Topic (iii): Software tools and international collaboration

Discussants:
Rudi Seljak (Slovenia)
Claude Poirier (Canada)
Topic (iii): Introduction

• Software tools enable modern data editing strategies
• Development of tools is complex, and costly
• Sharing tools becomes the solution, but not yet successful
• GSBPM and GSIM provide the concepts
• CSPA provides the development basis
• Collaboration has no more obstacles. Let focus on …
  - Open source tools; Moving from local to global solutions;
    International initiatives; Using Enterprise Architecture;
    Applications of the CSPA
7 Presentations

• Towards a generic approach to validation: ValiDat
  A pan-European initiative – from local to global solutions [IT, LT, DE, NL]

• ValiDat survey on the validation approaches in the ESS
  A comprehensive overview of validation methodologies [DE]

• Short Term Indicators - Data cleaning with X12 Arima
  Using common tools in a extremely tight timeframe [AS]

• A formal typology of data validation functions
  Defining measurement concepts and validation functions [NL]

• Integrated data entry and validation system (ADEL)
  On using ADEL to edit data from 130 surveys [HU]

• Usage of external software tools - lessons learned
  What to expect when using external tools to process official statistics [SL]

• Editing and Imputation in Household Based Surveys
  A successful use of external tools [BA]
Summary

[IT/LT/DE/NL] ValiDat - A generic approach to validation

- Validation has a high potential with respect to efficiency gains
- An attempt to find a balance between efficiency and quality
- Metrics of validation rules to monitor coverage & consistency
- Common language between countries and statistical domains
- The project includes 3 work packages:
  - Stock taking / Methodology / Language and Infrastructure

Q1 – Objectives are exemplary. How are you building on other initiatives (VOL1-2-3 , EDIMBUS, MEMOBUST, GSDEMs)?

Q2 – VTL sounds complex. Who can develop a VTL interface?
Summary (cont’d)

[Netherlands] ValiDat survey on the approaches in the ESS

• Clarifying Validation against Editing and Imputation
• Questions and results on …
  - validation processes
  - validation rules
  - validation workload

Q3 – Half of NSIs don’t have a centralized validation service. How should a Pan-European centralisation be governed?
Summary (cont’d)

[Austria] Short Term Indicators - Cleaning with X12 Arima

- Short-term statistics – tough timelines demands → very limited time for data editing
- X12-Arima procedures used for data editing (outlier’s detection and adjustment; missing data imputation)
- Fully automated procedure
- R package X12 was written, offering almost all functionalities of the original program

Q4 – How does the procedures deal with the revisions of the historical data?
Summary (cont’d)

[Netherlands] A formal typology of data validation functions

- Describing validation with a high level of abstraction
- Defining data, data set, data validation
- Using four dimensions to classify validation rules \((U\tau uX)\)
- Setting four levels of validation
- Although its abstraction, the typology is easy to follow. ValiDat uses it as a basis to map practices to theory.

Q5 – Should we deal with multiple sources through \(\tau\) or through \(X\)?
[Hungary] Integrated data entry and validation system (ADEL)

- Metadata driven system - integrated with other IT systems
- 130 surveys already uses ADEL system
- The IT tools are organized in alignment with the adjusted GSBPM model
- ADEL system is built modularly - the standard framework application can be expanded if needed
- User interface layer is provided for users

Q6 – How demanded is the system for end users? Any special training needed?
Summary (cont’d)

[Slovenia] Usage of external software tools - lessons learned

• NSIs want modern-harmonised-automated-cheap production
• External tools (Banff, Calmar, τ-Argus, Demetra) are costly
• General trend and perspective based on open source solutions
• SURS tools are ready to be shared (STAGE, SOP)

Q7 – When time comes to share a tool, should we look for open-source features or CSPA standards, … or both?
Summary (cont’d)

[Bosnia and Herzegovina] E&I in Household Based Surveys

• Description of the E&I procedure for the very complex HBS data
• Combination of quantitative checks, deterministic and probabilistic E&I procedures and qualitative checks
• Different modules of CONCORD, software developed by ISTAT was used to implement procedures

Q8 – What is the long-term perspective of usage of software, developed by another institution?
Questions / Discussion

International Collaboration

Are we going in the right direction? Is there any gap?

Q1 – ValiDat objectives are exemplary. How are we building on other initiatives (VOL1-2-3, EDIMBUS, MEMOBUST)? How does it complement the GSDEMs?

Q2 – VTL sounds complex. Who can develop a VTL interface? Can a partnership be considered?

Q3 – Half of NSIs don’t have a centralized validation service. How should a Pan-European service be governed?

Q5 – While using typology based on \((U\tau uX)\), should we deal with multiple sources through \(\tau\) or through \(X\)?

Are we ready to formalize the processing of multiple sources?
Questions / Discussion

Software tools

Q4 – How does the ARIMA based procedures deal with the revisions of the historical data

Q6 – How demanded is the ADEL system for end users? Any special training needed?

Q7 – Should we look for open-source features or CSPA standards?

Q8 – What is the long-term perspective of usage of software (CONCORD), developed by another institution?