Distr. GENERAL

30 October 2012

WP. 31
ENGLISH ONLY

UNITED NATIONS
ECONOMIC COMMISSION FOR EUROPE

CONFERENCE OF EUROPEAN STATISTICIANS

Seminar on New Frontiers for Statistical Data Collection
(Geneva, Switzerland, 31 October-2 November 2012)

Topic (v): Economies of scale from using common tools and methods

STATISTICS ESTONIA ON ITS WAY TO IMPROVING EFFICIENCY

Contributed Paper

Prepared by Tuulikki Sillajõe, Statistics Estonia

I. Introduction

1. Changes in the society are fast. The demand for official statistics is continuously increasing. Users need more and more statistics, and they need them quicker. Statistics are required to measure an increasing number of phenomena, at an increasingly more detailed level. At the same time, the overabundance of information poses a challenge for the visibility and integrity of official statistics, because new suppliers of information appear on the market. At the end of the day, users might be confused, data suppliers might suffer under the great response burden and statisticians might be irreparably overloaded.

2. To cope with the changing environment, Statistics Estonia developed the strategy "From data collector to information service provider" for the period 2008–2011. It has been a continuous movement towards a higher reputation of official statistics, higher efficiency of processes and better partnership with data suppliers.

3. The paper gives an overview of the statistical production system and the relevant tools at Statistics Estonia. More specifically, it presents the development of the data collection phase of the statistical business process. It also describes how centralisation of functions and usage of generic tools have allowed the organisation to cope with a decreased budget.

II. Organization of data collection function

A. Development of data collection function

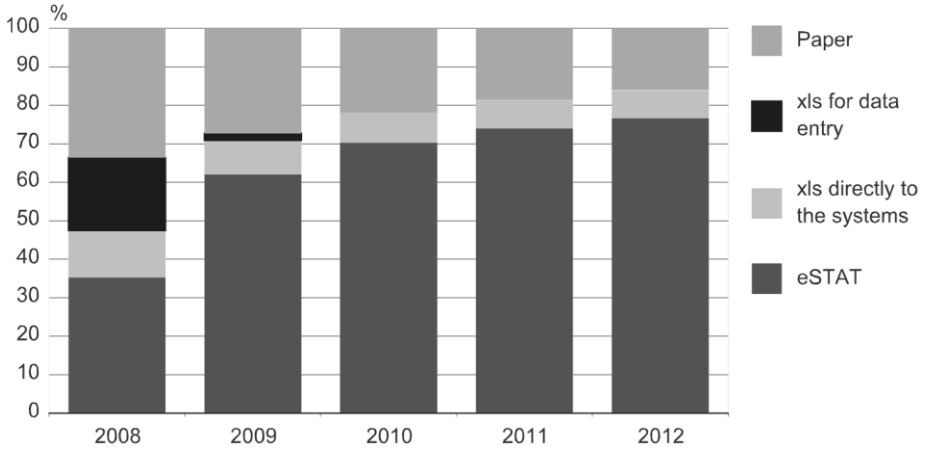
4. At Statistics Estonia, the data collection tasks of the value-chain of official statistics production were organised into a central functional department in 2004. The central department provides all specialised departments with data collected from both individuals and economic entities.

5. For a long time, the customer support for respondents at Statistics Estonia was questionnaire-based, i.e. in case of questions about a particular statistical questionnaire a respondent could contact a specialist in the relevant statistical domain. This way, a highly professional service concerning one particular questionnaire was guaranteed. In 2005, the user support centre was integrated with the Data Collection Department and the Contact Centre was established.

