



Economic and Social Council

Distr.: General
22 September 2016

English only

Economic Commission for Europe

Conference of European Statisticians

Group of Experts on Population and Housing Censuses

Eighteenth Meeting

Geneva, 28 - 30 September 2016

Item 5 of the provisional agenda

Methods for assessing quality and usability of registers and administrative sources

Outline for UNECE Guidelines on the use of registers and administrative data for population and housing censuses

Note by the Task Force on Register-Based and Combined Censuses¹

Summary

As a follow-up to the in-depth review of diversification of population census methodology and sources in October 2015, the Bureau asked the Secretariat to set up a Task Force on register-based and combined censuses. In this note the first activity of this Task Force is described: an outline of the new guidelines (draft 5 September 2016). The terms of reference for the Task Force on register-based and combined censuses can be found in Annex A.

¹ Prepared by Eric Schulte Nordholt, Chair of the UNECE Task Force on Register-Based and Combined Censuses.

1. Between 2012 and 2015 the UNECE Steering Group on Population and Housing Censuses coordinated the preparation of the Conference of European Statisticians Recommendations for the 2020 Censuses of Population and Housing. Nine Task Forces were working together with the Steering Group on the Recommendations.
2. The CES adopted the Recommendations for the 2020 census round in June 2015. The Recommendations are available in electronic format on the UNECE website² and are printed in English and Russian in 2016.
3. In October 2015, the CES Bureau conducted an in-depth review of diversification of population census methodology and sources, based on a paper by Finland and Turkey (ECE/CES/BUR/2015/OCT/3) and a note by UNECE (ECE/CES/BUR/2015/OCT/3Add.1). As an outcome of the review, the Bureau supported the preparation of new guidelines on the use of registers for population and housing censuses, and requested the Secretariat to prepare new terms of reference for the Steering Group on Population and Housing Censuses and for a Task Force on Register-Based and Combined Censuses (Report of the Bureau meeting: ECE/CES/BUR/2015/OCT/21).
4. This note presents the proposed outline of the new guidelines produced by the Task Force on register-based and combined censuses. To prevent confusion in the further work of the Task Force some of the basic concepts have been worked out. After the presentation of this progress report in Geneva in September 2016 a start will be made with the work on the complete draft of the new guidelines on the use of registers and administrative data for population and housing censuses

I. Scope of the new UNECE Guidelines and definitions of register-based and combined censuses

5. The scope of these new UNECE Guidelines is not on traditional censuses, but on register-based and combined censuses. Therefore, only definitions of register-based and combined censuses are given. More information about traditional censuses can be found in the Recommendations for the 2020 census round. In these new UNECE Guidelines different kinds of registers (on persons and buildings) are mentioned with a focus on those used in censuses.
6. For some of the definitions we can refer to those presented in the publication “Register based statistics in the Nordic countries”³. On page 15 of that publication a register is defined as a systematic collection of unit-level data organized in such a way that updating is possible. Updating is the processing of identifiable information with the purpose of establishing, bringing up to date, correcting or extending the register, i.e. keeping track of any changes in the data describing the units and their attributes. Administrative data sources are data holdings that contain information collected primarily for administrative (not research or statistical) purposes. This type of data is collected by government departments and other organizations for the purposes of registration, transaction and record keeping, usually during the delivery of a service. They include registers and possibly other administrative data without a unique identifier. Statistical registers are registers created for statistical purposes. They are typically created by transforming data from registers and/or

² <http://www.unece.org/publications/2020recomm.html>

³ http://www.unece.org/fileadmin/DAM/stats/publications/Register_based_statistics_in_Nordic_countries.pdf

administrative data sources. This transformation is often required to enable the registers or administrative sources to meet statistical definitions.

