

(Re)designing administrative data – towards register based census. Estonian experience

Doris Matteus, Kristi Lehto*

***Statistics Estonia**

Abstract: The paper gives an overview of the experience of Statistics Estonia regarding cooperation with administrative registers and redesigning administrative data in the process of preparing for the register-based census

Introduction

Statistics Estonia (SE) has begun preparations for conducting a register-based Population and Housing Census (REGREL) in the next census round. At this point, Statistics Estonia's goal is to use the data of administrative registers to the greatest extent possible.

The first stage of preparations entailed an extensive analysis. A total of 20 databases and the data collected in them were analyzed. The representatives of administrative registers played a significant part in the project. During the next stages of the preparations, cooperation between SE, register holders and Estonian Information System Authority will be even more intense. SE has been given a significant role in coordinating the preparations and also redesigning administrative data in order to meet statistical needs. Several other ongoing developments support this goal:

- Obligation to implement an address standard in all registers.
- Better documentation of register data.
- Data exchange layer X-Road, which enables secure Internet-based data exchange between the state's information systems.

Preparations for register-based population and housing census (PHC)

Statistics Estonia started preparing for a register-based census in 2010. The first stage of preparations entailed an extensive analysis which was started in the autumn of 2010 and finished in September 2013.

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, and also lawyers and analysts from Statistics Estonia. A significant part in the project was played by the representatives of administrative registers.

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 coefficient 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 Regulation (EC) No 763/2008 of the European Parliament and of the Council and in Commission Regulation (EC) No 1201/2009. The actual censuses of the countries (incl. Estonia) are often much more extensive than the list of obligatory topics.

The analysis was done in two stages:

- meta-analysis of the obligatory characteristics of the PHC;
- detailed analysis of the characteristics that need to be analysed in terms of data quality according to the results of the meta-analysis.

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

- legal analysis;
- compilation of methodological guidelines for creating a glossary;
- analysis of international experience.

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

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.

(Re)designing administrative data – the background

Describing the preparations for the register-based census, the general background has to be taken into account, as register-based census in Estonia is not considered as a project belonging uniquely in the domain of Statistics Estonia.

The general background in Estonia can be described as rather favorable.

1. Pursuant to the Official Statistics Act [1], all data collected in administrative databases is available to SE for producing official statistics. Furthermore, the act stipulates that when producing official statistics, SE has the obligation to primarily use data collected in administrative records and databases. Of course, the prerequisite of using administrative data is that such data has to allow the production of official statistics in compliance with the quality criteria of official statistics.

During the production of official statistics, SE has the right to establish statistical registers of data by collecting the same data from various administrative records and databases and other data sources. The data collected in a statistical register for the performance of statistical actions may be used for the production of other official statistics, regardless of the purpose for which the data were initially collected. SE also has the right to link micro-data collected from respondents and administrative records and databases by using characteristics that allow direct or indirect identification of statistical units. Of course, all along the process the confidentiality of the data has to be protected.

2. According to the Public Information Act [2], establishing separate databases for the collection of the same data is prohibited in Estonia. The processing of data which are collected as basic data by another database belonging to the state information system has to be based on the basic data of the other database. This fact narrows the possibilities for data quality comparisons and checks but also helps to improve the quality of the data collected. The more the data are used for different purposes, the better the quality due to stricter and more continuous checking.

3. Before the establishment of a database or changing the composition of the data collected in a database, introducing a database or terminating a database, the technical documentation of the database has to be approved among other institutions also by Statistics Estonia.

Furthermore, the producer of official statistics has the right to make proposals for amending the composition of data and the classifications used in the administrative records and databases, if the coverage of data and the composition, level of detail and quality of data in the administrative records and databases do not allow the production of official statistics complying with the quality criteria of official statistics.

4. The Public Information Act establishes several support systems to the state information system, including the classifications system, the system of address details, the data exchange layer of information systems and the administration system of the state information system.

The use of those support systems is mandatory for all state and local government databases.

An important support system is RIHA, which is the administration system for the state information system. RIHA serves as a catalogue for the state's information system. RIHA guarantees the transparency of the administration of the state's information system and helps to plan the state's information management.

RIHA gives information on the following subjects:

- Which are the information systems and databases that make up the state's information system;
- Which data are collected and processed and in which information systems;
- Which services, incl. X-Road services, are provided and who is using them;
- Who are the responsible and authorised processors of the information systems and databases, and who are the contact persons;
- On which legal basis are the databases operated and the data processed;
- The reusable components that ensure the interoperability of information systems (XML assets, classifications, dictionaries and ontologies).

