**IMPROVED QUALITY IN STATISTICAL PRODUCTION BY USING STANDARD PROCESSES IN PHASES DESIGN, BUILD AND COLLECT**

Biljana Ristevska Karajovanovic, State Statistical Office of the Republic of Macedonia (SSO), [biljana.karajovanovic@stat.gov.mk](mailto:biljana.karajovanovic@stat.gov.mk)

Valentina Todorova Hristovska, State Statistical Office of the Republic of Macedonia (SSO), [valentina.todorova@stat.gov.mk](mailto:valentina.todorova@stat.gov.mk)

# **Abstract**

*Metadata describe comprehensively statistical data and processes, assist in their interpretation and they are fully integrated in statistical processes and represent a prerequisite for developing advanced tools for statistical production, enabling high quality statistics.*

*Standardization and harmonization of statistical variables and code lists are essential for improvement and standardization of processes and at the same time, they represent a base for a single standardized production line for all surveys.*

*Electronic methods offer new challenges and opportunities to improve the efficiency of statistical processes and get high quality incoming data, reducing costs at the same time.*

*As a tool for electronic data collection, SSO is using eSTAT system. eSTAT is an IT system for automatic generation of web forms collection, on-line data collection, communication with respondents, data validation and loading collected data into final observation register.*

*All the logical expressions that are frequently used in the process of data validation have been covered by an incorporated engine for generation of the validation rules in eSTAT.*

*Statistical Task is responsible for creation of web form instance for the specified survey and its reference period, conducting the collection process according to defined dates and running mass e-mailer for confirmation and reminders for respondents. All the activities in the process of questionnaire generation have strictly determined order and they represent a highly standardized process in the eSTAT system.*

**Keywords**: statistical variables, statistical processes, standards, data collection, respondents

**1. Introduction**

The web-based survey offers significant advantages over more traditional survey techniques, however there are still serious methodological challenges when using this approach.

The process of collecting and processing data is the most important part of the statistical production, and modern business statistics faces with new challenges requiring new concepts and methods of data collection. As the Internet has a profound effect on the survey research industry, National Statistical Institutes surveys endeavor to replace traditional methods of survey data collection with electronic ones.

Electronic methods for data collection provide new opportunities for improving the efficiency of statistical processes, obtaining high-quality incoming data and reducing costs at the same time.

Electronic statistical data collection system (eSTAT) is an online application for completing statistical questionnaires. It includes infrastructure for automatic ‘metadata–based’ generation of intelligent web forms (web questionnaires). It also supports maintenance of metadata and raw-data database, maintenance of respondents’ database and mass e-mailer for confirmation and reminders.

The application is developed in C# on the .Net Framework 4.0. The underlying database is Microsoft SQL Server. The web application uses third-party controls and packages such as: NCalc for evaluating the mathematical expressions, ELMAH for error logging, AjaxControlToolkit for customizing and reusing web extenders and controls.

This system for automatic generation of web forms for on-line data collection has all the necessary data and network security measures. The eSTAT system design ensures high level of scalability, particularly from the point of view of volume of data and number of users.

Electronic data collection eSTAT is based on single standardization production line for all surveys, but it is mainly intended for survey respondents, which are business entities (enterprises).

**2. eSTAT workflow**

The SSO is faced with new demands from the users as well as with new developments in Information Technology. Nowadays, web methods for data collection provide new opportunities toward improving the efficiency of statistical processes and obtaining high-quality incoming data. The three most common reasons for choosing web survey over traditional paper-and-pencil approaches are decreased costs, faster response times and increased response rates.

eSTAT is an IT system for automatic generation of web forms, on-line data collection, communication with respondents, data validation and loading collected data into final observation register, and it is a comprehensive metadata-driven electronic data collection system which consists of two parts: 1. Metadata-driven generation of questionnaires and 2. On-line data collection, applicable for almost all standard statistical surveys.

*2.1 Metadata-driven generation of questionnaires*

Aware that Internet technology greatly extends the possibilities of questionnaire design, we have started research on web questionnaire design and the basic principles for designing web questionnaires. We have dedicated a lot of time analyzing existing paper questionnaires from various statistical surveys, where respondents are enterprises. We have captured all types of respondents’ answers with the following available major choices: single-response and multiple-response questions, response options into a grid or matrix.

As a result of our analysis, we have developed a global frame for a unique model for designing and generating web questionnaires. The eSTAT is a comprehensive metadata-driven system which contains a module for collection and maintenance of metadata.

The main activity of this module was determination of all necessary metadata, especially structural metadata: variables, code lists and standard classifications.

Depending on the types of respondents’ answers, the model recognizes two types of variables: elementary and multidimensional.

Elementary variable is a variable where for the relevant question only one answer is obtained (single-response).

