

**Economic and Social Council**Distr.: General
29 August 2016English
English and Russian 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**Securing Interoperability of State Information Systems for
the Register-Based Census in Estonia****Note by Statistics Estonia¹***Summary*

In Estonia, registers currently cover nearly all obligatory census characteristics. As the next step, Estonia wants to develop a means of data capture from registers that is based on such an information-technology solution that would ensure a data collection process that is automated to the maximum extent. The prerequisites for automatic data capture are the data format and quality.

In the business process, cost savings increase when data collection and organization of register data are no longer done file-based and manually, but are rather harmonized across the various databases.

This presentation introduces the need to ensure the interoperability of information systems for the register-based census in Estonia.

¹ Prepared by Diana Beltadze.

I. Introduction

1. In the past decade, automating business activity through information systems has caused the accumulation of data i.e. the so-called data revolution. The current aim is to create interoperable information systems which will enable the collection, mining, processing and organized presentation of necessary information.
2. This document will introduce the need to ensure the interoperability of information systems for the register-based census in Estonia. Interoperability is defined as the ability of disparate and diverse organizations to interact towards mutually beneficial and agreed common goals, involving the sharing of information and knowledge between the organizations².
3. The document covers the issue of ensuring interoperability and the possibilities for solving it, and Statistics Estonia's key role in achieving the necessary result for the next census. Other topics covered include the obstacles to ensuring interoperability of information systems for the register-based census in Estonia and the strategy which will help solve the problem methodically.
4. The lessons learnt from the two trial censuses (in 2014 and 2016) bring out the most important: in order to carry out the register-based census, data quality has to be improved, semantic capability must be enhanced and standards must be used in cooperation with partners. Ensuring the interoperability of information systems is easier if administrative information systems implement state-created reinforcing systems such as address data system or classifications system.

II. What kind of interoperability is needed for the register-based census?

5. New technologies reach users fast, and the volume and importance of data are increasing in many fields. The interoperability of information systems in producing statistics should be examined in connection with information integration in order to put emphasis on the relative increase of the volumes of information and information management in an information society³. Managing an increasing volume of information, transmitting it quickly and without losses and at the same time ensuring its integrity, manageability and quality have become the challenge for new technologies and an indicator of interoperability on the organisation level⁴. The topic of need of interoperability is generally relevant in the two following cases:

- (a) Data exchange that is based on ICT systems⁵;
- (b) Heterogeneity of information systems⁶.

² Vallner, U. (2009). Registrate koosvõime. Eesti Statistikaaltsi 21. Konverentsi ettekanne.

³ Klischevski, R. (2004). Information Integration or Process Integration? How to Achieve Interoperability in Administration. Electronic Government Volume 3183 of the series "Lecture Notes in Computer Science", pp. 57–65.

⁴ Vallner, U. (2009). Registrate koosvõime. Eesti Statistikaaltsi 21. konverentsi ettekanne.

⁵ Majandus ja Kommunikatsiooniministeerium (2011). Riigi Infosüsteemide koosvõime raamistik. [<http://www.riso.ee/et/koosvõime/raamistik>] 20.06.2016.

⁶ Ouksel, A., Iqbal, A. (1999). Ontologies are not the Panacea in Data Integration: A Flexible. Coordinator to Mediate Context Construction. Distributed and Parallel Database, 7, pp. 1-29.

6. In the case of the production of statistics, interoperability will provide the possibility to create more effective and economic solutions. In organizing censuses, the biggest costs are those connected to data collection. One way to reduce these costs is connected to the possibility of using register data.

7. The State of Estonia uses central information systems and registers which enable getting data about persons and dwellings⁷. As the next step, the interoperability of state information systems will have to be ensured by the census year 2020 and from there on.

8. Estonia wants to develop a means of data capture from registers which is based on such an information-technological solution which would ensure a data collection process that is automated to the maximum extent. For this, the following requirements must be met:

(a) All the Estonian addresses in databases originate from the Address Data System administration system of the Estonian Land Board;

(b) All the persons in the databases have been identified with a personal identification code (in the case of natural persons) or Commercial Register, Non-Profit Organizations and Foundations Register (Commercial Register) code (in the case of legal persons);

(c) The national data security policy requires that data capture via the data exchange layer X-Road will be used.

9. Statistics Estonia has presented the owners/administrators of registers a list of activities that need to be done based on assessments on the quality levels of databases and has also set deadlines by when the shortcomings should be eliminated. Statistics Estonia has announced that for the register-based census, data will be collected from registers following agreed data compositions, which need to meet the requirements of producing statistics.