Respondents can now call the common telephone number and send e-mails to the common e-mail address set up for questions and enquiries about any questionnaires to be submitted by economic entities irrespective of the domain (e.g. construction, trade, agriculture, wages and salaries, etc.). In 2008 the Contact Centre started handling contacts with the respondents of social surveys as well. That change was caused by the fact that individuals covered by social surveys started to ask for information which could not be given to them by interviewers who had been their main contact with Statistics Estonia for many years. The experience has shown that the staff of the central Contact Centre have been able to answer more than 99% of the respondents' questions. This way, statisticians have more time to concentrate on other parts of the statistical business process (processing, analysis, etc.). Also, respondents are satisfied with the level of the service they get – the 2011 respondents' satisfaction survey showed that general satisfaction with the services of the Contact Centre was rated 7.8 on a 10-point scale.

6. In 2006 Statistics Estonia launched eSTAT, which is a web-based channel for the completion and submission of statistical questionnaires of economic entities. The Data Collection Department is the owner of the eSTAT system and responsible for its development according to the needs of external and internal users. It is possible to present all (166 in 2012) statistical questionnaires of economic entities via eSTAT. As Figure 1 indicates, the share of data collection via eSTAT has reached 77% of the total number of statistical questionnaires received from economic entities by Statistics Estonia.

Figure 1. Questionnaires received from economic entities by channel, 2008–2012



7. eSTAT enables respondents to view the list of statistical questionnaires that a particular economic entity has to present to Statistics Estonia in the current year; to view the deadlines for the submission of these statistical questionnaires; to request e-mail reminders notifying users about upcoming deadlines; to complete statistical questionnaires, by filling in fields on screen or downloading and uploading CSV files (the latter is particularly meant for large enterprises with a big number of records, such as the Intrastat statistical questionnaire and some statistical questionnaires on salaries); to run controls, i.e. check whether the user has completed the statistical questionnaire as required; to correct statistical questionnaires immediately after these have been completed; to submit statistical questionnaires to Statistics Estonia; to view all earlier statistical questionnaires submitted to Statistics Estonia via eSTAT by the given respondent; to print out a paper copy of a completed statistical questionnaire; to administer users, i.e. to create, change and cancel rights and access; to accept or correct one's contact information, and so on.

8. In order to join eSTAT, an economic entity must file an application. Although the economic entity is the respondent, only individuals are allowed to log onto eSTAT. One person (a contact person) can submit the questionnaires of several respondents (e.g. an accountant working at an accounting firm) and, conversely, the questionnaires of one respondent may be submitted by several persons (e.g. the accountant, the human resources manager and the executive manager complete different questionnaires). An official representative of the economic entity shall determine who specifically can submit and correct statistical questionnaire on behalf of that

economic entity. In order to enable the contact persons of an economic entity to submit data via eSTAT, Statistics Estonia will create the so-called main user for the economic entity. The main user in turn has the right to create users. The use of eSTAT only by authorized persons ensures that each user sees only the data needed for his/her work. eSTAT enables secure data transmission. The security is guaranteed by the implementation of contemporary IT security measures. The connection between a user's computer and eSTAT is encrypted.

9. At Statistics Estonia the data collection from individuals has been completely paperless since 2005 – laptop computers (CAPI) have been used instead. The process is based on Blaise software. Web-based data collection (CAWI) was tested in a survey of doctorate holders in 2007, but not implemented in regular production. The first mixed-mode data collection from individuals also took place in 2007, when data for the General Household Module Survey were collected with CAPI and with outsourced telephone interviews (CATI).

10. Statistics Estonia decided to conduct the Population and Housing Census 2011 using a mixed mode: web-based data collection was followed by an interview period when enumerators used laptops to enter data not already submitted through the e-Census web site. Thus, new generic software for data collection had to be developed. The method and the software were tested in the Pilot Census in 2009 and used for the first time for the Agricultural Census in 2010. During the Agricultural Census, economic entities could present their data via the familiar eSTAT environment and individuals with an agricultural holding could use the new software, while in both cases respondents were visited by interviewers with laptops if some data had not been submitted online.

11. There is a single authorisation point on the Internet called eSTAT. After logging in, each respondent has access to all the questionnaires assigned to him or her as an individual or as the contact person of one or more economic entities. The respondent is not informed whether he or she is using the old eSTAT or the newly developed software. The difference lies in the suitability of a particular questionnaire for one of these software programs. For example, the newer software is better suited for social surveys on individuals.

