Economic Commission for Europe

Conference of European Statisticians

Workshop on Population and Housing Censuses
for countries of Eastern Europe, Caucasus and Central Asia
Geneva, Switzerland
26-27 September 2016

Item 1 of the provisional agenda 26 September 2016

Technological innovations for the 2020 census round

Information technology and its impact on productivity

By Diana Beltadze and Ene-Margit Tiit

Abstract

The presentation covers problems regarding the implementation and impact of information technology as illustrated by the population and housing census because a census is the biggest and most demanding statistical action of the entire decade and the problems related to the organisation of censuses are familiar the statistics producers of all countries.

Introduction

Population and housing censuses are usually conducted every ten years and this is the most reliable data source when it comes to describing demographic development. Statistics Estonia is planning to conduct the next population and housing census for the first time in the history of Estonia by relying on register data. This change in methodology provides solutions for improving the business logic of statistics production, and tries to make it more effective and cost-efficient.

The creation of prerequisites for a register-based census started during the preparations for the previous census (2011) when the census methodology that is based on an information technological approach was developed. In order to conduct the previous census, the producer of statistics had to prepare a statistics production environment which is based on an information technological solution and a statistics production process which is automated to the maximum extent and which would enable carrying out the statistical actions necessary for the population and housing census twice as fast for the user as in the 2000 census round.

In the process of preparing for the previous census, an effort was made to develop information systems based on sustainability principles with limited budgetary funds. Such aspirations have always been relevant during the preparations for each census. However, the history of censuses in recent Estonian history shows that the development of IT technology is not predictable. It was more rapid in the period of 1995–2011. The IT platforms and systems which were introduced in 1998 were no longer effectively usable in the 2011 census. This is what caused the 1.6-fold increase of census expenditure for the 2011 census.

If we comparatively study the expenses made for censuses, they were greater in the censuses of 2000 and 2011 and were mostly connected to conducting data collection (salaries of enumerators) and preparing the statistics production system (developing information systems).

When it comes to Estonia it is important to note that despite the sufficient funds provided for conducting censuses, enabling the institution to use modern information technological solutions for the census, cutting census costs has always been topical on the government level.

It is believed that working environments which are based on new technological capacities enable reducing administrative burden and the cost of actions, but when it comes to censuses it is difficult to calculate the costs of various activities, e.g. preparation activities. At the same time, we have been able to estimate that the introduction of new technologies entail an increase in the cost of certain actions, e.g. data processing (2011 census data, 2014 pilot census data, data of the first trial census of the Register-Based Population and Housing Census (REGREL) held in 2016). However, other actions have become cheaper, e.g. GIS-related actions, data collection etc.

When introducing technologies, it is important that the statistics producer of the census does not ignore the fact that the economy is not only digital nor is society only virtual (Sassen 2002: 1–24). Secondly, we can use the benefits of the information society provided that we have the technical resources necessary. Thirdly, for both individuals and the state, the most important issue for censuses has always been and will be security and, more specifically, privacy (Beltadze 2016), because in the process of conducting a census, an enormous amount of private and delicate information is collected and stored. Technological applications need to enable data protection at the highest level.

In conclusion, new technologies reach both the organisers of the census and the users quickly and the volume and importance of information are on the increase. Managing an increasing volume of information, transmitting it quickly and without losses and at the same time ensuring its integrity, availability and quality have become the challenges related to the implementation of new technologies in censuses. Experience shows that it is only worth investing in information technology if it helps create value for the organisation.

Uses of information technology in recent censuses in Estonia

In the implementation of IT solutions, the main objectives have been as follows:

1. saving on manpower;
2. saving time;
3. more efficient process management;
4. better quality of results;
5. higher level of security in data protection.

The first three objectives are directly related to the aim of achieving higher productivity, while the other two facilitate a rise in productivity only indirectly, as they reduce additional work towards raising quality and improving security.

The possibility of saving financial resources with the help of IT applications is another matter entirely. Here, general logic applies – even the best technology is bound to quickly become outdated both physically and morally, which is why the applications are basically not sustainable long-term. It is clearly felt in developing the software necessary for census purposes: since the censuses are held at an interval of 10 years, there is a great risk that an application that worked very successfully in one census needs to be replaced in the next census due to the technological basis having become outdated.

In addition, it needs to be taken into account that a census is such a special statistical action that the IT environment that has been developed for census purposes is difficult to use for other statistical actions. Thus, it is evident that creating the IT environment necessary for a census is expensive as a one-time project, especially in terms of costly IT manpower.