7. In some countries administrative data are used as a synonym of register data. In other countries a distinction is made. If administrative data are data collected for administrative purposes it is challenging for this data to be linked to other administrative data. However, sometimes this can be achieved by linking through a set of identifiers (e.g. name, sex, date of birth, numerical address) resulting in matches that can then be used to produce outputs similar to those produced through linking registers that do contain a unique identifier. If such successful links (i.e. links that are of good quality and thus not burdened by too many errors) are not possible, it is technically infeasible for a country to move to a combined or register-based census, but the administrative data may then be used for benchmarking purposes (see Annex B for an example in Ireland).

8. The register-based population census system is built around a set of basic registers that contain comprehensive data on the units that are to be described in the population and housing census (see page 26 of the Recommendations for the 2020 census round). The main challenge in some countries is the linking of data records from different registers. Some countries do not have a unified identification system, meaning that all relevant registers are using the same unique identifiers, preferably id numbers. Some register-based census countries miss some of the census variables in all of the available registers and choose to support their census by data from already existing microdata from sample surveys. All register-based census countries have in common that no census questionnaires are used to collect information about the population. Therefore, register-based censuses are in general much cheaper than combined censuses and especially than traditional censuses.

9. England and Wales are exploring the potential for moving to an “Administrative Data Census” after 2021⁴. Such a move will require a combination of

- Record-level administrative data held by Government;
- A population coverage survey;
- A population characteristics survey.

10. In a combined census statistics are created by using registers and other administrative sources, together with information from either sample field data for selected variables or full field enumeration for selected variables (see pages 12 and 24 of the Recommendations for the 2020 census round).

II. Quality of registers and administrative data to be used in censuses

11. There is an important difference between the quality of linkage with and without unique keys. If there is a unique identifier in all or most of the records the linkage is easy and the quality of the linkage is normally good. However, if such a unique identifier is absent, the quality of the linkage should be measured, and the impact on the resulting outputs should be assessed.

12. The quality in official statistics can be described on the basis of two kinds of quality:

⁴ See

<http://www.ons.gov.uk/file?uri=/census/censustransformationprogramme/administrativedatacensusproject/administrativedatacensusannualassessments/annualassessmentofonssprogressstowardsanadministrativedatacensuspost2021.pdf>

- Product quality (relevance, accuracy, timeliness and punctuality, comparability and coherence, accessibility and clarity)
 - Process quality (best methods, cost efficiency, low response burden).
13. In a quality framework⁵ the following three kinds of quality can be studied further:
- Quality of sources (supplier, relevance, privacy and security, delivery, procedures)
 - Quality of metadata (clarity, comparability, unique keys, data treatment)
 - Quality of data (technical checks, accuracy, completeness, time-related dimension, integrability)
14. A quality framework can typically be used to study different potential sources and to decide which sources are going to be used in the census. Finally, it should be mentioned that the ESS.VIP-ADMIN project is reviewing current practices on assessing the quality of administrative sources. This is of course relevant in a census context if these administrative sources are candidates to be used as input in the census.

III. Necessary conditions, advantages, disadvantages, opportunities and threats for register-based and combined censuses

15. P Necessary conditions (compare with chapter 3 of register-based statistics in the Nordic countries)
- a) Legal base (Statistics Act)
 - b) Public approval ('Big Brother is watching you')
 - c) Stakeholder approval (stakeholders typically want to have at least the same kind and detail of information as in the last census; if a change in census methodology will lead to a different financial distribution over the parts of a country this should be explained very carefully)
 - d) Cooperation between the statistical office and other authorities (mainly government organisations)
 - e) Comprehensive and reliable register system (administrative versus statistical quality)
 - f) Unified identification system (preferably unique ID-numbers; for countries where unique ID numbers for persons do not exist the ability to link data efficiently and accurately is a particular challenge)
 - g) Knowledge (a sound knowledge on the administrative sources available has to be built up in a country that moves from a traditional towards a combined or register-based census; although building up knowledge can be done en route, the efforts needed to make this process successful should not be underestimated).
16. Advantages