12. From Statistics Estonia's point of view, the new generic software has several outstanding features. It supports the data collection process of various surveys and censuses, dividing it into three sub-processes:

    (a) Preparation work for data collection;

    (b) Data collection on the web (CAWI) and fieldwork (CAPI);

    (c) Support and management of the whole process.

    Preparation of data collection allows to define questionnaires; to import samples from the statistical register; to assign the people involved in data collection and the hierarchy of management; to divide work tasks within the hierarchy; to grant access to systems and assign relevant roles for statisticians, survey managers, the call centre, interviewers, etc.; to pre-fill questionnaires with data from registers; to communicate with different people involved in preparation.

13. During CAPI it is possible for interviewers to navigate between different questionnaires, to use map info for easier subject location and work planning; to plan their work and to schedule contacts with interview subjects. The local database in the interviewers' laptops is encrypted; all data for offline work are synchronised over an encrypted channel.

14. Fieldwork managers have an overview of fieldwork progress and can easily and quickly distribute tasks between interviewers and approve filled-in questionnaires. The staff of the Contact Centre can schedule contacts and assist with questionnaire completion. There is also functionality for data operators who can classify answers in questionnaires and remove duplicate

questionnaires. Communication between managers of different levels, statisticians, and other parties is supported. Various activities in the system are centrally logged.

15. The new software was successfully used for the Population and Housing Census 2011. The census moment was 31 December 2011. The Census was carried out from 31 December 2011 until 31 March 2012. From 31 December 2011 to 31 January 2012, an e-Census took place where the residents of Estonia could fill in questionnaires online. From 16 February to 31 March 2012 enumerators visited those who had not participated in the e-Census. The participation rate in the e-Census was 66%. It was the first totally paperless census in Estonia.

16. In 2008 pre-filling of the questionnaires of economic entities with data from previous periods was introduced. A new functionality in data collection was introduced in 2011 – the annual statistical questionnaires of economic entities to be completed in 2012 (for collection of data on the reference year 2011) are pre-filled using administrative data. Respondents only have to fill in the gaps, i.e. information not available in the annual report. Statistics Estonia downloads annual reports from the Commercial Register every hour. The administrative data are converted into the statistical format by a new generic data processing system called VAIS. The information from annual reports is preloaded to eSTAT every hour as well. Pre-filling is currently used for annual statistical questionnaires for structural business statistics (EKOMAR), agriculture, forestry and fishing, financial intermediation and activities auxiliary to financial services and insurance activities and non-profit institutions.

17. As of the beginning of October 2012, 52% of the questionnaires for structural business statistics (EKOMAR) had been pre-filled from annual reports. The share was lower than expected, because accountants started to fill in EKOMAR questionnaires before their annual report had been signed by the board of the particular economic entity (Statistics Estonia can download the data only after the report has been approved). 80% of the fields were pre-filled, and respondents had to fill in the remaining 20% (only fields filled in with a number other than zero are taken into account). In addition, respondents corrected the data downloaded from the Commercial Register in 3% of the cases. The share of pre-filled fields did not depend on the size of the economic entity, but bigger entities did more corrections (15% of the pre-filled values were corrected). Corrections were relatively small, in total 1%.

18. Pre-filling from administrative registers and the Population and Housing Census 2000 was also used for the Population and Housing Census 2011. Some data were not even asked from respondents at all, but collected from registers.

B. The smallest changes may initiate bigger developments

19. During the previous strategy period, Statistics Estonia looked for efficiency gains from wider use of electronic data collection. The progress has been due to constant analysis of the office's actions and trying new ways of communication with respondents. In the implementation of central data collection, Statistics Estonia has changed its habits many times in order to encourage respondents to change their habits and switch over to electronic channels.

20. From the very beginning of central data collection, reminders have been sent to economic entities before the deadline of each statistical questionnaire, instead of after the deadline as in the past. The first two reminders are sent electronically and only the third reminder is sent out on paper.

