REPUBLIC OF SLOVENIA
STATISTICAL OFFICE RS

GENERALIZED SOLUTIONS FOR DATA EDITING AT SURS

Rudi Seljak, Kaja Malešič
Statistical Office of the Republic of Slovenia

4th Baltic-Nordic Conference on Survey Statistics
Helsinki, August 24 - 28, 2015

Contents

- Introduction
- Development of statistical data processing at SURS
- Main characteristics of data editing in MetaSOP
- MetaSOP IT environments
- Basic architecture of the editing module
- Graphical interfaces and examples
- Impacts of the new generalized approach
- Future challenges

Introduction

- Data editing is a process aimed at detecting and correcting errors
- Rationalization of statistical processes
- The need for transition:
  - from custom made solutions for surveys (stovepipe approach) to generalized process solutions
  - from domain oriented to process oriented production

Development of statistical data processing at SURS

- 2007 generic SAS programs
  - smaller generalized solutions for different parts of process (“building blocks“)
  - meta-driven approach

- 2014 - **MetaSOP** application (SOP =Slovenian acronym for Statistical Data Processing)
  - a general tool for data processing: data editing (in production), being tested or developed modules: aggregation and standard error estimation, tabulation and tabular protection, quality indicators

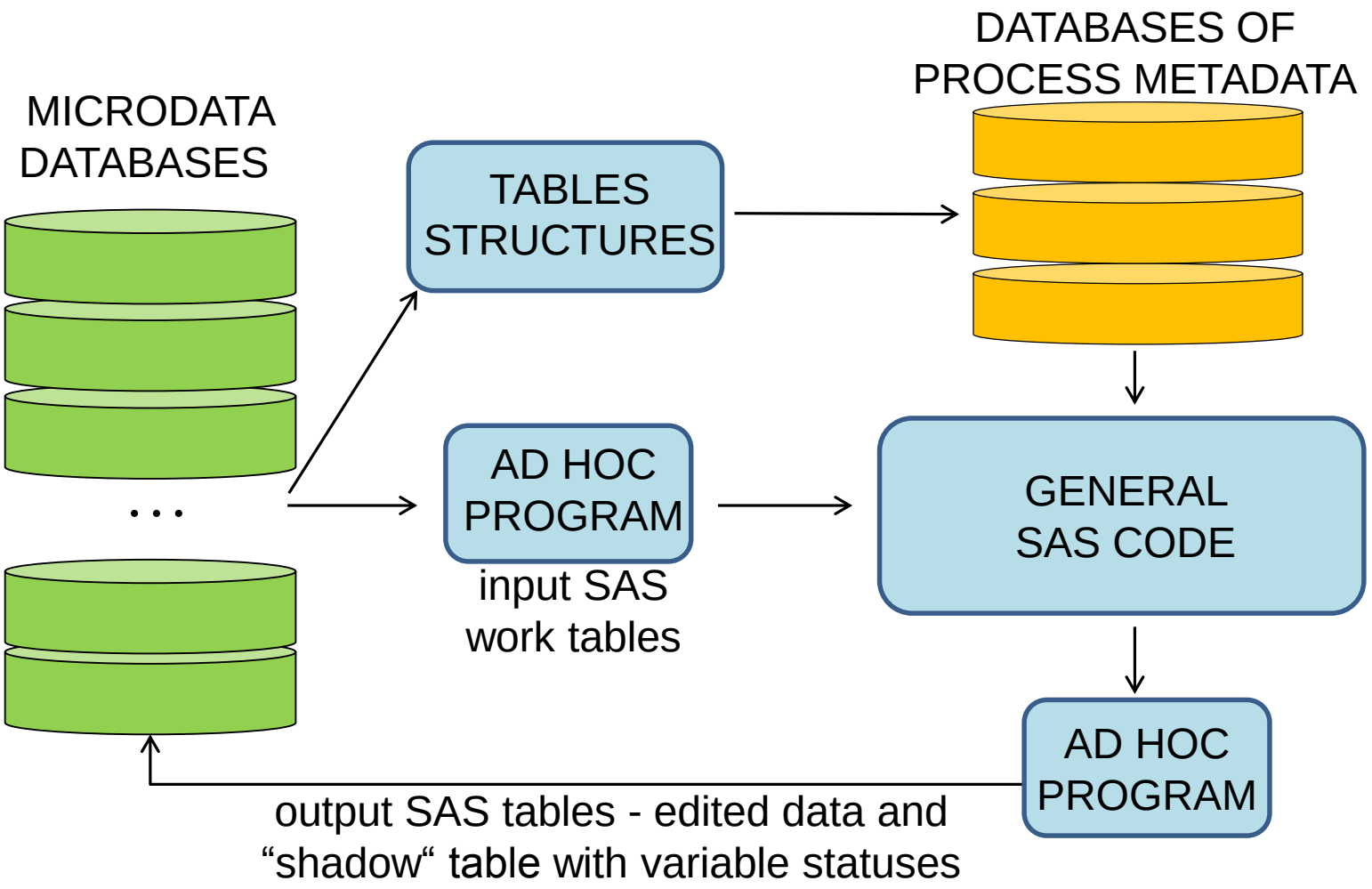
Main characteristics of data editing in MetaSOP