⁵ See e.g. <http://www.cbs.nl/nl-NL/menu/methoden/onderzoek-methoden/discussionpapers/archief/2009/2009-42-x10-pub.htm> and http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.41/2012/use_of_register/WP_16_Austria.pdf

- a) Smaller costs per inhabitant (in register-based censuses and in combined censuses without full field enumeration for selected variables)
- b) Quicker to conduct (shorter production time in register-based censuses as no field enumeration has to be conducted)
- c) Less problems with nonresponse (in case only or mainly registers are used, see also under opportunities)
- d) Possibility of a permanent (i.e. more regular) census (in case of annual census updates as a by-effect it is easier to keep the knowledge and IT-infrastructure up-to-date, see also under disadvantages)

17. Disadvantages

- a) Differences in concepts and definitions of censuses compared with registers and administrative data (full coverage of population may not be available on all administrative sources)
- b) Publication of small subpopulations sometimes difficult or even impossible because of limited information in already existing or census sample surveys (range and detail of outputs may be more limited than from a traditional census or a combined census with full field enumeration for variables missing in registers)
- c) How to keep knowledge and IT infrastructure up-to-date for a process that is running only a relative short number of years every ten years? (see also under advantages)

18. Opportunities

- a) Contribute to lower response burden (see also under advantages)
- b) Contribute to more efficient work with a lower budget
- c) Better relations with register holders (if they are open to that and conditional on the legal base; in some countries there is also a potential to improve the relations between register holders by introducing or extending register-based statistics)
- d) Better cooperation of units within the office (by moving from a survey organisation towards a register-based statistical office the traditional 1 to 1 relationship between sources and statistics is being replaced by a m to n relationship; by better integrating statistics the coherence in the statistical framework is improved)
- e) More time and money remains for innovations
- f) More flexible and responsive to new information needs
- g) New output, data and services can be made available

19. Threats

- a) Dependent on register holders (statistics is not their priority): failures in registers may cause failures in statistics
- b) Timeliness of registers
- c) Different reference dates in different sources (not all sources give the option to distinguish clearly between reference dates and dates of events, moreover even if that can be handled properly, the risk is that not all sources can be harmonised towards the same reference day → what is an acceptable difference in reference days?)
- d) Privacy and security concerns (often relating to linking different sources)
- e) Attention for the census (results) diminishes if people no longer fill in census forms

IV. Transition from traditional censuses to register-based and combined censuses

20. In this section some possible approaches for how to move from traditional censuses to combined and register-based censuses are presented. These examples are based on the experiences of countries that have moved, or plan to do so. Other countries can benefit from these experiences (see also under necessary conditions in the previous section).

21. To illustrate how transitions can take place and what kind of problems one has to overcome in practice some examples of recent moves towards combined and register-based censuses are given.

- Case study Austria
- Case study Estonia (see Appendix C)
- Case study Poland
- Case study Slovenia
- Other examples could be added

22. It is important to inform users of moves towards register-based censuses as such moves may have an impact on output availability. If there are planned moves to a different census methodology, it is good practice to be transparent share information on plans and tests as much as possible.

V. Essential features and quality of register-based and combined censuses

23. In this section information is given on essential features of register-based and combined population and housing censuses. Register-based censuses largely comply with the following five essential features defined by the UNECE.

- Individual enumeration
- Simultaneity
- Universality (within a precisely defined territory of a country)
- Small area data
- Defined periodicity

24. The quality of the output of register-based and combined censuses can be assessed in different ways.

- Quality and confidentiality issues (among other things the question is whether the quality of the output can be measured; in case sample surveys are used minimal cell frequencies have to be defined to guarantee that table estimates are accurate enough to publish, in case a country only relies on registers or complete enumeration confidentiality rules have to be defined to prevent disclosure of individual entities)
- Quality reports (in which results can be compared with e.g. LFS output; in the census context one typically finds that differences regarding demographic variables are relatively small and differences regarding economic variables (e.g. current

activity status) are relatively large; what is the impact of such quality reports on future censuses?)

