

DATA QUALITY ASSESSMENT METHODS AND TOOLS IN SSO – MACEDONIA

Dejan Peeski

State Statistical Office, Macedonia; dejan.peeski@stat.gov.mk

Abstract

Quality is of primary importance in the world of statistics. The State Statistical Office of Macedonia (SSO) is committed towards producing timely and internationally comparable statistical data with high quality.

The main aim of this paper is to share SSO's experience in the implementation of different methods and tools for data quality assessment: quality reports, quality indicators, self-assessment checklists and user satisfaction survey.

The production of quality reports in the SSO can be distinguished between user-oriented and producer-oriented quality reports. The user-oriented quality reports are based on the Euro-SDMX Metadata Structure and the SSO is publishing the reference metadata on the Internet in a standardized manner. The producer-oriented quality reports are based on the ESS Standard Quality Report Structure. The ESS standard quality and performance indicators are regularly calculated through the statistical process and are included in our quality reports. In order to further improve the production processes and statistical products for better compliance with the European quality standards, the SSO introduced the European self-assessment checklist for survey managers - DESAP as a process-oriented way to discuss and improve the quality of both: processes and products. Our Office is focusing on increasing the use of administrative and other secondary data sources for the production of statistics. This approach makes our Office highly dependent on the quality of those sources and this is the reason for introducing a checklist for the quality evaluation of the administrative data sources. User satisfaction survey is conducted regularly and the purpose of the survey is to obtain the opinion of different users about the products and services of the SSO.

As a conclusion, we can declare that the SSO has a well-defined systematic approach to quality assessment, and the quality is ensured by using different methods and tools, which are harmonized with the ESS methods, tools and methodological documents.

Keywords: quality reports, self-assessment, user satisfaction survey.

1. Introduction

The vision of the State Statistical Office of the Republic of Macedonia (SSO) is to be recognized as an institution that provides high quality, timely and relevant statistical information by adhering to the 15 principles of the European Statistics Code of Practice (CoP), which is considered as the main guidance in fulfilling this commitment. CoP, through the Law on State Statistics, is accepted as a document that establishes the framework for quality management in the SSO and at the same time it encourages other participants in the statistical system to implement the principles and to work together on increasing the confidence in official statistics. Production of high-quality statistics depends on the implementation of a quality assurance framework and the regular assessment of data quality. Without a systematic data quality management, there is a risk to lose control of the various statistical processes such as data collection, editing, weighting, etc., therefore having an impact on data quality. In the past, the quality of statistical products was observed mainly through the criterion of accuracy of the estimates, but nowadays the quality of statistical products has to be evaluated through the multi-dimensional approach, where the accuracy of the estimates is just one of the observed dimensions.

Since 2009, the SSO has worked systematically on data quality assessment and has significantly increased the number of produced quality reports. The preparation of the quality reports is now a common practice. In recent years, a lot of employees were included in the process of preparation of quality reports, especially user-oriented quality reports. But it is important to note that over the years there has been a substantial increase in the number of produced quality reports intended for producers. In order to further improve the production processes and statistical products to comply better with the European quality standards, since the end of 2013 the SSO has introduced the European self-assessment checklist for survey managers - DESAP as a process-oriented way to discuss and improve the quality of processes and products. Assessing the quality of data from the users' perspective is in line with the view that quality is to be decided by the user and in relation to the stated and implied needs of the user. In order to obtain the opinions of users on products and services, the SSO has been

conducting a User Satisfaction Survey to measure customer satisfaction since 2009. The SSO is also focusing on strengthening the cooperation with holders of administrative data in order to reduce respondent burden, and to increase utilization of administrative data. Using data from administrative sources has its advantages and disadvantages. One of the disadvantages of using administrative sources is that they can be of high risk if they have not been previously assessed. The quality of data from administrative sources is ensured via the adoption of appropriate standards and procedures for their implementation, and these standards are the framework for governance and accountability for data quality.