Multidimensional variable is a variable where for the relevant question a set of answers (set of data) is obtained and its expression is matrix-column or two-dimensional matrix.

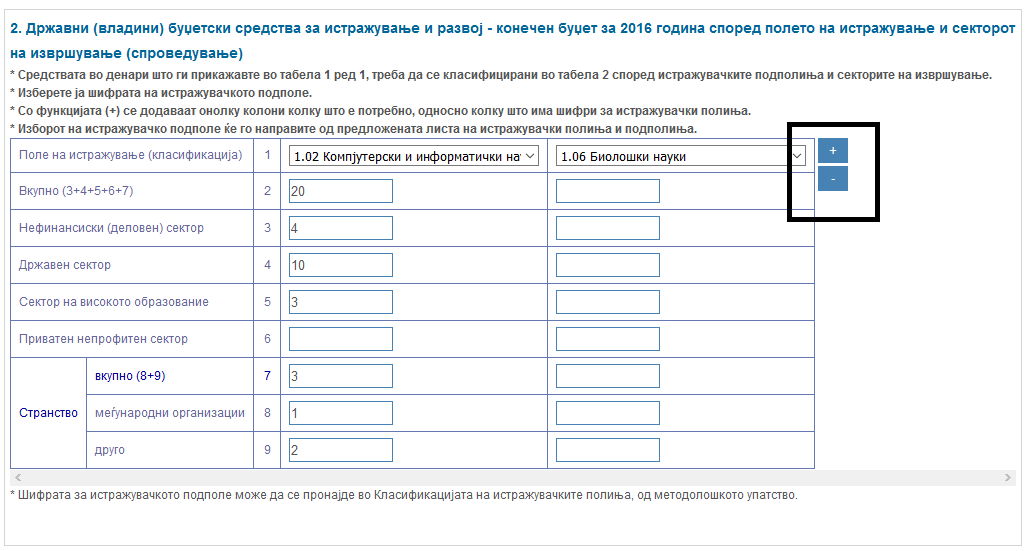
Every questionnaire contains a set of address data and some basic information on enterprises, called Identification variable, which is a multidimensional variable of matrix-column type.

The input tool in web questionnaires also serves as a visual guide for respondents. For example, radio buttons permit the selection of one and only one response. Drop-down lists similarly constrain the selection to one of the available options (code list). The code list is actually denoted as a variable value domain.

The matrices are mainly used for entering numerical data, and the system allows other input formats. For example, each matrix field can receive a value from a code list or a text value.

The eSTAT system enables the creation of dynamic matrix columns according to the needs of the respondents.

Figure 1: Dynamic matrix column



Text fields and text areas allow respondents to enter unformatted text. The text areas are presented by large open boxes with size defined by the designer.

For more practicability in redundancy and reusability of variables through different statistical surveys, we have decided to group the variables in three main categories: general variables, subject-matter area variables and single-survey variables. General variables are common variables for any statistical survey. Subject-matter area variables are common for a specific subject-matter area. Single-survey variables are suitable only for one statistical survey.

The system supports the creation of versions and variants of the code lists.

The functionality of the system is confirmed by the possibility of creating metadata objects, their versioning, multiple use, the possibility of establishing relations and linking between the metadata objects for generating web questionnaire final objects.

The process of generation takes place in two main phases. First, the web form is generated and it obtains the status “Completed Questionnaire”. It means that all questions, variables, and additional information are presented on the web form. Then, the responsible person confirms the web form and it is ready for validation rules implementation.

*2.2. Standardization of validation rules in data collection*

Validation rules in eSTAT are created by using the GUI called rule engine. Rule engine is a software [component](https://whatis.techtarget.com/definition/component) of the system that allows statisticians to add or change the validation logic to be applied in the data collection process. Rules are standardized in accordance with the business logic used in survey operations in daily statistical production. Rules are classified into seven categories according to their semantics and level.

At the cell level, we distinguish three categories of rules: cell marked as "no entry", cell marked as mandatory one and rule for auto-populating cell values from another cell value.

At the question level, routing rules are defined and applied to any subsequent survey question independently of the question style, format or response type.

Hierarchical relationship between the value domains is also defined at the question level in order to limit the possible answers for another question according to the chosen value domain item (e.g. items in municipalities code list to be limited depending on chosen region).