- Quality committees
 - Internal quality committees
 - Audits based on quality guidelines
 - External audits (to check privacy or peer reviews)
- Quality standards.

25. Finally, the ESS.VIP-ADMIN project should be mentioned again. The goal of this project is to help statisticians to make wider and better use of administrative sources in the production of official statistics. It is done by addressing the most typical challenges faced in the use of these sources: limited access to data, the lack of quality of sources, methodological issues related to the processing of data and the integration of several sources, e.g. estimation when combining sources. It also aims to ensure that the statistics produced using administrative data are comparable and are of sufficient quality by providing tools for assessing the quality of outputs based on administrative sources.

VI. Relevant literature

To be included at a later stage.

TERMS OF REFERENCE FOR THE TASK FORCE ON REGISTER-BASED AND COMBINED CENSUSES

I. BACKGROUND

1. In October 2015, the CES Bureau conducted an in-depth review of diversification of population census methodology and sources. As an outcome of the review, the Bureau supported the preparation of new guidelines on the use of registers and administrative data for population and housing censuses, and decided to set up a Task Force for this purpose (Report of the Bureau meeting: ECE/CES/BUR/2015/OCT/21).
2. This note presents the terms of reference for the Task Force on Register-Based and Combined Censuses.

II. MANDATE

3. The Task Force reports to the UNECE Steering Group on Population and Housing Censuses, which in turn reports to the Conference of European Statisticians and its Bureau. The Task Force is created for a period from 2016 to 2018.

III. OBJECTIVE

4. The objective of the Task Force is to prepare new *UNECE Guidelines on the use of registers and administrative data for population and housing censuses*, covering operational, practical, technical and legal aspects. The new guidelines should cover the use of multiple sources (including the so-called combined censuses) and include methods for measuring coverage and quality of the various sources, without specific focus on the comparability of the various census methods.

IV. PLANNED ACTIVITIES AND OUTPUTS

5. The Task Force is expected to prepare a set of Guidelines on the use of registers and administrative data for population and housing censuses. The planned output is a publication in electronic and printed format, in English and possibly in Russian (subject to availability of resources), to be finalized by April 2018.
6. The Task Force will start its work by collecting available material on the use of registers and administrative data for population and housing censuses. In particular, the Task Force should consider the publication "Register-based statistics in the Nordic

countries”⁶, the relevant text in the census recommendations for the 2020 round, and the papers prepared in the last years for UNECE expert meetings on censuses, describing the preparation and conduction of register-based and combined censuses⁷.

7. The Task Force should also establish cooperation with other relevant international activities, including in particular those conducted under the ESS.VIP.ADMIN project (part of ESS Vision 2020) on the assessment of the quality of administrative sources.

8. The available material should be reviewed, expanded and integrated with new text, in order to produce a comprehensive document covering the following topics (the list is tentative and not exhaustive):

- Definitions of register-based and combined censuses, and description of the methods more commonly adopted in the 2010 and previous rounds;
- How to assess the coverage and quality of registers and administrative data, and how to decide if and how they can be used for the population and housing census;
- Advantages, disadvantages, and necessary conditions for register-based and combined censuses;
- Transition from traditional census to combined or register-based censuses;
- Methods for assessing the quality of register-based and combined censuses.

9. The Task Force should use as much as possible the information available on national census practices, availability and use of data from registers and administrative sources, and plans for the 2020 round, including in particular the information collected by UNECE in the 2013 and 2015 surveys on population and housing censuses. If necessary, the Task Force may collect, through the UNECE, additional information from member countries, minimizing the burden on national statistical offices.

10. The Guidelines should present good practices identified by the Task Force that could be used as examples by other countries. The Guidelines may also include additional recommendations to complement those already provided in the CES Recommendations for the 2020 censuses.