2. Data quality assessment methods and tools

2.1. User- and producer-oriented quality reports

In early 2009, Eurostat published the ESS Handbook on quality reporting and ESS Standards for Quality Reports (ESQR) and it was easy to recognize that many of the quality and performance indicators from the ESQR can be clearly identified in the "Euro-SDMX Metadata Structure" (ESMS). The compilation of ESMS concepts in a standardized quality report is advantageous for supporting internal self-assessment, for reporting to Eurostat and also for user-oriented quality reporting because it puts considerable emphasis on output quality.

Reference metadata in the SSO exist in various kinds of shapes and storage; in particular there was no officially promulgated and adopted standard at the organizational level. Before actual implementation of ESMS started, in order to achieve better harmonization, SSO's three-level Theme/Statistics structure was converted into a five-level structure. The old Theme/Statistics structure consisted of: Statistical Area, Statistical Sub-Area and Statistics. The revision was done in accordance with the Eurostat Compendium structure: Domain, Theme and Module. Furthermore, a breakdown was made of the module level into two additional levels: group level and level of statistics. The practical side of the work done in the SSO since June 2009 resulted in the development of ESMS Database and ESMS Application. Approximately 90 user-oriented quality reports are produced in Macedonian and English language and all of them are available on SSO's website.

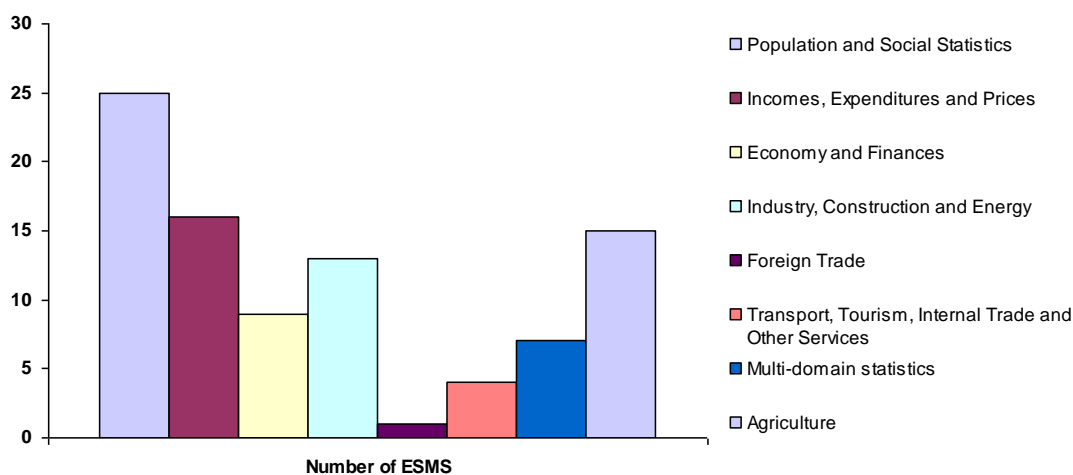


Figure 1: ESMS by statistical domains

Quality reports are important not only for users but also for producers of official statistics. Producers need to have a picture of product quality in order to see the results of earlier production developments and to identify the points of further improvements and for monitoring purposes of the management. That is why they need the most detailed quality reports and a number of quality indicators, involving the processes behind. Quality reports and indicators provide documentation of the quality features of statistical products. They are the key reference documents for quality assessment. For this reason, they form an important input for auditing and self-assessment.

Producer-oriented quality reports in the SSO are prepared according to the ESQR structure or other structures received from Eurostat. For most of them, ESS Metadata Handler is used for transmission to Eurostat. The SSO produces quality reports for several domains like: LFS, EARNINGS, LCS, EUSILC, INFOSOC, CIS, FSS and others. Also, there are still some reports that are prepared only for internal use, like reports for Retail and Wholesale Trade, Statistical Business Register and others. ESS Quality and Performance Indicators, such as sampling error, over-coverage, unit non-response, item non-response, imputation rate, punctuality - delivery and publication, time lag of first results, time lag of final results, etc., are regularly calculated, monitored and included in the quality reports.

2.2. DESAP self-assessment checklist

Self-assessments provide valuable input for the evaluation of statistical products and processes. In comparison with other assessment tools, self-assessments have the advantage of causing less effort and lower costs. Since 2007, three self-assessments according to the Common Assessment Framework (CAF) have been conducted in the SSO. The ESS CoP was officially accepted in 2009, and a Light Peer Review was undertaken in 2010. In the current year, a new round of Peer Review will be undertaken. What was missing in this quality assessment chain in the SSO was self-assessment of the survey managers.