10. For Statistics Estonia, the topic of interoperability of information systems is first and foremost connected to the protocols of data transfer, software, syntax, semantics, and structure of IT systems. The issue of ensuring the interoperability of information systems in producing statistics is closely connected to the expectation that with it, long-term sustainability is ensured for more statistical actions than just the census. Capability will be ensured by the following factors when they have been agreed upon with different parties⁸:

- Rules;
- Standards;
- Procedures;
- Methods;
- Common or compatible information systems;
- Cooperation networks.

III. Information integration as the first challenge

11. An information system can be considered an instrument that has a structure which is based on rules and the functioning of which is directed at the logic of gathering, storing and

⁷ Tiit, E-M. (2015). Ettevalmistused registripõhiseks loenduseks. Eesti Statistika Kvartalikirj nr 3.

⁸ Zhao, K. & Xia, M. (2014). Forming Interoperability Through Interorganizational Systems Standards. Journal of Management Information Systems /Spring 2014, Vol. 30, No.4.

releasing information⁹. Man is part of the information system and to achieve synergy, the logical compatibility of the organization's structure with information systems and connection with production processes has to be ensured¹⁰. If you change any component in the information system, the organization where people use the information system can change as well, therefore the quality indicators of the information system have to be assessed with regard to integrity, accessibility, continuity, coherence, accuracy and timeliness¹¹. Besides assessing the quality of the information system, it is imperative, from the census statistics production aspect, to ensure data quality, e.g. accuracy, correctness and timeliness¹². Low data quality can affect the interoperability of different information systems, but at the same time the compatibility of information systems might not necessarily cause problems¹³. For example, in census statistics, metadata that have not been updated can cause problems in data processing, but data transfer from one information system to another remains smooth. The questionnaire survey conducted in 2016 by the database working group showed that the issue of data quality is central in ensuring interoperability. It can be claimed that when conducting a census, data quality indicators have an impact on the quality of the information system, and vice versa.

12. If it is practical to enable information systems to exchange data for the register-based census, the possibility that the system is a collection of objects which work in the name of common goals that has information on the input and output of operations and the communication between systems takes place through an interface should be taken into account¹⁴.

13. In such a case, in the context of moving data it is important to ensure data quality in time: timeliness, right data adjusted for the census moment. Data have to be unambiguous, reliable and in a suitable format. Here, preparation activities have to be divided into two categories: work with information and work with systems¹⁵.

14. For the register-based census, cooperation has to be established with the information system owners or database administrators which are under the responsibility of the following ministries.

- Ministry of the Interior (RR – Population Register; ETR – Register of Residence and Work Permits; current VANGIS, future KIR – Register of Prisoners and Probationers; KMAIS);
- Ministry of Education and Research (EHIS – Estonian Education Information System);
- Ministry of Finance (EMTA – Register of Taxable Persons and TÖR – Register of Employment; KPR (KOPIS) – Register of Mandatory Funded Pension, SAP);

⁹ Mereste, U. (1987). Süsteemkäsitlus: Süsteemsest mõtlemisviisist majandusnähtuste käsitlemisel. Tallinn: Valgus, 264.

¹⁰ Hiie, I. (2014). IT-süsteemide üllahoid. Loengumaterjal. Tallinna Ülikool.

¹¹ Alberts, D. S., Garstka, J. J., Hayes, R. E., Siguari, D. A. (2001). Understanding Information Age Warfare. CCP, 312.

¹² United Nations Economic Commission for Europe (UNECE) (2006:17-90). CES Recommendations for the 2010 Censuses of Population and Housing. New York and Geneva: United Nations. [[www](http://www.unece.org/fileadmin/DAM/stats/publications/CES_2010_Census_Recommendations_English.pdf)]http://www.unece.org/fileadmin/DAM/stats/publications/CES_2010_Census_Recommendations_English.pdf

¹³ McGinnes, S., Kapros, E. (2015). Conceptual independence. A design principle for the construction of adaptive information systems. Information Systems. Vol. 47, pp. 33-50.

¹⁴ Rava, K. (2000). Eeldusi infosüsteemide projektide õnnestumiseks. A & A, nr 1.

¹⁵ Rava, K. (2000). Eeldusi infosüsteemide projektide õnnestumiseks. A & A, nr 1.

- Ministry of Economic Affairs and Communications (EHR – State Register of Construction Works);
- Ministry of Justice (KR – Land Register; ARIREG – Commercial Register; VANGIS (KIR) – Register of Prisoners and Probationers (in the future: RKOARR – State Register of State and Local Government Agencies; e-File));
- Ministry of Social Affairs (STAR – Register of Social Services and Benefits; PKR – State Pension Insurance Register);
- Estonian Health Insurance Fund (KIRST – Health Insurance Database);
- Ministry of Defence (KVKR – National Defence Obligation Register);
- Estonian Unemployment Insurance Fund (EMPIS – register of persons registered as unemployed and jobseekers, and of provision of labour market services);
- Ministry of the Environment (Address Data System of the Estonian Land Board);
- Estonian Road Administration (Traffic Registry).