- Traceability, repeatability and reproducibility
  - all changes in data are recorded with statuses of the record (information in which process (phase) the record was changed) and statuses for each variable (information how the change has been performed)
- Meta-data driven approach
- Microdata databases have to be designed according to standard rules
- Standard quality indicators are used to monitor the process

MetaSOP IT environments

- .Net WPF application - the visible part of the system used by subject-matter statisticians
- SAS macros as general programs for data processing
- ORACLE database of process metadata
  - information about tables and variables
  - edit checks and rules in SAS syntax which are entered in the .Net application

Basic architecture of the editing module

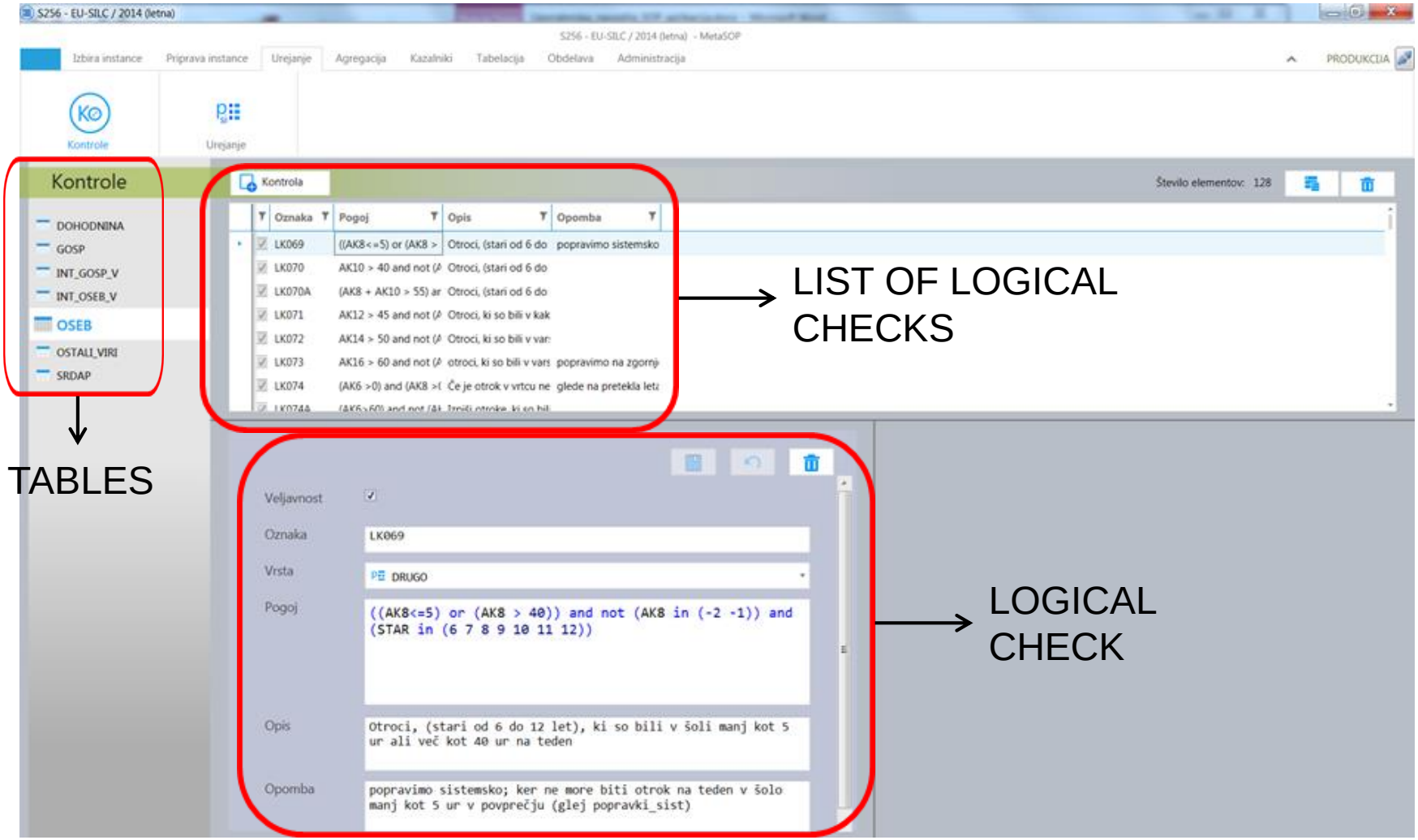


Graphical interfaces

Groups of graphical interfaces:

- Interfaces for selection and preparation of the survey and survey instance:
  - for importing metadata from previous instances
  - for information on variables and definition of new “process“ variables
  - for definition of processes (phases)
- Interfaces for management of the process metadata
  - logical checks
  - deterministic systematic and individual corrections
  - imputations
- Interface for running the processing (running the SAS macros)
- User administration interface