RIHA also serves as the procedural and administrative environment for the administration of reusable components, including classifications, dictionaries and ontologies.

5. According to the law, all exchange of data with the databases belonging to the state information system and between the databases belonging to the state information system has to be carried out through the data exchange layer of the state information system.

Data exchange layer X-Road is a technical and organisational environment, which enables secure Internet-based data exchange between the state's information systems.

According to the Public Information Act, the exchange of data with the databases belonging to the state information system and between the databases belonging to the state information system has to be carried out through the data exchange layer of the state information system. The X-Road allows institutions/people to securely exchange data as well as to ensure people's access to the data maintained and processed in state databases.

In the case of citizens, the X-Road enables using the services of the X-Road via different portals. That includes making enquiries from state databases and to check the information related to the person himself/herself. Such micro-level checking helps to improve the quality (accuracy) of data in administrative databases.

6. The Spatial Data Act [3] stipulates that all state and local government databases have to be integrated with the central Address Data System (ADS). The Address Data System is a support system for the maintenance of databases with the aim to ensure a unique identification of the address objects both at their location and in different databases, and to make geographical addresses that have been submitted in different times and according to different principles comparable.

The Land Board has created a central address data management system (ADS Management System) through which all address data throughout the country will be managed and the majority of address services will be offered. The system became fully operational on January 1, 2009.

Unique identification of the address object was the final link to create a registers and connections system for the register-based PHC. Unique identification for persons and enterprises was used already before.