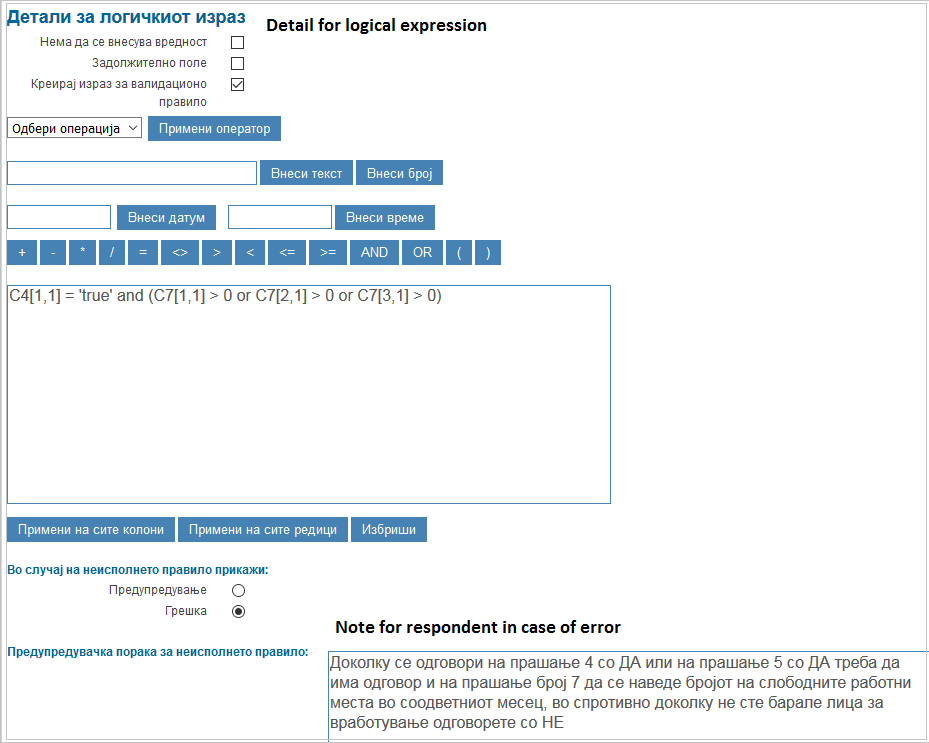
Any mathematical-logical expression is defined as a finite combination of standard math and Boolean symbols applied on different question values. The expression rule is written at the question level and is well formed according to context and math rules.

Dependency validation rule is a special type of rule created at the question level, as well. With this type of rule, possible question answers could be limited if defined mathematical-logical expression applied on multi-question values is satisfied.

Each validation rule could be defined at higher level (full) or at lower level (less), i.e. there is a possibility the web questionnaires to be run in less or full validation mode dependent on the role in the system (respondent or subject-matter staff). Also, each validation rule could be defined in the system as suppressible (only supported by informational message) or not. For expression rules and auto-populated values, user messages are customizable.

Validation rules are integrated in runtime production environment by using the NCalc open-source mathematical expressions evaluator for .NET platform. NCalc is able to parse any expression and evaluate the result, including static or dynamic parameters and custom functions.

Figure 2: Expression rule



**3. On-line data collection**

Considering survey costs and speed of response, email invitations and notifications are preferable: messages are sent and received immediately, costs are low and it is easy to offer or dispose any additional information. When email invitations are used, the web questionnaire is already at hand, only a single click is needed to move from the email message to the survey questionnaire in a web browser, especially when respondents have a permanent Internet connection. Sending out individual notifications to respondents is far more effective than general notification, so we have chosen list-based web surveys. List-based web surveys use individual invitations to respondents from a list. This list can be a sample from a representative sampling frame.

*3.1. Respondent management*

The eSTAT system contains a module for respondent management, which is responsible for maintenance of respondents’ database and mass e-mailer for notifications. Respondent lists can be imported into the system using a specified template, performing validation of the contents of data from respondent lists using system integrated procedure.

At the first appearance of the respondent in the imported list, the system creates a user name and password and sends main invitation to the respondent, so that all necessary information for using the application is provided. On the survey’s start day, the respondents get next notification whose goal is to motivate respondents and provide them with the necessary information for answering the web questionnaire. The respondents are reminded about the legal obligation to provide data. They are also provided with additional contact information and are given a deadline for response.

During the time period for data providing the respondents get one reminder notification and one warning and notification for confirmation after successfully competed web questionnaire. The large number of notifications may have a negative impact on respondents' responsiveness, and that is the reason we have decided to send them two notifications for reminding.

The most important and essential part of conducting web questionnaire and organization of the collection process is Statistical Task.

Statistical Task is responsible for creation of web form instance for the specified survey and its reference period, conducting the collection process according to defined dates and running mass e-mailer for notifications and reminders for respondents. All the activities in the process of questionnaire generation have strictly determined order and they represent a highly standardized process in the eSTAT system.

For each subsequent reporting period it is necessary only to create a new statistical task and link it with a version of an already existing web questionnaire.

*3.2. Automated communication with respondents and monitoring the process of data collection*

For a number of statistical surveys, data collection is optimized and standardized by implementing the eSTAT system. Standardized interface and working dashboard is available for respondents, statisticians and data entry personnel in the SSO. The questionnaire web forms and dashlets could be available in multiple languages (option provided by multilingual metadatabase).