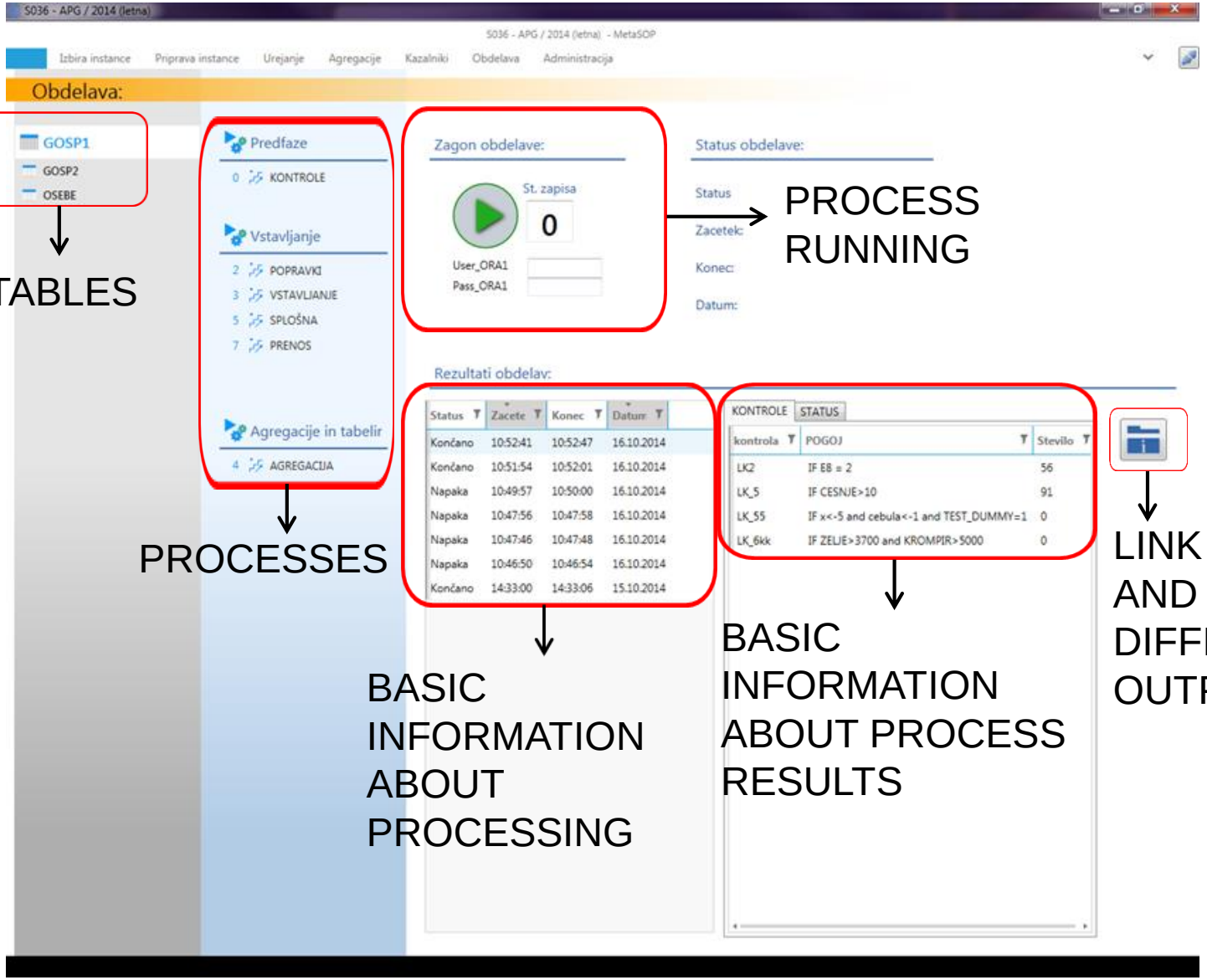
Example - Interface for logical checks



Example - Parameters for a „hot-deck“ imputation

- Variable
- Step (of the imputation process)
- Condition (determines for which units the imputation procedure will be performed)
- Condition for donor (determines which units can serve as the suitable donors)
- Method of imputation (HD in this case)
- Stratum (stratification variables for donors)
- Matching variable
- Minimum number of donors
- Minimum percentage of donors
- Rounding rule
- Note

Example - Interface for process running



Impacts of the new generalized approach

- Main goal is rationalization of statistical processes and overall improvement in data quality.
- Changes in work organization:
  - Different distribution of work among subject-matter statisticians, general methodologists and IT experts.
  - Subject-matter statisticians are trained to write the edit checks and rules themselves in the form of SAS syntax.
  - Surveys are less dependent on an individual subject matter statistician, methodologist or IT expert.
- Elimination of errors in consistency between entered rules and variables.
- Microdata databases must be designed according to standard rules.

Future challenges

- All modules of MetaSOP introduced into production (aggregation and standard error estimation, tabulation and tabular protection, quality indicators)
- Procedures for selective editing
- Implementation of data editing for all surveys at SURS with MetaSOP Application