V. TIMETABLE

Feb 2016	Establish the Task Force and draft detailed work plan with activities, timing and division of work
Mar-Apr 2016	Collect and review available material
Apr-July 2016	Prepare outline of the Guidelines, and first draft chapters
Sep 2016	Present progress report at UNECE-Eurostat Expert Meeting on Censuses

⁶ http://www.unece.org/fileadmin/DAM/stats/publications/Register_based_statistics_in_Nordic_countries.pdf

⁷ <http://www1.unece.org/stat/platform/display/censuses/Collection+of+papers+on+transitions+to+new+census+methods>

Oct 2016-Jun 2017	Revise draft and prepare complete draft Guidelines
Sep 2017	Present draft Guidelines at UNECE-Eurostat Expert Meeting on Censuses
Sep-Dec 2017	Revise draft Guidelines based on discussion at Expert Meeting
Dec 2017	Submit final draft Guidelines for review by CES Bureau in February 2018
Feb 2018	Review of draft Guidelines by CES Bureau
Feb-Mar 2018	Electronic consultation among all CES members
Apr 2018	Expected endorsement by the CES plenary session

VI. METHODS OF WORK

11. The Task Force is expected to work in English only, mainly by e-mail and telephone conferences. The UNECE Census Wiki platform can be used by the Task Force to share work material, develop documents, and possibly organize discussion forums.

12. Face-to-face meetings of the Task Force could be organized in connection with Joint UNECE-Eurostat Expert Meetings on Censuses, or other events attended by a significant number of members of the Task Force.

VII. MEMBERSHIP

13. The following countries and organizations participate in the work of the Task Force: Austria, Canada, Estonia, Germany, Ireland, Israel, Italy, Netherlands (Chair), New Zealand, Norway, Poland, Republic of Korea, Slovenia, United Kingdom, United States, Eurostat, FAO and UNECE. The secretariat is provided by UNECE.

Annex B

This note is based on a paper "The Irish Statistical System and The Emerging Census Opportunity" presented at NTTS 2015.

Contents

- Environment
- Census background
- Census 2021 intentions
- Census 2026 intentions
- Useful references

Environment

CSO, Ireland has made significant progress with developing administrative data sources for statistical purposes. This work is underpinned by the developing national data infrastructure (NDI) in Ireland. In summary, NDI is a conceptual framework promoting more efficient use of data across the Irish public sector. It promotes and advocates for the use of common official identifiers for persons, business and property across all official systems (the Nordic model). A central population register with up to date information on where everybody lives is not envisaged.

There exists a person identification number that is generally used across all person based administrative systems (schools, welfare, employment, etc.) that is assigned at birth or when a person enters the country to live. This identification number is stored on a master file or register. This register is maintained and updated by the department of social protection (DSP - government ministry that oversees social welfare administration); however there is no requirement for organisations to provide updates to the DSP.

Recently, a project has been undertaken to summarise each person's annual activity on key public administration systems with a simple yes/no indicator in a Person Activity Register (PAR). The PAR employs a Protected Identifier Key (PIK) to reduce privacy risk while preserving the linking possibilities over time and across data sources. Key administrative data sources include births, children's benefit, education (early, primary, secondary, higher and further), employment, unemployment, occupational pensions and social welfare (including state pension). The purpose of the Par is threefold:

- to enable longitudinal analysis of population cohorts across different administration systems.
- to explore population structures over time.
- to provide a summary master key with respect to different administrative data sources in order to examine the feasibility of different potential projects.

The potential to provide for longitudinal analysis of specific population cohorts has proven to be of significant value in promoting the concept of "joined up government needs joined up data" across the Irish public sector. In the strict sense of the definition of a register the PAR is not a register and is simply a population frame.

There are a number of official identification numbers in use for businesses on the main tax, employment and company registration systems. These numbers are linked and available on the CSO business register for exploiting data sources for statistical purposes. The tax authorities also have these numbers linked. Currently, it is planned to expand the use of official identification numbers for businesses across the public sector.

A linked employer employee statistical file has been developed based on employer employee tax returns. This file captures every paid employment in the state and facilitates linking of business and person based pillars of NDI.