The original DESAP self-assessment questionnaire was developed in European cooperation in 2002 and 2003. In order to improve the production processes and statistical products further to comply better with European quality standards, the SSO has introduced the European self-assessment checklist for survey managers. It is not the original version, but the extended German version of DESAP, which incorporates a number of improvements based on a series of pilot self-assessments in Germany. It is a generic questionnaire for systematic quality assessment of surveys and is considered as a valuable tool for taking a systematic and in-depth look at statistics and the underlying production processes. It has been designed as a tool for survey managers and should support them in assessing the quality of their statistics and considering improvement measures. Its main goals are the comprehensive quality assessment of a single statistical domain as well as the generation of improvement ideas. It follows a process-oriented approach, which means that it goes along with you in thought during the whole process of a survey from the very early stages until the final steps. The questionnaire is structured according to the Generic Statistical Business Process Model (GSBPM) and is divided into 9 chapters.

DESAP has been identified as a useful tool for the systematic process-oriented assessment of statistics and has been introduced at the SSO in 2013. The pilot use of DESAP aims specifically at increasing the quality awareness of staff of all levels while at the same time getting a better overview of the production processes used in the different sectors. It was translated into Macedonian language and two pilot self-assessments based on the this version

of the questionnaire were assessed: Household budget survey and Monthly report on the value of completed and contracted construction works, total number of employees and effective hours worked. Based on the pilot experiences, the questionnaire was further improved and several issues were concluded: relevance was assessed slightly weaker than other areas mainly because of lacking knowledge about user satisfaction, the subject-matter statisticians in both surveys need to invest considerable effort in the identification and improvement of erroneous raw data. When evaluating the pilot self-assessments all participants in the team groups agreed on a positive appraisal of the self-assessments. However, DESAP has been considered as an occasion-based quality instrument and will be used for: surveys with major changes to concepts and/or methodology, new surveys after the 2nd or 3rd wave and surveys with known quality problems.

Also, a procedure exists for filling-in the self-assessment questionnaire and it contains: plan and timetables, necessary actions and measures for calculating and providing quality indicators, discussion and consensus on the findings, report on the self-assessment defining strengths and weaknesses, action plan, involving teams for implementation of the action plan, submission of the report to the top management, implementation of the action plan.

2.3. Quality evaluation of the administrative data

The SSO is focused on increasing the use of administrative and other secondary data sources for the production of statistics. This approach imposes the need of assessing the quality of those sources. In the last two years, the SSO intensified the work associated with the holders of administrative data sets. In order to strengthen the cooperation with the holders several work meetings were held to define the necessary standards for using administrative data for statistical purposes. Memoranda and agreements for electronic data exchange were signed with the Ministry of Interior, the Public Revenue Office, the Employment Service Agency, the Pension and Disability Insurance Fund, the Health Insurance Fund, the National Bank, the Customs Administration, the Civil Registry Office, the Ministry of Agriculture, Forestry and Water Management, the Central Register, the Chambers of Commerce, the Economic Chamber of Northwest Macedonia, etc.

Before signing agreements for data exchange with administrative data holders the following activities (procedures) are done:

- Detailed analysis of data and their coverage, content, timeliness and frequency
- Comparison of statistical and administrative concepts, definitions, classifications and nomenclatures
- Assurance of quality of administrative data i.e. control procedures used by the provider

The above-defined activities (procedures) were implemented for several statistical surveys. One example is the use of administrative data on labor market from the Public Revenue Office (PRO). The administrative data which were provided from the PRO were previously extensively reviewed and analyzed, but at entity level. Additional analysis of the variables for the given data was done and in the further contacts with the PRO the possibilities for delivering the data at individual level in the entities required for the Structure of Earnings Survey were discussed. On the basis of the analysis it was concluded that certain data could not be delivered from the PRO in the foreseen deadline. Therefore, it was decided to establish a combined method for data collection for the Structure of Earnings Survey, which implies using certain data from the PRO's database and collecting additional data via paper questionnaire facilitated by the variables obtained through administrative sources.

