

**UNITED NATIONS ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS**

Workshop on Statistical Data Editing

18-20 September 2018

Neuchatel, Switzerland

**Topic (v): Standardization of editing and imputation methods
and systems (including practical applications)**

Topic organizers: Pedro Revilla Novella, INE, Spain and Jeroen Pannekoek, CBS,
Netherlands.

Topic (iv): Standardization - Introduction

Topic about the use of standards in concepts, methods and processes for data editing.

Standardization offers several advantages.

- Makes it possible to share knowledge and re-use systems with reduced costs of development and maintenance.
- Makes it easy to apply (already implemented) modern state-of-the-art methods.
- Greatly increases transparency of communication about the editing processes and (resulting) data quality.

Standardization has been a topic in most SDE workshops.

And international projects: ESSnet Validat. GSDEMS

Topic (iv): Standardization - Papers

All papers in this topic show the possibilities of using generic tools that can be applied in different statistical processes.

They also show the importance of metadata in this respect

- To configure generic tools to make them suitable for specific steps in an E&I process (edit & imputation rules & method parameterization).
- To drive the process flow. What is the next step (process metadata)
- To describe to quality of the incoming and outgoing data (quality metadata).

Topic (iv): Standardization - Presentations

1. Use of Commercial-off-the-Shelf (COTS) Software at the National Agricultural Statistics Service - ***Darcy Miller (USA)***
2. A standardized approach to editing: Statistics Finland's metadata-driven editing and imputation service - ***Rok Platinovsek (Finland)***
3. A generic validation report for the ESS - ***Olav ten Bosch (Netherlands)***
4. The evolution of BANFF in the context of modernization - **with software demonstration - *Darren Gray (Canada)***
5. Management of data processing processes using metadata - **with software demonstration - *Marek Panfiłow (Poland)***

Enjoy the presentations!

Topic (v): Standardization of editing and imputation methods and systems (including practical applications)

Discussion

Five papers

1. Use of Commercial-off-the-Shelf (COTS) Software at the National Agricultural Statistics Service - ***Darcy Miller (USA)***
 - From in-house programs to COTS software. Possibilities IVEware, PROC MI, PROC SURVEYIMPUTE
2. A standardized approach to editing: Statistics Finland's metadata-driven editing and imputation service - ***Rok Platinovsek (Finland)***
 - Metadata describing the parameters of E&I steps, process flow and quality indicators.
3. A generic validation report for the ESS - ***Olav ten Bosch (Netherlands)***
 - A generic harmonized validation report that prevents misinterpretations and data ping-pong.

Topic (v): Standardization - Presentations

4. The evolution of BANFF in the context of modernization - [with software demonstration](#) - ***Darren Gray (Canada)***
 - Process flow using standardized Banff procedures. Possibility to integrate external (open source) modules.

5. Management of data processing processes using metadata - [with software demonstration](#) - ***Marek Panfiłow (Poland)***
 - Software to manage the sequence of process steps driven by process metadata.

Questions

[Q1] USA- NASS investigated IVEware (ISR) and PROC MI (FCS) methods. Both are multiple imputation methods. What are the pros and cons of using MI at NSI's? Are other institutes also considering multiple imputation?

[Q2] How is the generic validation report received by Eurostat and other agencies? Should such a standard validation report be a mandatory part of quality guidelines?

[Q3] What approximate percentage of StatCan business survey use Banff?

[Q4] Are StatCan considering introducing selective editing methods within Banff?

Questions (cont.)

[Q5] Several papers discuss metadata-driven E&I processes. Are there possibilities for standardizing the metadata. For instance for specifying edit and imputation rules (VTL?), record status, quality reports, metrics?

[Q6] How to integrate the processes of recontact with the informants with those of editing and imputation?

[Q7] Which are the most useful theoretical paradigms (e.g. Fellegi & Holt) for generalized systems?

[Q8] How to integrate commercial and open source software?