A new postcode system has recently been rolled out in Ireland. The postcode uniquely identifies each letter box in the state. At present the use of the postcode is not mandatory. It will likely take a number of years for this postcode system to become a unique reference for properties available on all official systems. In Ireland over 30% of address strings are not unique and are only distinguishable by using a person's name and the postman's local knowledge of who lives where to deliver the mail. The postcode system and its underlying register of properties is the basis of a sampling frame and master address frame for social surveys and census operations (2011 and 2016).

The obvious gaps in data sources for conducting a census using administrative data relate to population coverage and where the population resides.

Census background

Ireland typically has high population movements and migration flows that necessitate it conducting a census every 5 years (4.6 million persons living in 1.6 million houses in 2011). Ireland typically carries out a traditional 'de facto' census in that approximately 4,000 enumerators hand deliver a census form to every household in the state to be filled in by the head of household with respect to every person present on census night. The forms are then picked up by the enumerators to be compiled and collated centrally by the CSO. In 2011 the census cost €55m or €12 per person living in the state. In 2016, experience found that it is becoming harder to contact each household.

Census 2021 intentions

The intention in 2021 is to include an internet response option as part of the traditional census. This could happen as follows: census forms or letters with a url, unique id and pin will be delivered to each property (by hand or post) and the head of household will be asked in the first place to complete the form online (in the first instance), return the form by post or give it back to the enumerator when they call. There is an intention to use administrative data sources to assist with person/household non response.

Census 2026 intentions

It is not known how feasible a census based on administrative data sources and existing surveys would be in 2026. But if Ireland were to conduct such a census it might be undertaken along the following lines.

Step 0. Ensure an up to date master address file. This master address file would also collate building characteristics from previous censuses and administrative data holdings to assist in the housing component of the census. There may also be a possibility to identify the occupancy status (vacant/occupied) of houses based on the records of utility companies.

Step 1. Identify all persons interacting with public administration systems for the census year and summarise this on a register against a person identification number. Let's call this register the person activity register (PAR). The PAR will contain a list of all persons that are active in a given year i.e., 'signs of life' approach. The key administrative data sources in the Irish context will include births, children's benefit payments, primary school database of enrolments, higher and further education sources of enrolments and awards, social welfare, employment including self-employment data sources from tax authorities, persons registered on property rental leases, persons claiming medical benefits (or registering in a given year), social welfare and occupational/state pensions.

Step 2. Use statistical methods to correct for under coverage and attach a correction factor to each record. The statistical methods may rely on existing sample surveys or another second independent source. One such method is currently being explored where driver licence renewals are used as a second independent source to identify under coverage in the person activity register using capture-recapture methods.

Step 3. Allocate each identified person with an address/postcode from the master address file. Use a suitable decision tree algorithm that is capable of incorporating situations where more than one address is identified with a person in administrative data sources. Household relationships identified in administrative data sources (e.g., children's benefit) may also feed into this decision tree algorithm.

Step 4. Form household relationships using relationships identified in administrative data sources and persons identified as living at the same house.

Step 5. Estimate and include attributes for each person on the PAR using existing surveys, administrative data sources and appropriate methodologies.

Step 6. All census outputs are now compiled from the PAR as updated by the above steps.

References

MacFeely and Dunne, "Joining Up Public Service Information: The Rationale for A National Data Infrastructure" Administration, Vol. 61, No. 4 (2014), pp. 93–107 http://ipa.ie/pdf/forum_vol_61_4.pdf.

Dunne, "The Irish Statistical System and The Emerging Census Opportunity" NTTS/SJIAOS 2015.