21. Also, in December each year, economic entities are simultaneously informed about all the questionnaires they have to complete in the following year. The information is sent directly to the economic entities, but also published on the website of Statistics Estonia and in the official publication *Ametlikud Teadaanded* (the web-based publication for official notices).

22. The preparation of questionnaires for economic entities was centralised from statistical departments to the IT department and then over a few years to the Methodology Department. Due to the fact that all 170 questionnaires are designed by only a few people, a great amount of

standardisation and simplification has been done. These efforts have been appreciated by accountants who have to fill in the questionnaires. Usually the same accountant has to fill in several statistical questionnaires and any differences between definitions or explanations cause confusion, while developing a complete understanding of the differences takes time and prolongs the process of presenting statistical questionnaires.

23. So far, the biggest rise in the share of electronic submission of statistical questionnaires of economic entities was observed after the termination of the practice of sending empty printed questionnaires to economic entities in 2008. It meant that instead of a thick envelope with empty statistical questionnaires for the whole year, economic entities received thinner envelopes with the list of statistical questionnaires they had to submit for the production of official statistics. That change was initiated by respondents' surveys which showed that respondents were reluctant to join eSTAT because they were used to receiving printed questionnaires from Statistics Estonia. That action saved more than 100,000 euros due to reduced paper and printing costs, but hopefully also some 100 trees were saved from being cut down each year. The Estonian society welcomed that action and Statistics Estonia was recognised as an environmentally friendly institution.

24. Individuals are still sent a printed notice informing them about their inclusion in the sample of official statistical surveys.

25. In 2008 measures were taken to allow the generation of statistical questionnaires in the PDF-format from eSTAT. These files can be used for all kinds of administrative purposes, such as submission for approval or publication online, but also as a print-ready original if needed. This innovation meant that Statistics Estonia could stop the process of manually creating statistical questionnaires in the XLS-format, which was a parallel process to description of statistical questionnaires in eSTAT. Consequently, at least one man-year could be saved and utilised for other tasks, not to mention the manpower of staff who no longer have to compare whether the statistical questionnaires created during the two separate processes are identical. An additional win was that economic entities could no longer download the Excel files from the web site of Statistics Estonia, modify them and send them in by e-mail with different formatting. Guaranteed input in a certain format has saved a great deal of time that used to be required for formatting within Statistics Estonia.

C. Enablers and catalysts of changes

26. The main enabler of changes in data collection has been the development and wider use of information technology. In Estonia 75.5% of households have a computer and almost as many (75%) also have an Internet connection at home (2012 data). 97% of enterprises in Estonia use computers in their work and almost as many (96%) have an Internet connection (2012 data).

27. We cannot underestimate the importance of the creation of the Generic Statistical Business Process Model. Although Statistics Estonia started to develop its own business process model at about the same time inspired by the example of Statistics New Zealand, it has now implemented the generic version. Understanding that the production process is the same for all statistical domains and independent of data source (surveys, censuses, administrative records, and other non-statistical or mixed sources) has given plenty of room for more efficient operation. One has only to keep in mind that the model is not linear, i.e. the elements of the model may occur in a different order in different circumstances.

28. A constant driver of change in the modern world is the desire to reduce the administrative burden of respondents as well as attempts to expand the reuse of administrative data and commercial data. The recent economic crisis and intense competition have forced enterprises to cut costs, incl. those related to the submission of statistical questionnaires.