One of the priorities of the SSO, stated in the Strategic Plan 2016-2018, is strengthening the cooperation with providers of data and holders of data sets. Main objectives are: inclusion of the SSO in the process of creating and amending the data sets, greater use of administrative data sources in regular statistical production, development of new data collection systems, reducing respondent burden, signing agreements for cooperation and data exchange with holders of data sets and measuring the satisfaction of data providers. For achieving the planned priorities and objectives the following program has been defined: Strengthening the cooperation with providers of data and holders of data sets, sub-program: Establishment of a system for measuring the quality of administrative data sources. "Preparing documentation for measuring the quality of administrative data sources and their application via testing" will be used as a success indicator for assessing the implementation of this program.

The SSO has recently started the activities for assessing the quality of administrative data sources in a systematic manner. Using administrative data has its advantages and disadvantages. Main disadvantages are: they are collected for different purposes (not statistical), collection and storage is outside of the control of the SSO, the existence of incompatibility of variables, definitions, classifications, concepts, storage and protection of data, etc. Hence is the need to define indicators to assess the quality of any administrative source. For that reason, the SSO has prepared a Framework Questionnaire intended for administrative sources. The questionnaire contains general questions that need to be asked of any new or already used administrative source, for making decisions related to the quality of administrative data. The components that are included in the questionnaire are related to: institution, relevance, concept and definitions, privacy protection and data confidentiality, data delivery and treatment of data. Based on the answers in the questionnaire, the SSO will monitor the process of evaluating the answers. The process of evaluating will be done through the “Checklists of indicators”. The Checklist assesses the three dimensions of quality: source, metadata and data. Each dimension has its own quality components, and each component consist of quality indicators. Quality indicators are measured by one or more quantitative or qualitative methods.

The SSO, together with representatives from 18 other institutions, holders or users of administrative data sets, took part in the project “Improving the quality of data in the systems of state institutions”. The outcome of this project was a document that describes the standards for data quality in state institutions systems. The data quality recorded in the registers is ensured via the adoption of appropriate standards and procedures for their implementation. The standards are the framework for governance and accountability for data quality, with continued commitment to providing high quality data in any organization. The quality standards of data contain mandatory guidelines that should be applied in the creation of new electronic registers by state institutions. Existing electronic registers must comply with the standards within three years after they take effect. The following eight characteristics of data quality are determined for the purposes of the scope of these standards: accuracy, validity, reliability, timeliness, completeness, relevance, consistency and conformity.

2.4. User satisfaction survey

Assessing the quality of data from the users' perspective is in line with the view that quality is to be decided by the user and in relation to the stated and implied needs of the user. In order to obtain the opinions of users on products and services, the SSO conducts a survey to measure customer satisfaction. The first survey was conducted in 2009, the second one in 2012, and in 2015 the SSO conducted this survey for the third time. The questionnaire was designed to obtain information about the habits of users in the use of statistical data, their views on the statistical data, the principles of the European Statistics CoP relating to statistical outputs, the services of employees with regard to provision of data and information, the web site, as well as to obtain basic demographic data on participants in the survey.

The survey included several types of users that are recorded in the registry of users: public administration, business entities, students, scientists, media, NGOs, international organizations with representatives in Macedonia and private users. 2098 users recorded in the registry of users with email were invited to take part in the survey. 337 persons or 16.6% of the total number of invited persons participated, which is slightly higher than 2012 when 308 invited persons participated in the survey. According to the group of users, the majority of respondents are from the public administration and local government, 30.9%, followed by business entities with 22.0% and scientists with 11.3%.

