

**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

**DATA COLLECTION IN THE STATISTICAL OFFICE OF THE
REPUBLIC OF SLOVENIA (SURS)**

Contributed Paper

Prepared by Dr. Mojca Noc Razinger in consultation with Tatjana Novak, Bojan Vovk, Irena Krizman,
Statistical Office of the Republic of Slovenia, Republic of Slovenia

*Key words: streamlining of data collection, lean production, process improvement, quality guidelines,
single entry point, project work.*

I. Introduction

1. SURS has a long tradition in organising the data collection in a single division. The process flow has been organised in cooperation with statisticians, IT experts and other support services. That enabled lean production of statistics and smaller demands on resources. “Stove pipe” processes exist only for some surveys. The reason is their specifics, mostly developed in the process of EU integration. The plan to adjust them to the general process was prepared. The other big challenge in streamlining the data collection process is coming from the intensive use of administrative data in statistical production. Successful implementation of a completely register-based population census in 2011 has established several good solutions, but we are going to continue with the process improvement. The objective is to get administrative data through a single point and lean the organisation of data editing and use of data by different divisions.
2. One of the most important parts of such organisation is good documentation. Detailed description of the process flow, clear organisational responsibility, good communication, standardisation and especially application of standards are also very important factors of a lean data collection process as well. Recently, SURS has produced a publication describing the process of the execution of statistical surveys and quality guidelines for good application. The GSBPM model was taken as reference.
3. The paper will present the main features of the current organisation of data collection, lessons learned from new developments (implemented and projects in progress) and challenges for the future.

II. Current organisation of data collection

A. Basics of the overall data collection process

4. The Data Collection Division is organised according to the type of reporting units and mode of data collection. Considering this we developed three different internal process flows that are intertwined in some parts: data collection from enterprises, data collection from administrative sources and data collection from persons, households and farms. Data collection from persons, households and farms is not formally in the Data Collection Division, but it is still part of data collection.
5. The Data Collection Division is responsible for preparing data collection, implementing data collection and communicating with reporting units, taking over administrative sources, data capture and manual micro-level editing. For surveys on persons and households, we also manage the interviewers (recruitment, control and instructions).
6. The central metadata system supports work planning with process flow description, time planning and responsibility for all activities in surveys. It enables cross reconciliation of used resources and encompasses detailed guidelines for work in the Data Collection Division. Work planning for each survey in the system is done by subject matter statisticians at the end of each year for the following year in consensus with performers: the IT Division and the Data Collection Division.
7. The IT Division prepares software solutions according to the work plan. The solutions are tested by the users in the Data Collection Division and used in corresponding activities. In the case of problems the solution can be quickly implemented.

B. Data collection from businesses entities

8. The Data Collection Division is responsible for preparing address lists from the Business Register and created samples. We introduced a new system into the production process in mid-2009. The system enabled us to prepare address lists in a common place for all surveys. We can track in which surveys the business entity is included, the mode of reporting and in addition it holds the information for tracking eligibility codes of entities and enables further development in this part of the data collection process.
9. Most of the surveys in the field of wage statistics, balances, Intrastat and Extrastat are managed by our partners in the statistical system. The Customs Administration of the Republic of Slovenia manages Intrastat and Extrastat, while the Agency of the Republic of Slovenia for Public Legal Records and Related Services manages surveys in the field of wage statistics and balances. In both cases advanced solutions for electronic data collection are utilised. Other more demanding data collection from business entities is still mainly on paper as we do not have our own infrastructure. But we have an ongoing project for implementation of the e-reporting system that will be in production in the next year.
10. In the data collection process from business entities we have five check points that the Data Collection Division needs to achieve: address list preparation, sending questionnaires, reporting deadline, end of data collection, data entry and manual logical controls at micro-level.

C. Data collection from administrative sources

11. For optimization of the data collection from administrative sources, we prepared internal regulations with defined roles of involved employees. We defined a content trustee who takes care of contents of the administrative source and a technical trustee who takes care of technical things and data transfer procedure. We have over 50 administrative sources. Technically speaking, they are not harmonized entirely.
12. The data collection process is determined in the agreement signed by SURS and the administrative institution. We have 48 signed agreements with 42 institutions. Agreements are published on the intranet site for all employees. The agreement encompasses the common part

and the annex, which is mostly technical and includes names of persons involved, time plan, data type and mode of data transfer.

13. According to the time plan or notification from the administrative institution, the employee responsible for data transmission accesses the data and stores them into the controlled environment in SURS.
14. Recently, we have prepared the metadata records of data collection from administrative sources and enabled access to internal users. In the records they can find the information on who is responsible, which data are in the source, time plan of the collection, etc.

D. Data collection from persons, households and farms

15. Data collection from persons, households and farms is the most demanding data collection process regarding financial resources.
16. For most of the surveys on persons, households and farms data collection is prepared and organised by SURS's unit for data collection from persons and households. Only in the case of extensive data collection, e.g. Farm Census, data collection is done with outsourcing and managed by subject matter statisticians.
17. Main tasks of the unit for data collection from persons and households are assuring the interviewers network (recruitment, instructions and surveillance), and organisation of data collection including data quality control.
18. Basic data sources for sample frame preparation are the Central Population Register and the phone book. We utilise CATI and CAPI modes of data collection. We do not have CAWI yet, but we are planning the beginning of the project in the next year for development and introduction into regular production.
19. If possible, data collection is done only by phone; otherwise we use data collection modes in sequencing order (from less to more expensive). In longitudinal surveys the first wave is done in the field, for the other waves we use the sequencing order (from less to more expensive). The only exception is household consumption, because of the diary, where we perform only field data collection with interviewers.
20. For interviewing we hire external workforce. Procedures for interviewer selection, instructions and control are being constantly improved.
21. We are facing changing rules for hiring interviewers from the government (restriction of some employment forms) and we are forced to prepare different solutions which include also the possibility of involvement of our employees in the interviewing. As a result we are considering the reorganisation of work in SURS.