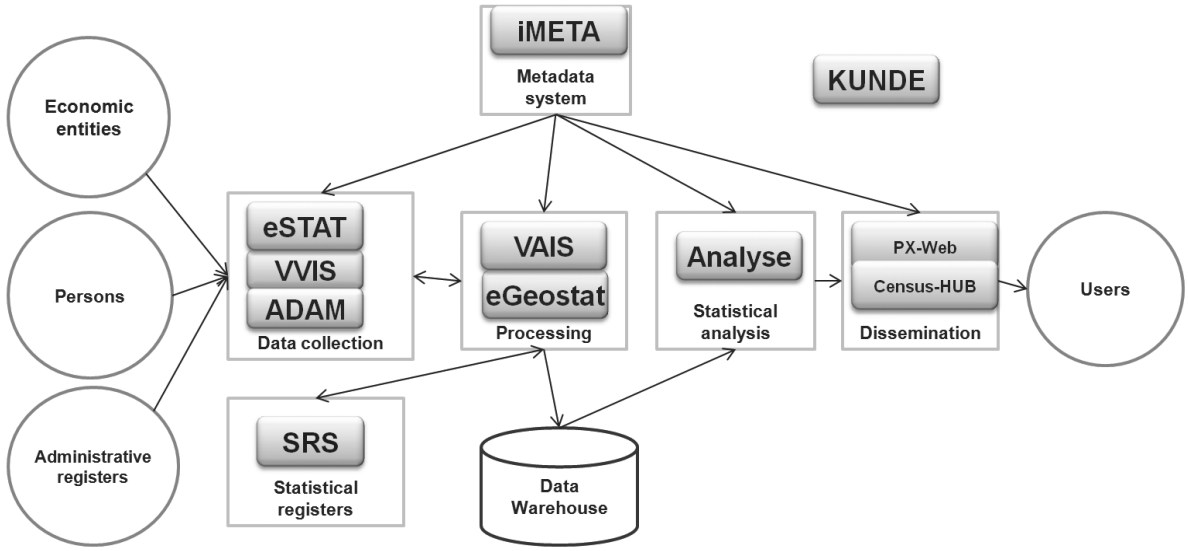
29. Statistics Estonia has had to deal with considerable budget cuts in circumstances where the society is asking for more and more official statistics with a greater level of detail. It has put the organisation in a situation where it has to constantly and repeatedly justify the cost of each and

every position and calculate the ROI of each investment into technology. In 2012 the annual expenditure from state revenues (excl. expenditures of the Population and Housing Census) of Statistics Estonia is roughly on the level of 2006, but the office has not noticeably reduced the output.

III. The production system as a whole

30. The data collection phase cannot be approached separately from the eight other phases of the statistical business process. At Statistics Estonia, generic office-wide software has been developed or is under development for other phases as well.

Figure 2. Architecture of the information system



31. The architecture of the information system is presented in Figure 2, where the acronyms stand for the following information systems:

(a) iMETA – integrated metadata system (2011);

(b) SRS – system of statistical registers (2012);

(c) Kunde – customer relationship management system (2006);

(d) eSTAT – data collection system for economic entities (2006);

(e) VVIS – data collection system for private individuals (2011);

(f) ADAM – data collection system for administrative data (2011);

(g) VAIS – template-based data processing system (2012);

(h) Analysis system (2013);

(i) PX-Web – output database (2001), will be replaced by .Stat (2012 and onwards);

(j) Census Hub – dissemination tool for European statistics (2013–2014).

32. In the context of data collection, the data collection system for administrative data should be described in more detail. At Statistics Estonia, the Data Collection Department is not responsible for that type of data collection. This responsibility is divided between two other central departments of the office: the Methodology Department (established in 2004) and the Data Processing Systems Department (also established in 2004, but reorganised on the basis of the IT Department in 2011). By the way, in case of Statistics Estonia, the bulk of traditional IT-functions are supplied by another institution – the IT Centre of the Ministry of Finance (established in 2012). It means that the Ministry of Finance has centralised IT-functions in its area of administration.

33. The Methodology Department consolidates the needs of statistical domains within Statistics Estonia for data available in registers; conducts negotiations with the holders of administrative and other registers, and organises the conclusion of agreements with these holders. The Methodology Department is also in charge of the description of data in a central metadata system. Technically, the data from registers enter Statistics Estonia via a single entry point which is under the responsibility of the Data Processing Systems Department. This department runs pre-agreed data processing and makes the data available for in-house applications. This ensures that there is no duplicate collection of data and the data are ready for statistical analysis.