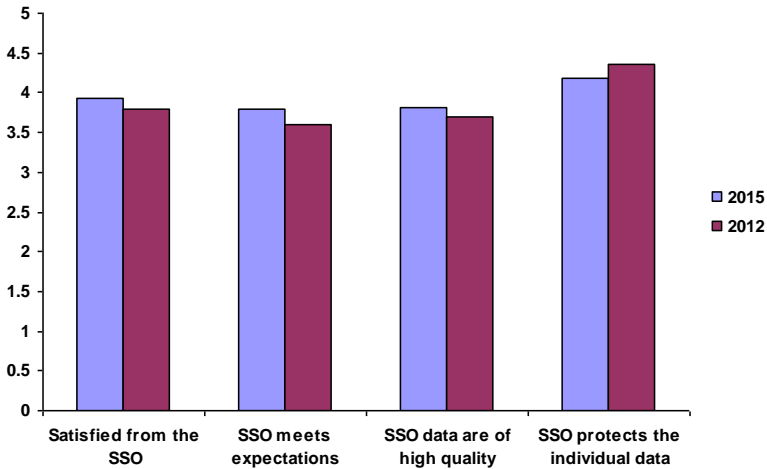


Figure 2: Questions related to SSO

In the set of questions related to the SSO (Figure2), users evaluated the degree of agreement with statements related to SSO general satisfaction, expectations, quality of statistics produced and the protection of individual data.

In the set of questions related to statistical data produced by the SSO (Figure 3), users evaluated various aspects of basic statistical principles such as the availability of data, use of appropriate statistical methods, clarity, information on statistics, data comparability and timeliness.

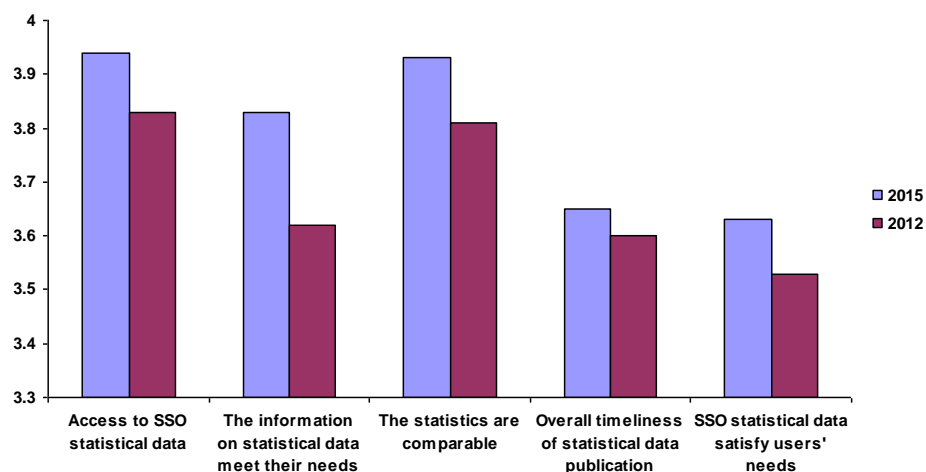


Figure 3: Questions related to statistical data

According to the answers of the respondents on questions related to the habits, the greatest part, or 39%, use the data less than once a month, while 34% of the users use the data very often (on a daily basis or at least once a week).

3. Future challenge

In order to decrease the reporting burden of employees for production of the two existing ESS report structures, ESMS and ESQRS, the SSO should consider the implementation of a “Single Integrated Metadata Structure” where all statistical concepts are included and streamlined, by assuring that all concepts therefore will be reported upon only once. The broader implementation of DESAP should be accompanied by a plan for implementation and

knowledge about the instrument, and the self-assessment approach needs to be thoroughly spread in the SSO. Also, the SSO should continue to increase the use of administrative data sources for statistical purposes. In order to strengthen the cooperation and communication with users the objective of the SSO in the near future is to simplify the access to statistical data, to modernize the manner of disseminating statistical data, to increase the amount of published data, to provide data with longer time series and to continue with measuring user satisfaction. Another challenge this year will be the forthcoming Peer Review, which can be classified as a special case of an external audit, taking as a basis the principles and indicators of the European Statistics CoP.

4. References

Eurostat (2007), Handbook on Data Quality Assessment Methods and Tools

Papazoska H., Karajovanovikj B., Lipikj S. (2010), The value of adopting implementation of ESMS structure in Macedonian State Statistical Office

SSO (2015), Strategic plan 2016-2018

SSO (2015), Report from the User satisfaction survey

SSO (2015), Assessment of Quality of Administrative Data Sources

MISA (2016), Standards For Data quality in State institutions systems

Internal SSO documents