15. We determined the factors which obstruct the establishment of a stable data transfer. In 2016 they are the following:

(a) The documentation on the interfacing of information systems is not sufficiently detailed (contrary to the statistics producer's view);

(b) Data composition descriptions are very labour-intensive and their coordination is time-consuming;

(c) Instead of data, excerpts are transferred to the database of the register (e.g. State Register of Construction Works), which requires the later selecting of information based on needs;

(d) The adaptation of changes (updates) requires constant monitoring and manual intervention;

(e) On the technical level, there have been volume limitations (Estonian Education Information System) when transferring data via the data exchange layer X-Road.

16. Therefore, to improve the situation, more emphasis has to be put on work with information on the organization level. It is in the interests of the data provider to have a clear understanding of the format in which information has to be transferred to the recipient. For the integration of information, data transfer has to be improved. For that, the data transfer process has to be divided into independent components¹⁶, after which the independent components have to be identified and analysed from the aspect of production capacity (e.g. LEAN methodology).

17. Cooperation has to be organized between several institutions; here, the movement of information and communicative integration on the social level are of critical importance. The problems of the information integration of information systems stem from the history of the creation of registers. Estonian registers are new and they were not created for producing census statistics.

18. The information integration necessary for the census is based on the capacity of systems to co-function, exchanging information, content and format for that purpose. Such

¹⁶ Infotehnoloogia sõnastik, Eesti Standard EVS-ISO/IEC 2382

a capacity requires the register administrators' interoperability in the field of semantics¹⁷. Once automatic data capture is achieved, the interoperability capacity of information systems will also be created.

19. In the business process, cost savings accrue when data collection and organizing of register data are no longer done file-based and manually, and when it is not different for every database. Transfer to a register-based census in statistics requires, on the organization level, the creation of an environment which would enable increased productivity, the use of various dynamic information sources, and access to accurate and relevant information.

20. Therefore, the objective of information integration – data flow and ensuring access to structured information – is complicated because the production environment is becoming more complex, heterogeneous, it involves existing systems, semantics, technologies etc. The key issue in this new situation is managing the heterogeneity, the respective interoperability, ensuring it on a semantic, syntactic, structural, system level.

21. Is it possible to formalise the information integration process? The challenge is to take into account the impact originating from the reasons why one or another register was created. It must be taken into account that in the case of a census we are dealing with many information sources which are constantly being developed due to the updating of hardware and software platforms. When ensuring interoperability for the census it is important to keep in mind the socio-cultural aspect of the process, i.e. the ability to be flexible and understand the cooperation partner's development needs and capacity.

IV. Process integration

22. Process performance and ability to create new services is central to the interoperability process integration. Solutions need to be sustainable and support increasing the volume of producing current statistics based on registers by optimising working methods and technologies. The objective is to produce census data at least twice as fast than in the previous census, i.e. collect and publish all data within a year.

23. Process integration requires the management of services. The management of interfaces is easier when there is configuration management for separate components in systems and processes. It is useful to monitor the directing of messages between services and possibilities of interpreting the service.

The requirements for the interfaces of information systems can be universal:

- All the interfaces of the information system have to be realised in the application layer as web services;
- All contacts with the interfaces have to be securely authenticated and authorised;
- The interfaces have to be documented following the valid model;
- External interfaces need to be transported with the descriptions of WADL or WSDL services.

24. Process integration strategy could follow the information integration strategy. That is a strategy which in Estonia is connected to the statistical institution's strategic plan's measure to speed up the publication of statistics and to reduce the administrative burden.

¹⁷ Ouksel, M., Sheth, A. (1999). Semantic Interoperability in Global Information Systems. A brief introduction to the research area and the special section. SIGMOD Record, Vol. 28, No1, pp. 5-12.

25. In the first stage, in the period 2016–2018, the goal is to create the ability of information integration for the census.

V. Automatic data capture as an indicator of state information systems' interoperability

26. Ensuring the interoperability of information systems in the data collection process of the register-based census requires determining certain principles that will form basis for capturing data from registers.

27. Automatic data capture can work in two methods (pull or push), a universal data format is a prerequisite. However, the trial census experience shows that despite tests, the possibility of manually transferring data (excerpts of data) has to remain.