34. In 2012 Statistics Estonia used about 100 different administrative registers (Population Register, Estonian Education Information System, Register of Construction Works, Health Insurance Database, Commercial Register, Register of Taxable Persons, State Register of State and Local Government Institutions, etc.) for the production of official statistics. In case of those registers which are more widely used within the office or from which data is taken more often, automatic extraction of detailed personalized data takes place using X-road (national data exchange layer). FTP services or similar are used in case of the rest of the external registers. A special data acquisition application (ADAM) has been created for transition to register-based statistics production.

35. Statistical registers have been an important prerequisite for standardisation and improved efficiency of processes at Statistics Estonia. In 2011 the office started to develop the System of Statistical Registers (SRS). SRS will integrate existing statistical registers (economic entities and agricultural holdings) and new statistical registers (persons, and buildings and dwellings) into a common system.

36. The current strategic approach of Statistics Estonia has been to develop only generic office-wide software, but whenever possible to use commercial statistical software (SAS, SPSS, etc.) or software developed by other members of the international statistical community. So, for the output database, PC-Axis has been used for many years and .Stat will be used from 2012 onwards; µ-Argus and Ţ-Argus have been used for disclosure control, CLAN for variance estimation, IVEware for imputation, and Demetra for time series analyses. Statistics Estonia believes that this is an example of how scarce resources have been used efficiently. More statistical domains have benefitted, even though at the same time great efforts have been made for the standardisation of methods and working routines.

IV. Efficiency gains in the data collection of the Population and Housing Census 2011

37. As mentioned before, mixed-mode data collection (CAWI + CAPI + registers) was used for the Population and Housing Census (PHC) 2011. For that purpose, a considerable part of the information systems at Statistics Estonia were modified. A new data collection system was developed which can be used as a generic system for other statistical domains, but ensures cost-efficiency even if used only for data collection for PHC 2011.

38. Table 1 compares the costs of data collection of PHC 2000 and PHC 2011. The data in the first row of the table are based on real costs of PHC 2000, taking into account the consumer price

index of the past ten years (meaning that if in 2011 the same data collection method had been used as in 2000, the cost would have been 13.2 euros per enumerated person). If in 2011 new generic software and a different method of data collection had been used (compared to 2000) and the online participation rate had been 25%, the cost of data collection would have been 8.14 euros per enumerated person. The real cost per enumerated person was 6.50 euros, but it might be even less come 12 December 2012 when the final number of enumerated persons will be published.

Table 1. Costs of data collection of the Population and Housing Census 2011

	e-Census participation rate	Number of enumerated persons (preliminary)	Number of enumerators	Cost of data collection, euros	Cost of data collection per enumerated person, euros
PAPI (prediction)	0%	1,340,000	4,592	17,710,000	13.2
CAWI+CAPI (prediction)	25%	1,340,000	2,970	10,910,000	8.14
CAWI+CAPI	66%	1,294,236	2,131	8,410,000	6.50

39. The cost per enumerated person might seem high compared to countries with a bigger population where the economies of scale are higher. But it was at least twice as small as it would have been if PAPI had been used; and it was at least 25% lower than predicted for use of the new method of data collection. So, in the Estonian context, efficiency has been reached.

40. The main reasons for lower costs of data collection during PHC 2011, compared to PHC 2000, were as follows:

(a) There were no special data entry costs;

(b) Automatic checks diminished mistakes during interviewing in both cases (CAWI and CAPI);

(c) Less time was required for interviewing than planned, and therefore fewer enumerators were hired;

(d) Printing, communications and archiving costs were considerably smaller;

(e) Management of data collection was much cheaper because much fewer employees were needed, thanks to the management module of the new software.