One solution is to develop the IT environment as modules, with each module being usable in other statistical actions as well. Additionally, some modules can also be used in several censuses, regardless of the fact that the ideology and organisation of censuses might change and that information technology will have developed further. Example: Estonian Population and Housing Census 2011 – a paper-free statistical action, which was carried out fully as an IT project.

A modern census is a grandiose IT project. Since the lifespan of IT solutions is generally not long due to the solutions becoming morally outdated, it often means that a new software solution is created for each census. The previous census held in Estonia was carried out using a *mixed method.* IT-related innovations can be classified in various ways. The collection of data was the most important census activity, where the implementation of information technology proved productive.

1. Use of register data
   Some of the data (e.g. information on studying) were retrieved from registers and not asked from respondents; some data (e.g. legal marital status) were taken from registers and entered into the questionnaire (pre-filling) and, based on the rules, the respondents were asked to check the accuracy of the answers.
2. Technologies for responding
   Three data collection methods were used: self-completion of online questionnaires, computer-assisted interviews (the enumerator entered the answers into the questionnaire on a laptop) and paper-and-pencil interviews. The last method was a back-up solution in order to reduce safety risks (e.g. a cyber-attack).
3. Digital mapping of addresses during self-enumeration
4. Using the GPS in determining spatial coordinates during census interviews
5. Creation of an e-questionnaire in three languages
6. Using GIS technology in the organisation of the census
   The Estonian territory was divided into three regions, which were further divided into census districts, supervision areas and enumeration areas. An enumerator's job was organised based on the enumeration area. The census interviews were held (in a total of 2,000 enumeration areas) based on the enumeration lists compiled for the enumerators using the address data recorded in the Population Register. The enumerators had the task of enumerating all non-enumerated persons, dwellings and households, and clarifying the spatial coordinates of all dwellings (incl. those enumerated online) using the GPS. They also had to check the data recorded in the e-census and occasionally to fix some errors made in the online census.

A total of 8.41 million euros were spent on data collection. A total of 9 million euros were saved in data collection activities thanks to the introduction of IT solutions.

What was novel in this census was the data verification and organisation (incl. coding) done simultaneously with data collection, but also the automatic modification of the survey questionnaire according to the answers, all of which saved both the interviewers' and the self-enumerators' time. This was followed by a more difficult stage, data processing, which was conducted in VAIS, a new information system designed for processing survey data. Unfortunately, the IT solutions used at this stage were not effective, but today the information system is used in several other surveys as well as in the activities of the new census round.

There were several reasons for why the information system that was developed for data processing did not work effectively during the census. The main problems concerned ensuring the quality of data in the case of recording addresses during the census. It was the first time in the history of censuses in Estonia that respondents could write down their address themselves. The clean-up of address data proved to be a very labour-intensive task for data processing operators and also prevented the speedy publication of census data. The address standard, which was being implemented during the census and is fully effective now, contains unique codes and spatial coordinates of addresses.

The use of IT solutions also affected the saving in the general costs of the census, with a total of 5.43 million euros saved. This included 2.9 million euros saved on personnel and transportation costs thanks to the e-census; 0.72 million euros saved on the development costs of VAIS (the system for statistical activity data processing), which were lower than planned; 0.25 million euros saved on linking spatial information thanks to using GIS solutions. The saving on postage, printing, advertising, PR, training and licensing costs amounted to 1.17 million euros. (Report to the census commission of the Government of the Republic 2013"

The wider societal and economic impacts of the e-census are still unknown because the e-census should be regarded together with the effect produced by the development of public services in Estonia and with the implementation of e-services.

What were the effective new IT solutions used?

A Survey Fieldwork Information System (VVIS) was created, which enabled us to create enumeration lists based on enumeration areas, to compile questionnaires both for the online census and laptops, to supplement the questionnaires with additional materials and checks, to manage the roles of the census team, to monitor the census and exchange of information between the members of census teams of different levels, to compile current statistical overviews, organise initial data (identify, encrypt, etc.). As a result of the operations performed in VVIS, a first version of the source dataset is compiled based on census data.

The system was also introduced in sample surveys.

What necessitated the need to update the system of monitoring and managing personal surveys?

1. The need for up-to-date information during the 2011 Population and Housing Census.
2. The limited time available for performing census activities and the great number of interviewers (ca 2,000) working simultaneously.
3. The need to manage the fieldwork of various levels.
4. The use of several data collection channels: internet-based interviews (CAWI) and computer-assisted personal interviews (CAPI).

