

GENERALIZED SOLUTIONS FOR DATA EDITING AT SURS

Rudi Seljak

Statistical Office of the Republic of Slovenia, rudi.seljak@gov.si

Kaja Malešič

Statistical Office of the Republic of Slovenia, kaja.malesic@gov.si

Data editing is a process aimed at providing accurate, complete and consistent information. In general, it encompasses all activities related to the detection of errors, correction of inconsistencies and imputation of missing values in the observed data. Due to its complexity, it is mostly very time and resource consuming. It is therefore rational for statistical organizations to harmonize different editing processes within a variety of surveys with appropriate general technical solutions.

The paper presents the main characteristics of the new approach to data editing implemented at the Statistical Office of the Republic of Slovenia (hereinafter SURS) with the use of a generalized software tool and its impact on the statistical process. This tool, the so-called MetaSOP application (SOP is an acronym in Slovenian for statistical data processing), will encompass also other parts of data processing (e.g. aggregation and standard error estimation, tabulation and tabular protection, quality indicators). At the moment the module for data editing is already in production, while other modules are being tested or developed. The editing module combines three IT environments: ORACLE process metadata database, SAS macros as general programs for data processing and .Net WPF application. The whole system is based on a metadata driven principle. Its core is a general program code consisting of SAS macros for logical checks, deterministic, systematic and individual corrections and imputations. For usage in a particular survey, this general code is parameterized with appropriate metadata entries.

After the implementation of the new editing approach into the process of first surveys we have already noted the first advantages, such as elimination of errors in consistency between entered rules and variables. We expect other positive results, such as rationalization of statistical processes and overall improvement in data quality (also due to different organizational procedures and expected increased number of surveys having the automated editing), to be visible later. The goal is that all surveys at SURS will be gradually included in the generalized process.

Keywords: *metadata driven systems; statistical data editing; generalized software solutions.*

References

Seljak, R. (2009). Integrated statistical systems and their flexibility – How to find the balance?. NTTS Conference, Brussels, Belgium, 5-7 March, 2013.

Seljak R. (2014). Metadata driven application for data processing – from local toward global solution. Conference of European Statisticians. Paris, France, 28-30 April, 2014.