Annex C

Estonian case study

During the 2000s, Estonia worked very productively on systematic development of registers, incl. identifying the data in all personal registers on the basis of personal identification codes and linking them with the X-way system to facilitate exchange of information. However, an initial analysis indicated that Estonia's registers were not yet ready for a successful conduct of a population and housing census in 2011. The main reasons for this were as follows:

- at least 20% of the addresses specified in the Population Register were not the actual places of residence of the people concerned;
- the Education Information System only contains data on young people (general and higher education diplomas from 2000 onwards; vocational and other certificates from an even later date);
- no register contained information on the occupations of persons;
- addresses have been recorded differently in different registers, with a variable degree of specificity, making the data incompatible;
- registers had been used only for a short period of time and their quality and adequacy had not been verified;
- the consistency of definitions used by different registers and information technological compatibility of registers had not been analysed.

The final decision on unfeasibility of a register-based census in this census round was made with the support of the Estonian Academy of Sciences. On 9 September 2008, the Census Committee of the Government of the Republic approved a combined methodology as the method to be used in this census. It is also important that, at the same time, the Government of the Republic was working towards improvement of the registers in Estonia, motivated also by a need to harmonise the records of spatial positioning of objects. Therefore, an address data system (ADS)⁸, incl. an address-standard, was approved in 2008 to harmonise the use of addresses. The Spatial Data Act⁹, adopted in 2011, also harmonised the use of spatial data.

The circumstances surrounding the census were finalised with the adoption of the Official Statistics Act¹⁰ on 10 June 2010, replacing the earlier Census Act. The envisaged census methodology was also approved by the Scientific Council of the Population and Housing Census.

Combined census methodology was approved for the Census 2011.

Meaning for the combined census methodology:

1. Coordination of data sources. Previously created data sources, i.e., registers are used in censuses alongside with face-to-face interviews. In PHC 2011, registers were used in three ways: as a tool for preparing the census (preparation of work lists and census sheets), pre-filling of questionnaires, and supplementation of

⁸ Address Data System. (2007). Government of Estonia Regulation. Riigi Teataja, part I, No 71, art. 439. [www] <https://www.riigiteataja.ee/akt/12901083> (22.04.2013).

⁹ Spatial Data Act. (2011). Riigi Teataja, part I, No 2. [www] <https://www.riigiteataja.ee/akt/128022011002> (22.04.2013).

¹⁰ Official Statistics Act. (2010). Riigi Teataja, part I, No 41, art. 241. [www] <https://www.riigiteataja.ee/akt/13332259> (22.04.2013).

census results in case of missing data. The information on the studies of enumerated permanent residents was taken from EHIS the Information System of Education and was not included in questionnaires.

2. Combined survey methodology. Unlike previous Estonian censuses, self-completed questionnaires were used in 2011 as a method alongside interviews. This required compilation of extensive instructions and provision of comprehensive guidance and training to the persons enumerated, as well as the enumerators.
3. Combined data collection methodology. All previous censuses in Estonia have been conducted using paper questionnaires, or census sheets. The data of persons covered by the census – the answers to the census questions – were entered on these paper or cardboard sheets by enumerators, using a writing instrument, e.g., a special machine-readable pencil.
4. Two new technologies were introduced in PHC 2011: self-completion of questionnaires on the internet and entry of answers directly in laptop computers during census interviews. The option of using paper questionnaires was kept as a backup for emergency situations. The possibility of telephone interviews was also foreseen for particularly exceptional circumstances, especially in cases when there was a need to specify the answers to a questionnaire or a census location was extremely difficult to access (e.g., on a small island).

The background of register-based censuses in Estonia

Already in 2010 preparations began for a register-based census. Here, the experience of those countries who had already conducted a register-based census was taken into account. So a project REGREL was started to develop a methodology for a register-based population and housing census.

Preparatory works during the period of 2010-2016

2010-2013 REGREL methodology project

2010-2011 meta-analysis of obligatory PHC characteristics

2012-2013 detailed analysis of characteristics that required data quality analysis (as indicated by the meta-analysis)

2014 REGREL pilot census

2016 Census trial

REGREL methodology project, 2010-2013

In 2010 Statistics Estonia started preparations for the transition to register-based population and housing censuses (the REGREL project). The first stage of the preparations was extensive analysis, which began in autumn 2010 and was completed in September 2013.