The following was achieved with the introduction of the information system:

1. Data processing was sped up (and thus the publication of statistics also became faster).
2. The coverage of the survey was improved and losses were reduced.
3. The quality of raw data was improved.
4. The effectiveness of the monitoring of the data collection process was increased.
5. Information exchange during fieldwork was sped up.
6. The cost of census interviewing was reduced.

7. Labour needs were reduced (there were fewer enumerators in the 2011 census than in the 2000 census).

8. The data collection for personal surveys was supplemented with new means of collection: CAPI (2011 census), CAWI, CATI.

9. The quality of official statistics was raised.

The changes made to the process of statistics production were as follows:

- A permanent online connection between Statistics Estonia and an enumerator. Data are transmitted automatically when an Internet connection is established;
- The information necessary for finding a respondent is linked to the enumeration list (previously to the questionnaire);
- Customer support – all contacts made with a respondent, incl. cases when the respondent has contacted Statistics Estonia, are recorded and the records are kept with the enumeration list;
- Less time and money are spent on looking for objects (addresses) thanks to the map application together with the GPS;
- The supervisor of interviewers has a better overview of the status of tasks and of keeping to the schedule.
- A fast exchange of notices and information (clarifications and changes concerning the enumeration list, contacts made with Statistics Estonia by the respondents).
- If necessary, an overview is available on the census as a whole or on the strata required for all surveys and regions. It is possible to get a full picture (graphs/maps) of any period in a fast and easy manner.

Since 1922, spatial information in the form of maps has been used in Estonia in preparing for and conducting censuses, and in disseminating census results. Nowadays, the majority of countries use geographic information systems in preparing for and conducting censuses and in analysing and disseminating the results. This was also the case in Estonia in 2011. Implementing GIS technology enabled Statistics Estonia:

1. to adequately plan the number of enumerators.

2. to effectively monitor the census fieldwork, as the enumerators knew their working areas.

3. at any moment, to give a fast and clear overview of the census to the directors, the census monitoring steering group and the media.

4. to provide the enumerators with census maps necessary for giving them a clear overview of the working area in order to help them orientate themselves in the area.

5. to provide more detailed census results (detailed, grid-based datasets) that would meet user needs better.

6. to present census results to Eurostat based on the "Localities" method

7. in the dissemination of census results, to omit one source of self-financing which was significant for users (digital census maps, digital grid maps of census results)

8. to provide field enumerators with data on the location of persons having participated in the e-census; generally, no respondent was contacted unnecessarily.

GIS technology was used already in the 2000 census round, but the use of this technology had to be changed for the 2011 round. The main lessons learned in 2000 from the tasks related to the GIS technology and maps can be summarised as a negative experience because:

- the census maps (spatial information) used in 2000 were not manageable (ca 100,000 different files) because there was no database-centred management;
- the attribute information of census maps was missing sufficient metadata which would have ensured the sustainable use of data;
- the link between the spatial information of the census map and registers was not retained, which is why basically the same investments which were already made during the preparations for the 2000 census round had to be made for the new census in 2011;
- spatial information has not been updated (there is no exchange of information with spatial databases).

At the same time, in 2011, it was possible to identify spatial coordinates with much greater accuracy than in 2000.

Monitoring

The progress of the census was monitored via an electronic monitoring system. During the online census, each day, Statistics Estonia's website featured information on how many people have had themselves enumerated online and on the share these people amounted to among all persons to be enumerated. The same information was also published for each county. Thus, a sense of competition between the counties was felt, encouraging the inhabitants to have themselves enumerated online.

During the interview phase, the supervisors of enumerators could monitor the progress of each enumerator and react to cases where some enumerators significantly lagged behind their work plan or performed their tasks unnaturally fast, raising doubts about the quality of their work. The enumerators' expenditure on transportation (both on private cars and public transportation) were also calculated electronically.

The monitoring system as a whole substantially increased the effectiveness of the census, but it also played a great role in presenting the census to the public and in motivating the population to participate in the census.

Data protection and ensuring data security

Great attention was paid to data security. The census questionnaire could be accessed on the Internet only by using an ID-card, online banking passwords or a mobile banking password, which prevented third-party access to the data.

The interviewer who filled out a person's questionnaire lost the opportunity of seeing it as soon as the questionnaire was completed. Since the data stored in laptops were encrypted, there was no risk of an outsider taking a look at the data even if the laptop had been lost. The channels used to transmit the data to the database were also secure and encrypted.