III. Lessons learned and future plans

A. Lessons learned

22. Transition of address lists for business entities from separate files into the central system enabled us to have a deeper insight into reporting unit burden. This system is also the basis for further plans for development of communication with reporting units and the basis for development of coordinated sampling. This year we are modernising a contact centre for units cooperating in surveys on business entities. This will enable tracking of communication and better insight into eligibility of the business entity. We would like to improve communication with the respondents and motivate all users to be actively involved.

23. We are starting to improve the phase of manual micro-level editing. It will be introduced into regular production next year. Almost half of the staff in the Data Collection Division is involved in this phase. With further introduction of semi-automated micro-level editing only harder mistakes will be manually edited. Semi-automated editing is implemented already in data from censuses, some data collections done by our partners and a few surveys conducted by SURS (SILC, ICT, etc.) and will be further introduced into the data editing processes. This will open the possibility of transferring the employees' knowledge towards enhancing the communication with reporting units.
24. With the increasing number of surveys where e-reporting is enabled, we will offer the units more friendly reporting. On the other hand, SURS will reorganize the processes and gain resources for enhancing the cooperation with important reporting units. Most of the development connected to e-reporting processes in the future will be oriented into simplification and standardization of questionnaires and modernizing and standardizing the process flow within SURS according to the principles of lean production. This will form a good foundation for less "stove-pipe" oriented processes.
25. We have established good cooperation with administrative sources regarding communication channels and correction of recurring logical failures in the data in some cases. We intend to improve the process by streamlining the data flow and we need to set up a single entry point. Different administrative input data flows will be managed centrally using a common system for exchanging governmental data.
26. Stability of employment is one of the key factors for assuring quality. The global crisis had an impact on reducing costs also in this part of the process. The consequence is higher turnover of interviewers.
27. Despite the work already done in decreasing the consumption of resources in data collection from persons, households and farms with optimization, increased control and lowering the payment, we still need to improve optimization of the resources. We are seeking the solutions for management of interviewers, which is the most demanding activity in terms of time.
28. The documentation of the processes developed and enhanced throughout the years. Though we never really tackled standardisation, we strive for it. With the development and current situation we need to further strengthen standardisation. A step forward is the publication describing the process of the execution of statistical surveys and quality guidelines for them. This is the basis for regular verifications of survey process flow according to the guidelines and improvement of deficiencies in surveys. We introduced into regular production the standard for preparation of paper questionnaires that is strictly controlled.

B. Future challenges and plans

29. With the introduction of e-reporting for business units as well as initiating e-reporting for persons, households and farms, we expect to stop the decrease in response rates and improve the reporting process. In addition, we will transfer the internal workload from less demanding tasks to enhancing communication with important reporting units. We are already involved in the projects dealing with improving the communication (BLUE-ETS and some grants) and have performed some cognitive tests. We are planning to enhance cooperation with business entities by taking into account common advantages, their expectations, concerns and motivation and actively engaging in seeking solutions. We are planning to establish a team of field assistants for important units.
30. For decreasing the reporting burden on business units, we will continue to work on coordinated sampling. The source of data will be address lists where we have the data in which surveys the unit is included.

31. Nonresponse is from the point of view of quality the most important factor. Unfortunately, it is connected with higher costs. We are performing different activities for preventing or decreasing nonresponse (tracking and considering by results, i.e. weighing and imputations). We should improve this process more systematically in the following years.
32. Information is another part that can help in streamlining the data collection process. Thus we will focus on new possible “big data” locations and investigate data flows in the environment. As with every other data collection, we will need to determine new standards for methodological and technical aspects of the data collection process.
33. So far we have focused on central points for data collection from business entities and from persons, households and farms, but in the future we will have a standardised single entry point for data collection from administrative sources. This will enable the execution of some data protection procedures, especially for sensitive data, already in this point and will enable improvement of information security throughout the statistical processes. The single entry point for the administrative data should become also the single exit point for exchange of micro-data with authorized producers of the Slovenian statistical system. This will enable us to exert better control over the data that should be protected.
34. We will supplement the system for translation of PIN into the Statistical Identification Number (SIN) with protection of data on names and addresses. With this we will achieve better optimization and security of data in SURS.
35. As administrative sources are more and more significant in statistical data production and the government is striving towards lowering administrative burden, we should take into account proper communication at all levels with administrative sources, e.g. tracking changes. We will try to find ways to further strengthen the cooperation with administrative sources and better organisation of all partners involved in the process in SURS (statisticians, IT experts, methodologists).
36. New technology solutions (i.e. new e-reporting system, single entry point) for supporting data collection processes will follow internationally comparable process-oriented models and standards (i.e. GSBPM, GSIM). Introduction of a new statistical business register will improve data collection processes. Common process phases will be identified as foundation for development of standardized reusable components, solutions and methods. A big challenge is to develop such electronic system, solutions and e-services to really improve efficiency, reduce burden and enhance statistical quality.