28. On the technical level, before capturing data, the data composition of the census has to be described in the metadata for every register and also the register variables that cover the entire capturable dataset have to be described. Register variables must be connected to classifications, indices and dictionaries. The second step would be to create XSD on the push-method, save and assign a unique URI. The method enables automatic feedback on data quality, i.e. errors. The pull-method application would be the opposite.

29. There are various architectural patterns of transmitting data. Upon transmitting data, Statistics Estonia uses a data architecture model which takes into account the function, concept and format of data¹⁸. To plan the following data capture activities, requirements in four categories have been developed for register administrators:

(a) Requirement for the primary capture of data: all data have to be presented with metadata, including the classification codes;

(b) Requirements for adopting updates;

(i) All presented data must come with the metadata, including the classification codes;

(ii) Pursuant to the agreement between Statistics Estonia and the chief processor or authorised processor of the database, Statistics Estonia will adopt updates either through the X-Road service created by the chief processor or authorised processor of the database, or the chief processor or authorised processor of the database will do it him-/herself using Statistics Estonia's X-Road service;

(iii) Data are submitted in XML format and the creator of the X-Road service will submit the description of the data as XSD. Upon capturing data, the chief processor or authorised processor of the database of Statistics Estonia will use the submitted description in XSD format;

(iv) The updates must include the time of the presentation moment; updates must be possible to adopt as of the last presentation moment.

(c) Requirements for identifying objects;

(i) All Estonian addresses in the databases must originate from the Address Data System administration system of the Estonian Land Board or be connected to the ADS descriptions of the Estonian Land Board (address object identifier ADS_OID and address identifier ADR_ID);

¹⁸ Kütt, A. (2014). Andmete saatmise arhitektuuri mustrid. Maa-amet. (2013). ADS käsiraamat.

(ii) All the persons in the databases have to be identified with a personal identification code (in the case of natural persons) or Commercial Register, Non-Profit Organisations and Foundations Register (hereinafter Commercial Register) code (in the case of legal persons);

(iii) Requirements for the classifications and comparison tables in the database foresee that they will have been coordinated with Statistics Estonia following the requirements of the Administration System for the State Information System (RIHA).

(d) Requirements for ensuring data quality

(i) Databases need to take into account the results of Statistics Estonia's periodic quality assessment.

(ii) The data captured from databases must comply with the characteristics agreed upon with Statistics Estonia, be complete and correct.

(iii) Databases must check the internal coherence of the data (e.g. identify and correct duplicate entries and errors occurring due to characteristics, etc.).

30. The activities necessary for meeting the requirements for all databases are: improving and monitoring data quality centrally; introducing the use of ADS administration system data instead of addresses originating from other sources; providing data with classifications and creating the X-Road service or introducing the use of Statistics Estonia's X-Road service.

31. The producer of statistics must ensure in the extent of census activity:

(a) Technical documentation of the composition of databases and implement the measures of information security;

(b) Signing agreements between the chief processor of the database and the authorised processors of databases regarding technical conditions;

(c) Cooperation between the information system and databases, regarding technical issues with state central level information systems and databases.

32. Meeting the requirements of data capture and carrying out activities will ensure the sustainable interoperability of state information systems for the register-based census.

VI. Conclusion

33. On the most general level, the prerequisite for register-based censuses is the presence of data in registers and possibility of linking the data.

34. In Estonia, there is now a situation where nearly all obligatory census characteristics have been covered with registers, except for the occupation and place of work. There is a network of information systems to enumerate persons and dwellings. There is a functioning system of personal identification codes and an address data system is being implemented, steps have been taken to use the data exchange layer X-Road in data capture.

35. Using the developed standards in the infrastructure of state information systems is a prerequisite in the transfer to a register-based census. It is important to ensure data capture from registers as well as the updating of data. If data collecting is automatic, we can talk about the interoperability of state information systems. Census data can be captured from registers within a few days, there are only a few exceptions which have to do with the retrospective registering of events in registers, e.g. births and deaths.

36. The prerequisites for automatic data capture are the technical environment, data format and quality. In the future, efforts need to be made to ensure interoperability for work with the data compositions of different registers. Capture which is based on the links between unique data would reduce burden regarding the receipt and storing of data. Quality criteria must be developed to assess data which was captured (e.g. number of technical errors, missing values etc.). The systematic check of quality in the data collection stage will facilitate reducing errors in register data. At the same time, methodology must be developed on how the producer of statistics can help improve data quality in registers and the database administrator can organize register data without a breach of procedure.

37. 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.

38. Taking into account Statistics Estonia's strategic views regarding data capture methods, data collection from registers must be reorganized – there is no other way to make the data collection process more effective and to reduce costs.
