improvement, since it only requires making a simple IT-related improvement to 85–90% of the register. Additional information would be needed about an estimated 10–15% of the persons. The need for additional information can prove to be even smaller in reality because, besides the relationships between spouses and those between children and parents, the Population Register also enables to determine kinship of somewhat greater degree (e.g. grandchild-grandparent). Because of the aforementioned reasons, it would be sensible to supplement the dataset of the Population Register by adding the relationships between the members of a dwelling-based household. More specifically, the Population Register needs the creation of new characteristics which describe the relations (spouse, child, parent, sister-brother, grandchild, grandparent, other relative, non-relative) between the persons living in one dwelling or, respectively, in a dwelling-based household.

F. Improvements in recording data on persons arriving from abroad

33. With regard to several census topics (marital status, level of education, information on children and parents, which is the basis for the characteristics of household and family), the analysis highlighted problems with registering data on persons arriving from abroad. For example, marital status was not recorded in the Population Register in the case of approximately one-half of those citizens of EU Member States who are more numerous in Estonia (Finland 58%, Sweden 55%, Great Britain 53% et al.).

34. In order to receive these data, the recording of data should be improved for foreign persons taking up residence in Estonia, and the submission of missing data (marital status; level of education; information regarding children and parents, which is the basis for the characteristics of household and family – the list is incomplete) in a reasonable amount of time should be made obligatory.

G. Refining register data on earlier arrivals in the country

35. According to the REGREL analysis, the quality of register data is lower than the average in the case of foreign-born persons who have arrived in Estonia before the Population Register was created. Obtaining data (marital status; level of education; information regarding children and parents, which is the basis for the characteristics of household and family; year of arrival in Estonia) from them is more complicated in comparison with new arrivals. In the case of earlier arrivals, data collection cannot be joined with the first application for a residence permit or with submitting a notice of residence, and requires a separate procedure.

36. In order to bother these persons less, and to disperse the burden of data submission over a longer period of time, the refining of register data could be joined with renewing identity documents or with cases, where a person addresses a vital statistics official or a register employee for some other reason.

H. Improvements related to data on dwellings

37. Such solutions need to be developed for the State Register of Construction Works (EHR) that would guarantee the coverage and relevance of the register.

38. EHR also needs data quality projects which would identify the objects with missing and controversial data (incl. double objects) in the register. In some cases, the removal of problems may require addressing the owners or possessors; for this, supplementary resources should be designated for EHR and/or for local governments.

39. As for the organisation of housing and the type of dwelling, it is important to specify the data on private households and institutional households living in common dwellings.

40. In the case of the dwellings' technical data, the analyses ascertained that the rate of accordance with surveys is low. Such a situation can suggest that the data in EHR is out of date. For EHR, simplifying the procedure related to processing permits (time, complexity, and costliness) should be considered. If an address object exists, submitting technical data in a simplified way should be considered for dwellings.

41. With regard to dwellings' technical data, the classifications (e.g. water supply system, toilet facilities) used in EHR need to be revised as well, and supplemented if needed, based on the needs of register-based statistics and the population and housing census.

I. Improving the compatibility of registers, more systematic data quality checks, refining classifications and better documentation of register data

42. With several registers, the analyses of the REGREL methodology project highlighted various technical problems, the majority of which were not directly linked to the register-based census or statistics. Recounting them in detail would take up too much space, which is why only a few more important ones are mentioned.

(a) Implementing an address standard in all registers, which would ensure that databases can be linked without any problems.

(b) More efficient implementation of data monitoring.

(c) Creation of a continuous classification for Estonian room units, which would enable making the location data collected according to earlier administrative divisions compatible with later structures.

(d) Better documentation of register data.

J. Comparison of register-based characteristics with the respective characteristics of the 2011 census

43. During the REGREL methodology project, it was not possible to make all the necessary comparisons between the register-based characteristics and the respective characteristics of the 2011 census (the opportunity to use detailed data from the census emerged only in the final stage of the project). The comparison of register-based characteristics and the characteristics of the 2011 census cannot be based on the default presumption that if a discrepancy occurs, reality is reflected more exactly by the results of a survey-based census.

44. In the case of a discrepancy, the answers given in both sources should be critically analysed (in order to determine the distribution of mistakes between the sources, the analysis might require contacting the respondents of the census based on the sample).

K. Switching from census to register-based statistics: corresponding development works in Statistics Estonia

45. The main aim of the REGREL methodology project was to determine the possibilities for the register-based creation of the obligatory European Union census characteristics. Besides focusing on the census as one of the most large-scale and costly statistical actions, it is important that introducing register data into statistical use would be as extensive as possible.

46. Just like the Nordic countries, Estonia should start using register data more than before when doing continuous statistics in the various sectors of population and social statistics (e.g. register-based employment statistics, register-based household and family statistics, register-based education statistics, register-based housing statistics, register-based income statistics, etc.).

L. Statistics Estonia co-operating with registers, research institutions and experts

47. The completion of the methodology project does not signify readiness for a register-based census, but is a thorough mapping of the current readiness in terms of obligatory census characteristics. The list of further tasks shown above was quite long. The key role and responsibility in performing these tasks lies with Statistics Estonia. In order to develop register-based statistics, Statistics Estonia will have to show more initiative and greater competence than before.

48. The REGREL methodology project, however, clearly indicates that the efforts of Statistics Estonia alone are not enough to achieve the aim. The success of register-based statistics, including the population census, requires more extensive co-operation between offices, and developing a corresponding general mindset. In order to guarantee co-operation with registers, Statistics Estonia needs effective support from the Ministry of Finance, but also from other ministries.

IV. Conclusions

49. Considering all of the above, the prerequisites for a register-based census can be summed up as follows:

(a) There is a common decision and agreement to use the data collected in administrative databases to cover the needs of statistics.

(b) Permanent residence data will be corrected.

(c) Missing characteristics will be added to the registers.

(d) Recording data on foreign arrivals (incl. earlier arrivals) in the registers will be improved.

(e) The coverage and relevance of housing data will be improved.

(f) The compatibility of registers will be improved, data quality will be checked and improved systematically, classifications will be improved and register data will be documented better.

50. The results of the REGREL methodology project show that in order to achieve a register-based census, there is still plenty to be done. Although the activities leading to a register-based census need to be guided and co-ordinated by Statistics Estonia, the focus of the necessary changes lies outside the statistical office, concerning registers on the one hand, and respondents, people and enterprises on the other.

51. Everything depends on how effectively the various census characteristics achieve such readiness that the register-based creation of characteristics guarantees the former completeness and accuracy of census data.

V. References

REGRELi metoodika väljatöötamine. Projekti lõpparuanne (2013). A. Puur, L. Sakkeus, S. Aben, http://www.stat.ee/regrel (Final report of the methodology project, available only in Estonian)

Annual report 2013.(2014).Statistics Estonia, http://www.stat.ee/annual-report

Roadmap to a register-based census. D. Matteus. Quarterly Bulletin of Statistics Estonia 4/13. Direct link http://www.stat.ee/65376

(Re)designing administrative data – towardsregister based census. Estonian experience. D. Matteus, K. Lehto. (Paper presented at Q2014 in Vienna in 2014) Direct link http://www.q2014.at/fileadmin/user_upload/Q2014_paper_Matteus_Lehto_20140515.pdf

Opportunities and challenges of register-based census: Observations based on methodology development for the 2020 round census in Estonia. A. Puur, L Sakkeus (Presentation at 25th Conference of the Estonian Statistical Society in 2013). Direct link http://www.stat.ee/public/yritused/ess-25/1/12.Puur_Sakkeus.pdf