Business entities that are included in surveys can access their personalized dashboard by logging in the system with their username and password. The dashboard is the communication channel with the SSO concerning the surveys in which they are involved through on-line reporting. Data collection information flow starts with presenting survey information on respondents' dashboard and by sending notification e-mail. On the dashboard, web forms for active surveys and their statuses are displayed and respondents can submit survey data to the SSO until a predefined date for closing data collection. For submitted data, they can view and export web questionnaires with already reported data to the SSO. Information from respondents' previous reporting periods is also available for retrieving and preview.

Respondents are able to complete the questionnaire on the browser and securely exchange confidential information back to the SSO. After data are submitted, validation rules are implemented as defined for certain type of respondent, and either message for successful submission or an error message appears. If the error is defined as suppressible, the respondents have an option to ignore the message and continue to submit data.

All data after successful validation are transferred to a database on SSO server. The duration of the session is limited by certain time of inactivity (15 min). On session expiration, auto-save sequence is performed for the questionnaire to be saved. User identification data from the current authentication session are collected automatically with the data. The complete history of changes is kept, together with metadata showing user ID, date and time of modification.

Submitted questionnaires cannot be changed again by the respondent. If survey verificator permits a second try, then the respondent is able to correct the questionnaire. Target users get different types of e-mail notifications that are also presented on the dashboard screen.

Two types of survey verificators are defined and implemented into the system. Verificators have the same personalized dashboard as respondents. All information and web questionnaires related to active surveys are shown by statuses. Verificators have the possibility to mark "New" questionnaires as paper ones and to unlock the "Submitted" ones for further correction by respondents.

Only web questionnaires with status "Submitted" are editable by first-level survey verificators, while other types can only be previewed. Corrections are made directly in the submitted data based on implemented high-level validation rules, and statuses of web questionnaires are changed accordingly. Questionnaires verified by first-level verificators are only editable by second-level verificators.

Second-level verificators are able to correct web questionnaires only if they have already been checked or verified at the first level, and statuses are changed in accordance with the actions taken.

The system automatically generates new questionnaire instances with appropriate statuses after two important dates defined in the eSTAT flow: date for closing data collection and final survey closing date.

Data could be exported from the system at any time since the survey process begins. Data correction is possible until the final survey closing date, either enforced by high-level validation rules or by manual implementation of validation lists of additional (batch) control applied on exported data.

**4. Conclusions**

eSTAT started its production in September 2014 and, until now, 33 different statistical surveys with monthly, quarterly and yearly periodicity have been conducted. The experience with the еSTAT system confirmed the need for continuous standardization of metadata in order to obtain high-quality e-questionnaires.

Collected metadata using a module for collection and maintenance of metadata can be used for further analysis. It also represents a basis for the establishment of Metadata repository. The coding process is almost eliminated with imported standardized code lists in the system.

The eSTAT system enables active communication with respondents and continuous monitoring of the process of filling in web questionnaires and requirements for data corrections. Of great benefit are the inbuilt validation rules for respondents, which provide increased data quality.

Compared to other types of survey systems, the eSTAT system is cost-efficient because there are no costs for administrative and field work. Another advantage of the system is the speed of data collection because it is constantly up and running with immediate access to results, which also has an effect in improving the accuracy of data collection. The response rate in eSTAT is constantly growing in comparison with surveys conducted with different methods.

Web-based surveys are not without problems, questionnaires do not look the same in different browsers and on different devices, respondents may have different levels of IT expertise, e-mail survey delivery often requires large numbers of e-mail messages to be delivered over a short period of time, etc. All these issues could address potential technology problems to be solved. Therefore, web-based surveys present a constant challenge for improving the system with updated and additional functionalities included.

**5. References**

Edith D. de Leeuw, Joop J. Hox, Don A. Dillman (2008), International Handbook of Survey Methodology

[Katja Lozar Manfreda](https://onlinelibrary.wiley.com/action/doSearch?ContribAuthorStored=Manfreda%2C+Katja+Lozar), [Zenel Batagelj](https://onlinelibrary.wiley.com/action/doSearch?ContribAuthorStored=Batagelj%2C+Zenel), [Vasja Vehovar](https://onlinelibrary.wiley.com/action/doSearch?ContribAuthorStored=Vehovar%2C+Vasja) (2006), Design of Web Survey Questionnaires: Three Basic Experiments, Journal of Computer-Mediated Comunication

Milana Karaganis, Marc St-Denis (2012), Electronic questionnaire collection at Statistics Canada, Conference of European Statisticians Geneva, Switzerland WP. 32

Valentina Todorova Hristovska (2015), Standardization and harmonization of statistical variables as prerequisites for electronic data collection, Summer School, Ohrid, Macedonia