41. Talking about PHC 2011 in Estonia, one of the considerable achievements was the outstanding participation rate in the e-Census. It could not been achieved without the widespread use of Internet in Estonia, but was significantly influenced by the well-planned, strictly targeted and executed public campaign. The public campaign was deeply integrated and coordinated with the data collection activities. A special awareness survey was launched about a year before actual data collection started. Based on that survey, the PR activities were corrected and targeted. So, in February 2012, after the e-Census, awareness of the census among the population aged 15–74 was 99%, compared to 98% in January 2012 and 57% in August 2011. At the same time, 95% considered the census necessary (72% answered ’necessary’ and 23% ‘quite necessary’); 90% considered their knowledge about PHC ‘good’ or ’very good’, compared to 45% in August 2011. Already in November 2011, 60% of the respondents planned to participate in the e-Census.

42. One of the communication activities during PHC 2011 was reporting about the progress of the Census on the web site of Statistics Estonia. The number of enumerated persons in total and by county was updated hourly. This information was highly appreciated and widely discussed by the society as a whole, especially in the media. The comparison of counties started to serve as an additional driver, motivating people to enumerate themselves online.

43. Another effective feature on the web site of Statistics Estonia was a tachometer with a green, yellow and red zone for indicating the workload and traffic on the census page. The green zone indicated the best time for self-enumeration, the yellow zone marked a good time for self-enumeration, and the red zone meant possible disturbance or slowness of the system.

44. In case of PHC 2011, the social media also played an important role. During the data collection period, Statistics Estonia's PR staff worked on Facebook virtually for 24 hours per day. Information was quickly disseminated. People's questions were answered when the lines of the Contact Centre were busy or e-mails were not answered as fast as expected.

V. Conclusions and further developments

45. Statistics Estonia has observed economies of scale based on centralisation of functions. The currently centralised processes are dissemination (1993), data collection (2004), methodology (2004) and IT (2004). Centralisation of data processing is under serious consideration.

46. Inside the function of data collection, increases in efficiency have been achieved through the integration of data collection from economic entities and individuals. Efficiency has been found with the help of close cooperation between the data collection and dissemination functions. In the future, the stronger leadership of methodologists, supported by systematic implementation of common software for all statistical domains, could lead to a level of efficiency which has not been reached ever before. These gains are imbedded not only in standardisation, but also in bringing in new sources of data and using more complicated methods of data collection (e.g. data mining from existing sources, etc.).

47. Statistics Estonia was able to successfully conduct PHC 2011 at a two times lower cost per enumerated person than the use of the method of the previous census would have allowed. It was achieved due to the change of method and implementation of newly developed software.

48. Based on the software developed for PHC 2011, Statistics Estonia plans to start introducing CATI for data collection from both types of respondents (economic entities and individuals) step by step from 2012, but also to introduce CAWI for other surveys on individuals starting 2013. Pilot surveys using CAWI are scheduled for the 4th quarter of 2012.

49. There are also plans to increase the training offered to data suppliers (economic entities, individuals, registers, etc.), underlining their personal and public gains from official statistics.

50. There is still room for further simplification of statistical questionnaires (harmonisation of concepts, deadlines, practices, etc.). A specialised laboratory could find a way through the existing problems.

51. And last but not least, Statistics Estonia has plans to develop infrastructure for selling data collection services. This infrastructure should include questionnaire design, sample design, interviewing, management of interviewers, and processing and primary analysis of collected data.

VI. References

Annual report 2011. (2012). Statistics Estonia. [www] http://www.stat.ee/annual-report (26.10.2012)

Sillajõe, T. (2012). Implementation of the web-based data collection channel eSTAT for economic entities. [Unpublished paper presented at the NTTS conference "New Techniques and Technologies for Statistics" in Brussels in 2012]

Statistical Database of Statistics Estonia. [online database] http://pub.stat.ee/px-web.2001/dialog/statfile1.asp (26.10.2012)

The Generic Statistical Business Process Model. Version 4.0, approved by the METIS Steering Group in April 2009. [www] http://www1.unece.org/stat/platform/display/metis/The+Generic+Statistical+Business+Process+Model (26.10.2012)