



REPUBLIC OF SLOVENIA
STATISTICAL OFFICE

SURS

We count. Today for tomorrow.

Standard error estimation – how to do it quickly, efficiently and correctly

Session number: 25

Date: 3.6.2016

Authors: Rudi Seljak, Jerneja Pikelj, Petra Blažič

Affiliations: Statistical Office of the Republic of Slovenia, Ljubljana,
Slovenia

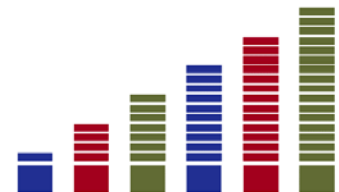
Email: rudi.seljak@gov.si, jerneja.pikelj@gov.si, petra.blazic@gov.si



Madrid, May 31 - June 3

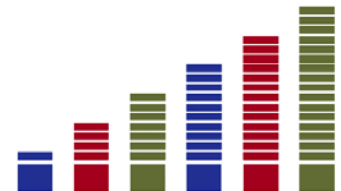
Contents

- ❖ Introduction
- ❖ General solutions – main characteristics
- ❖ Basic architecture
- ❖ Standard error estimation
- ❖ Challenges associated with defining process metadata
- ❖ Conclusions



Introduction

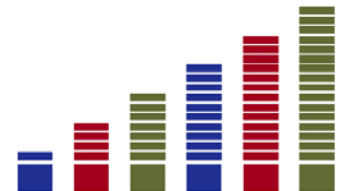
- ❖ The aim of data processing in official statistics is the production of quality statistical data
- ❖ Rationalization of statistical processes
- ❖ The need for transition:
 - From custom made solutions for surveys (stove pipe approach) to generalised process solutions
 - From domain oriented to process oriented production



General solution – main characteristics

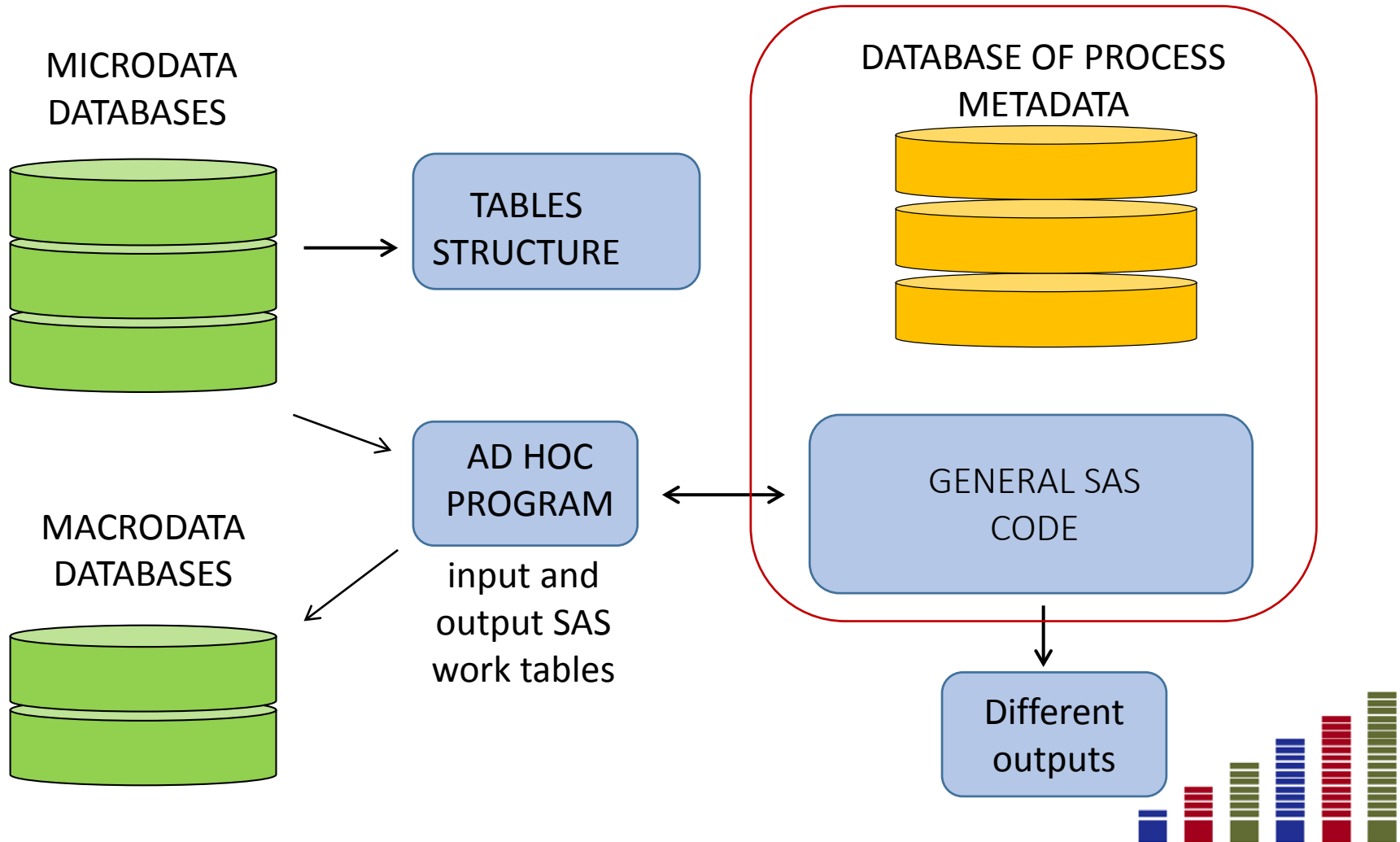
General tool is based on small generic solutions – building blocks