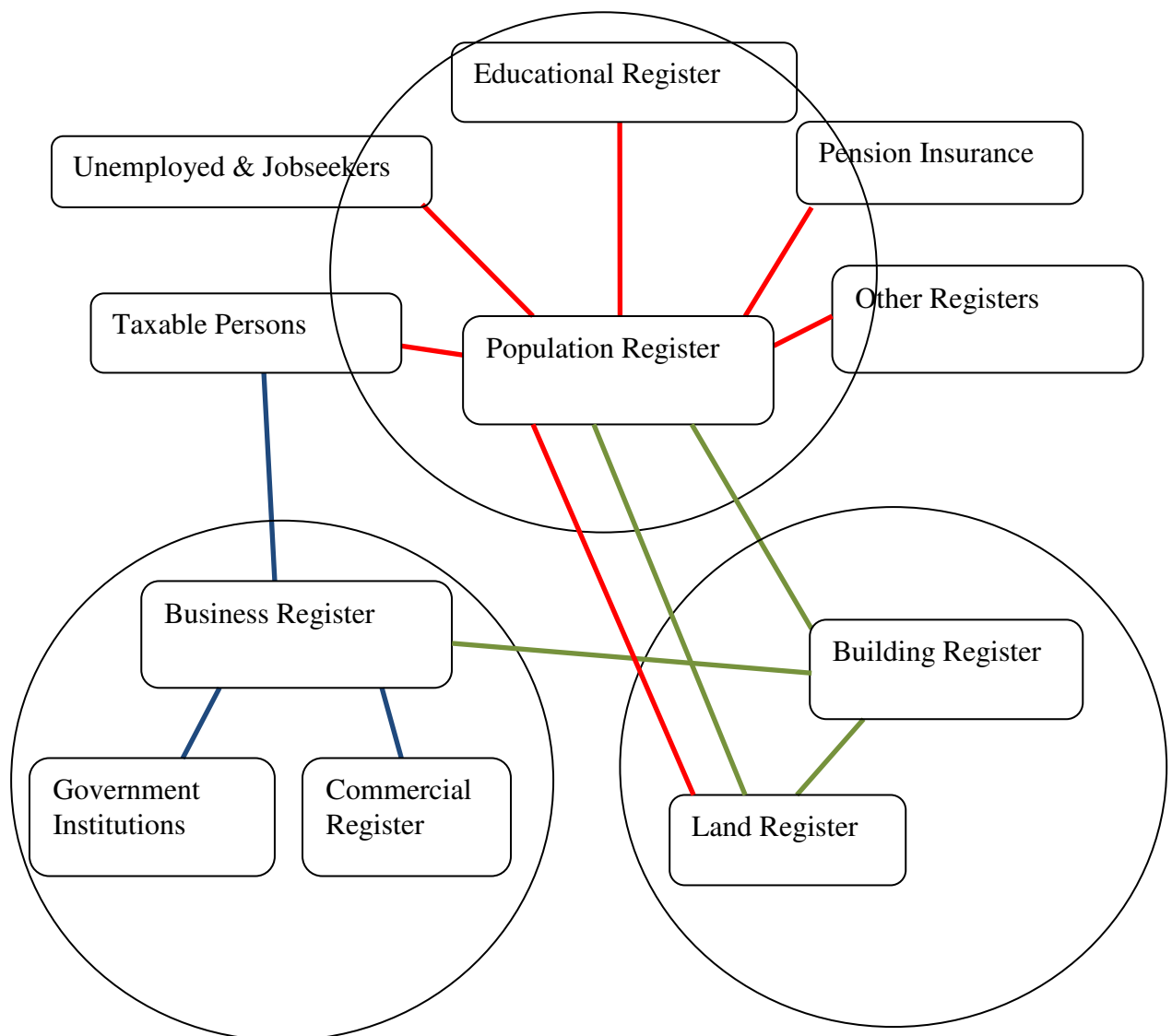


Figure 1 Registers and connections (red line – personal ID, green line – address object ID, blue line – enterprise ID)

(Re)designing administrative data – experience

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

Firstly, it is necessary that register-holders would take the needs of statistics into account from the beginning, already in the process of designing and redesigning the registers. In fact, contrasting administrative and 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, the needs of censuses or statistics are often not so specific. Of course, 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.

Experience gained from co-operation between the statistical office and register-holders in recent years has also proven that the needs of statistics and administrative purposes can be successfully combined and constructive co-operation between state agencies maintained.

Example 1: adding missing characteristics to the registers

As mentioned before, SE has the right, stipulated in the law, to make proposals for amending the composition of data and the classifications used in the administrative records and databases. This right has been even more reinforced in the transition process to the register-based PHC. Namely, the Official Statistics Act states that for the transition to the register-based population and housing census, the government will approve a list of necessary actions by administrative records and databases together with the terms of action. Although the mentioned list has not yet been approved, several actions have already been taken in order to improve the usability of register data for census purposes.

There are two main obligatory topics that are not covered by any registers: occupation and workplace location (local kind-of-activity unit).

With regard to occupation, the first suggestion of the methodology project was to add the variable 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).

Regarding the workplace location, the problem is that these data are not collected in the registers with the level of detail of the 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.

As the above-mentioned Employee Register is still in the development process, Estonian Tax and Customs Board involved SE in the process. Although the two mentioned variables will not be collected about all employees at the beginning, the technical readiness to do so will be developed. The register will be fully integrated with the ADS and the full official classification of occupations ISCO (International Standard Classification of Occupations) will be used, although this is not essential for tax purposes.

Example 2: taking into account statistical needs

During the preparations for the transition to the register-based PHC, several instances of constructive co-operation between register-holders and SE have occurred.

First of all, register-holders have acquired the habit of co-ordinating their planned improvements and changes in registers with SE. Although, as mentioned before, the establishment of a database or changing the composition of the data, introducing a database or terminating a database have to be approved among other institutions also by SE, the official approval is not enough to ensure compliance with the needs of official statistics. A better practice would be active and proactive co-operation and involvement.

There are several good examples of active involvement:

- The Estonian Register of Buildings has taken into account the suggestions of the methodology project in developing a new register. The suggestions concerned both the composition of data and the classifications (e.g. water supply system, toilet facilities etc) used to collect the data.
- The Register of Prisoners also consulted SE, before starting to develop a new register.
- SE was involved in the process of developing a new register for the Social Insurance Board.

Example 3: the use of the ADS

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 ADS. SE has been actively involved in the process. 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.

Example 4: classification

With several registers, the analyses of the REGREL methodology project highlighted various technical problems. One of the problems concerns the classifications used in the registers. In recent years, in several registers, the classifications used have been synchronised.

SE is the holder and administrator of all classifications and therefore has also the responsibility to advise the registers. The last example of this work is the adoption of ISCED2011 (International Standard Classification of Education) in the Population Register. The Population Register also recoded all their education data.

Summary

The REGREL methodology project clearly indicated 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.

(Re)designing administrative data in order to meet statistical as well as administrative needs is a long and continuous process, which requires rethinking many practises, including the essence of co-operation between state agencies. The goal has to be set and agreed in close co-operation between all register-holders and other relevant state agencies in order to ensure long-term success. This requires serious efforts to be made mainly by SE, the body responsible for official statistics and the PHC. Rethinking the usual processes and practices will be a long and demanding journey.

References

[1] Official Statistics Act

<https://www.riigiteataja.ee/en/eli/ee/Riigikogu/act/526032014002/consolide>

[2] Public Information Act <https://www.riigiteataja.ee/en/eli/514112013001/consolide>

[3] Spatial Data Act <https://www.riigiteataja.ee/en/eli/ee/Riigikogu/act/514112013024/consolide>