Topic (v):
International Collaboration and Processing Tools

Discussants:
Pedro Revilla (Spain)
Claude Poirier (Canada)
Gergely Horvath (Hungary)

UNECE - CONFERENCE OF EUROPEAN STATISTICIANS

Work Session on Statistical Data Editing
(Paris, France, 28-30 April 2014)
The development of stronger international collaboration to improve the efficiency of statistical production can help national statistical offices when facing resource constraints.

Organizations are beginning to share tools. Interchangeable modules, interoperable “plug & play” tools, and open source approach acquire the utmost importance.

A working group has been created by the HLG for the Modernization of Statistical Production and Services.
Topics of interest

• Collaboration initiatives targeting software development
• Governing data editing open source tools
• Competing goals of local programs against international initiatives
• Modernization using Enterprise Architecture
• Managing strengths, weaknesses, opportunities and threats
• International collaboration to solve practical problems and to share knowledge and solutions
• Moving national successes into an international perspective
• Editing and Imputation in the Memobust Handbook on Methodology of Modern Business Statistics [NETHERLANDS]
• Metadata Driven General Application for Data Processing From Local Toward Global Solution [SLOVENIA]
• On Implementing CSPA Specifications for Editing and Imputation Services [ITALY]
• Toward Generic Analyses of Data Validation Functions [NETHERLANDS]
• New Feature of VIM - Visualization and Imputation of Missing Values [AUSTRIA]
• SAS Enterprise Guide Project for Editing and Imputation [FINLAND]
• Migration of a Large Survey onto a Microeconomic Platform [NEW ZEALAND]
Summary

[Netherlands] E&I in the Memobust Handbook

- Methodology of Modern Business Statistics
- To be used as guidelines
- Modern, accessible, detailed, refereed
- Compliant with GSBPM, GSIM, EDIMBUS, SDE glossary
- Must stay up-to-date

Q1 – Updates: How to allow wide contributions to Memobust?
Q2 – Relevance: How to promote its use?
[Italy] On Implementing CSPA Specifications for E&I Services

- Sharing methods and tools involved serious works
- CSPA now helps collaboration and sharing services
- Italy successfully wrapped the R package "editrules"
- Compatibility of CSPA with local platforms
- We support the next step: wrapping "rspa"

Q3 – Are other NSOs ready to implement CSPA specifications?
[Netherlands]  Toward Generic Analyses of Data Validation Functions

• GSBPM classifies activities / GSIM defines objects
• CSPA defines services / DDI defines storage standards
• Now VALS offers a generic representation of editing rules
• It offers an implicit Measure of impact of edit failure
• The generic representation offers new ways to monitor edits

Q4 – Should VALS representation drive software development?
Summary (cont’d)

[Slovenia] Metadata driven application for data processing

• Sound methodology
• Building blocks: independent and generalized programs (SAS macros?)
• Process metadata bases (Oracle) data processing programs (SAS macros) GUI to manage (.NET)
• Changes in the organization (NSI):
  - IT, Method: support team
  - Subj. Matter stat.: independent, but needs new knowledge
  - From domain oriented to process oriented

Q5 – New skills of statisticians: programming, databases?
Q6 – How to select the right tool: .NET? SAS – GUI, scripting?
[Finland] SAS Enterprise Guide project for E&I

- SAS is a good statistical software; SAS EG is a great tool to develop complicated processes or ad-hoc, quick analysis
- SAS EG as the workbench for statistics
- Including editing programs: SELEKT and BANFF
- Ingenious use of "Flowchart" in a step-by-step method
- Comprehensive documentation

Two SAS-based applications: two different approaches

- Using an existing application – develop a new GUI

Q6 – How to select the right tool: Custom GUI (.NET?) with SAS or SAS EG?
Q7 – To invest in SAS? Or will it decline (in the near future)?
Summary (cont’d)

[Austria] New Feature of VIM

• Useful tool for missing data analysis and imputation
• Can use survey objects
• New imputation methods
  - KNN, iterative robust regression, hot-deck
• Easy learning and reproducibility: scripts to export
• VIMGUI: new user interface

Q8 – Place of R in official statistics: experimental or production?
Summary (cont’d)

[New Zealand] Migration of a Large Survey onto a Micro-economic platform

• Experiences in the implementation of a software tool
• General processing tool
• Use of Banff and SEVANI
• Ability to handle a complex data collection (AES)
• Editing in the AES
  • Banff procedures (Errorloc and Deterministic)
• Estimation:
  • SEVANI

Q9 – How was the support organised during the introduction of the system – both for the staff (training) and IT (managing technical problems, bug reports, etc.)
Questions / Discussion

International Collaboration

Q1 – How to allow wide contributions to Memobust?
Q2 – How to promote the use of Memobust?
Q3 – Are other NSOs ready to implement CSPA specifications?
Q4 – Should VALS representation drive software development?
Questions / Discussion

Processing tools

Q5 – “New” skills of statisticians: programming, databases?

Q6 – How to select the right tool: Custom GUI (e.g.: .NET?) with SAS or SAS EG?

Q7 – To invest in SAS? Or will it decline (in the near future)?

Q8 – Place of R in official statistics: experimental or production?

Q9 – Self-service processes in official statistics (something similar as the “self service BI”)

Q10 – Training the staff for the new tools, and now roles for the IT