- ❖ General SAS macros for e.g. data validation, imputations, systematic corrections, **aggregation and standard error estimation**, tabulation, quality indicators
- ❖ Can be plugged to different datatables in different environments (e.g. ORACLE, SAS)
- ❖ Designed as metadata driven systems



Basic architecture

Graphical interfaces for management of process metadata



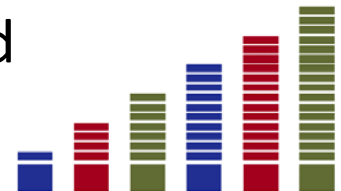
Standard error estimation

❖ Past:

- survey dependent standard error estimation
- direct estimators of standard errors for the key statistics and key domains
- simple linear models for the other statistics and (sub) domains

❖ Present:

- general rules for standard error estimation for different types of estimators and different types of sampling designs
- new rules for dissemination and presentation of the standard errors
- general application provides quick, efficient and unified standard error estimation system



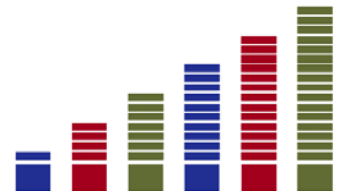
Structure of the metadata

❖ Statistics:

- `stat_label` – label of statistics
- `variable` – name of the variable, needed for the calculation of the statistics
- `type` – type of the statistics

❖ Domains:

- `domain_label` – label of the domain
- `dom_var1` , `dom_var2`, ... - list of the variables, which define the dimensions of the domain



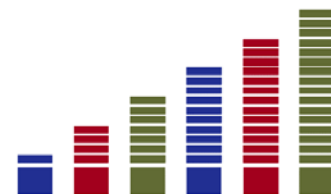
Process metadata – example (1)

❖ Community Innovation Survey (CIS)

❖ We want to estimate the number of innovative enterprises by size classes

- $INOV = \begin{cases} 1; & \text{if the enterprise is innovative} \\ 0; & \text{otherwise} \end{cases}$

- $SIZE_CLASS = \begin{cases} 1; & \text{if the number of employees is greater than 250} \\ 2; & \text{if the number of employees is between 50 and 250} \\ 3; & \text{if the number of employees is less than 50} \end{cases}$



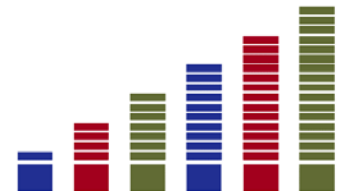
Process metadata – example (2)

❖ Statistics

Stat_label	Variable	Type
STAT1	INOV	01
STAT2	ONE	01

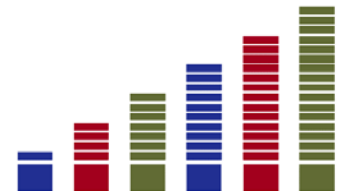
❖ Domains

Domain_label	Dom_var1	Dom_var2
DOM1	SIZE_CLASS	
DOM2	SIZE_CLASS	INOV



Conclusions

- ❖ Importance of the definition of the process metadata – different definition could give us different results
- ❖ Challenges associated with the form of the microdata input table
- ❖ Implementation of statistical processing for all surveys at SURS with SOP
- ❖ Future challenges:
 - Macro editing
 - Aggregation and tabulation of indices
 - ...



Thank you for your attention!