The REGREL methodology project (with about 80% of the project funded by the European Social Fund) was a partnership between Statistics Estonia, the Estonian Institute for Population Studies (at Tallinn University) and the consultancy AS Ernst & Young Baltic. The analysis was carried out by a few dozen scientists and experts from the University of Tartu and from Tallinn University, by lawyers and by analysts from Statistics Estonia. A very important role was played by the representatives of databases and registers, who took an active part in the process.

The analysis was carried out in two parts:

- meta-analysis of obligatory PHC characteristics;

- detailed analysis of characteristics that required data quality analysis (as indicated by the meta-analysis).

In addition to the meta-analysis and detailed analysis of characteristics, the team also made other preparations for register-based censuses:

- legal analysis;
- preparation of methodological guidelines for the creation of a census glossary;
- analysis of international experience and practice.

All in all, the project team analysed nearly 20 registers and the data collected by these. One of the most significant outcomes of the methodology project was the network of main registers and databases for REGREL (containing data on the obligatory characteristics).

The results of the REGREL methodology project showed that there is still much to do to prepare for register-based censuses. Statistics Estonia will manage and coordinate these activities. However, the most important work is to be done outside of Statistics Estonia – this means registers, on the one hand, and the respondents, residents and enterprises, on the other hand.

Census pilot in 2014

The main goal of the REGREL pilot census was testing of the production system with selected mandatory census characteristics (EU Regulation no 763/2008):

1. Place of usual residence
2. Sex
3. Age
4. Legal marital status
5. Country and place of birth
6. Citizenship
7. Relationships between household members
8. Level of education

Action plan during the period of 2016-2020

The main preparatory work for register-based census is related to the following tasks:

- Data acquisition from registers (contracts, description of the data set, checks on data quality and the acquisition procedure);
- Formation of census characteristics, programming of the necessary rules;
- Testing the statistical system as a whole;
- Testing statistical registers system and filling it with data 2015-2018;
- Currently, the registers do not cover the census characteristics in the entire population, and it is also unclear whether data are updated fast enough.

Statistics Estonia has developed a set of legal and organisational measures to improve the quality, timeliness and coverage of the data sets necessary for census. The set of measures are submitted to the ministries.

Currently, the registers do not cover the census characteristics in the entire population, and it is also unclear whether data are updated fast enough.

2016 Census trial

Period of pilot census activities: 02.01–08.12.16

Trial census population:

- Usual resident population of Estonia;
- Dwellings regardless of occupancy and occupied non-conventional dwellings located in Estonia.

Objective: to practice conducting the register-based population and housing census (REGREL) following Eurostat's quality requirements and rules on censuses. For this, data must be acquired from 22 national databases, followed by processing and analysis.

During the next census trial in 2018 the preparations for the register-based census are focused on datasets and data flow.

In Estonia, preparations have to be made for the register-based census, in the course of which data will be captured from various databases following the rules of data protection and statistical security. The requirements developed for databases are sufficient for ensuring the interoperability of state information systems if:

- All data presented are submitted with metadata, including classification codes;
- Capture and data updates take place via the X-Road service;
- Data are presented in XML format and the description of data will be submitted by the creator of the X-Road service as XSD and updates include the time of presentation.

The primary data in the databases are required to meet the quality requirements in order to guarantee that census objectives are fulfilled. Quality is indicated by the following:

1. The coverage of the registers needs to be at least 97% for the population and 95% for single characteristics;
2. 95% of the data need to be linked with the classifications registered in RIHA (Administration System for the State Information System).
3. All legal and private persons and citizens of foreign states need to be assigned a code.

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 – towards register 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

The project for developing the methodology of register-based censuses in Estonia (2014). K. Lehto. (Paper presented at UNECE meeting - Group of Experts on Population and Housing Censuses - in Geneva in 2014) Direct link

http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.41/2014/mtg1/WP16_s3_Estonia.pdf