The data of questionnaires which had no errors in them were transmitted to the database even without the operator having seen them. On the one hand, this ensured data security; while on the other it also made the work more efficient.

However, if it was necessary to clarify certain data in the questionnaires (e.g. to specify address data or to add a missing personal identification code), it was done by the data processing team, where the roles were divided as follows: the operator responsible for personal identification codes would deal with that code, the operator responsible for address data specified addresses, etc. The operators did not generally have access to data that did not concern their role. Thus, the operators were not able to see the content of the questionnaire of a specific person either. This was ensured by the system of automatic checks, which was implemented in the data processing system and prevented all kinds of mix-ups and thus increased the efficiency of work.

Future plans

Estonia is still in the learning process, as we want the solution developed for the new register-based census round to be not only cost-effective but also of high quality, and also aim for a varied implementation of the developed solutions in the production of statistics in the future.

The costs of the register-based census are predicted to be 10 times smaller than those of a regular census (Laihonen 1996). However, census costs may not be worked out immediately after the first register-based census.

It is not easy to calculate census costs, as it depends on calculation methodology and primarily on the starting position of the statistics producer upon conducting the census in one way or another. The difference may result from the specific character of the system of registers in the country, population density, people's attitudes and loyalty, income level, etc.

One of the aims of Statistics Estonia's strategy for 2015–2020 is to increase the timeliness of statistics and the speed of publishing data (Strategy of Statistics Estonia 2015–2020). If data are collected in the shortest time possible, they can be published faster, enabling the state to react faster and more effectively to economic and demographic changes.

In order to conduct a register-based census, the producer of statistics has to develop a production environment that is based on an IT solution and a production process that is automated to the maximum extent, enabling the statistical actions necessary for the Population and Housing Census to be carried out based on registers. The preparations for the register-based census are focused on datasets and data flow.

Summary

In the previous census, Statistics Estonia aimed for maximum coverage and cost-effectiveness. It has to be noted, however, that it was difficult to give an overview and assess the reasonableness of the

costs of data collection in the previous census because there was no methodology for calculating the monetary value of public services.
Our cost-benefit analyses were cost-based. As the most accurate data on the costs of data collection were available for the 2000 census, these data were taken as the main basis for comparison in assessing the costs of the 2011 census.
At present, it can be said that technological, legal, and organisational prerequisites guaranteed success in the 2011 census.
Evaluating effectiveness is a more difficult problem because there is no methodology for assessing actual impacts in the production of statistics. We are still looking for an answer to the question whether the impact of IT solutions that was desired has actually been achieved in the process of statistics production.

What can be considered an indirect benefit is that the functionality of the developed VVIS and VAIS systems are, either fully or to a great extent, universal for collecting all data gathered through Statistics Estonia's network of fieldwork.

Data processing information systems and GIS solutions are already in use in the preparations for the next census (in the pilot census of 2014, the trial censuses of 2016 and 2018).

Sources

UNECE 2007. *United Nations Economic Commission for Europe (UNECE) (2007). Register-based statistics in the Nordic countries. Review of best practices with focus on population and social statistics.[www]http://www.unece.org/stats/publications/Register_based_statistics_*

in_Nordic_countries.pdf

UNECE 2008.*United Nations Economic Commission for Europe (UNECE) (2008). Measuring Population and Housing – Practices of UNECE countries in the 2000 round of censuses. New York and Geneva: United Nations.*

UNSD 2009. Main Results of the UNECE-UNSD Survey on the 2010 Round of Population and Housing Censuses. Conference of European statisticians. Twelfth Meeting. Economic Comission

For Europe: Geneva

Hagedoorn, J. (1993). *Understanding the rationale of strategic technology partnering: Interorganizational modes of cooperation and sectoral differences*. Strategic Management Journal, 14, 5.

Laihonen, A. (1996). *Register-basedpopulation andhousin sensus systemin Finland.* WorkshopCensus2001: Methos and Content. Report. Lisbon. Portugal.

Sassen,S. (2002). Towards sociology of information tehnology. Current Sociology. CSI. Vol. 50, No3 1-24

Schulte-Nordholt, E., Hartgers M., Gircour R. *(2004). The Dutch virtual Census 2001. Analysis and Methodology. The Hague: Statistics Netherlands. [www] http://www.cbs.nl/nr/rdonlyres/d1716a60-0d13-4281-bed6-3607514888ad/ 0/b572001.pdf*