



# Generic Statistical Data Editing Model (GSDEM)

Version 2.0

SDE2020 3.9.2020

Virtual UNECE Workshop on statistical data editing 2020

Daniel Kilchmann, Swiss Federal Statistical Office



# Background (1)

## • GSBPM

- “...describes and defines the set of business processes needed to produce official statistics.”
- “...should [...] be seen more as a matrix, through which there are many possible paths”.

## • GSIM

- “...information object framework supporting all statistical production processes...”.
- Standard way to understand and develop statistical data editing (SDE) processes?

# Background (2) - GSDEM

## ➤ Generic Statistical Data Editing Model (**GSDEM**)

- Version 1.0: Released in October, 2015.
- Update to align with revisions of other HLG models and reflect progress in editing and imputation.
- Great opportunity to prepare a supporting document for project managers, methodologists and IT...
- Latest version, June 2019.

# Authors

1. Version, 11 people	Country	2. Version, 12 people	Country
S. Oinonen, P. Ollila, M. Pyy-Martikainen	Finland	A. Kowarik	Austria
E. Gros	France	D. Gray	Canada
		K. Walsdorfer	Germany
M. Di Zio, U. Guarnera, O. Luzi	Italy	F. Rocci, S. Rosati, G. Simeoni	Italy
L.-C. Zhang	Norway		
J. Pannekoek	Netherlands	O. ten Bosch, J. Pannekoek (advisor)	Netherlands
		D. Kilchmann	Switzerland
		F. Leather	U.K.
T. Kolomiyets, S. Vale	UNECE	I. Choi, T. Gjaltema, S. Studer	UNECE

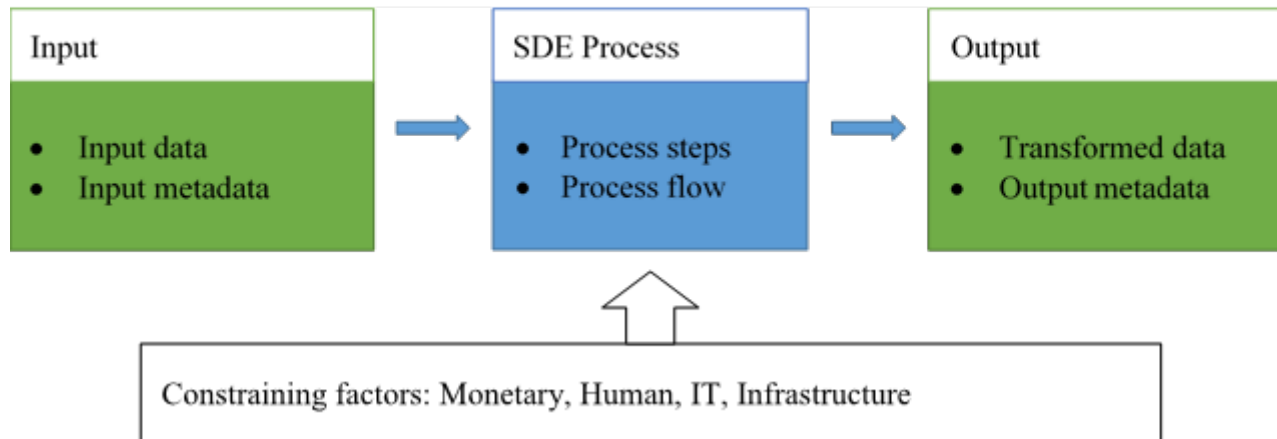
# GSDEM – Chapter description

1. Executive summary.
2. Background and Introduction of fundamental definitions.
3. Description of most common methods and functions using a generic terminology to explain data editing functions (review, selection, treatment).
4. Description of metadata needed to define and describe the data editing functions → fostering automation and better control of the data editing process.
5. Description of the elements defining a SDE process flow representing the sequencing of the different process steps.

# 2. Background: SDE process

- Grouping of data editing activities to form a "fixed segment": one point of entry and one point of exit.
- E.g. GSBPM Sub-processes 5.3 "review and validate" and 5.4 "edit and impute".

Figure 1. Statistical Data Editing Process



# 3. SDE function types

## •Review

- Functions that examine the data to identify potential problems.

## •Selection

- Functions that select units or fields within units for further treatment.

## •Treatment

- Functions that change the data in a way that is considered appropriate to improve data quality.
- The modification of specific fields within a unit (filling in missing values/changing erroneous ones) is referred to as imputation.

# 3. SDE function types examples

- Review

- value plausibility;
- logical consistency;
- plausibility of macro-level estimates.

- Selection

- unit selection for specific treatment;
- selection of influential outlying values for specific treatment;
- selection of variables for treatment by specific imputation methods;
- localising erroneous values among those that are inconsistent.

- Treatment

- imputation of missing or discarded (erroneous) values;
- correction of systematic errors;
- adjustment for inconsistencies.



# 4. SDE metadata types

- **Process input metadata**

- Information objects describing the SDE process input (conceptual/structural: units, variable, value domain, data set,...);
- additional information needed to apply SDE functions: auxiliary data and parameters.

- **Process steps and process flow metadata**

- Information objects describing SDE process itself
  - For each process step: functions and methods;
  - Routing among process steps: process controls.

- **Process output metadata**

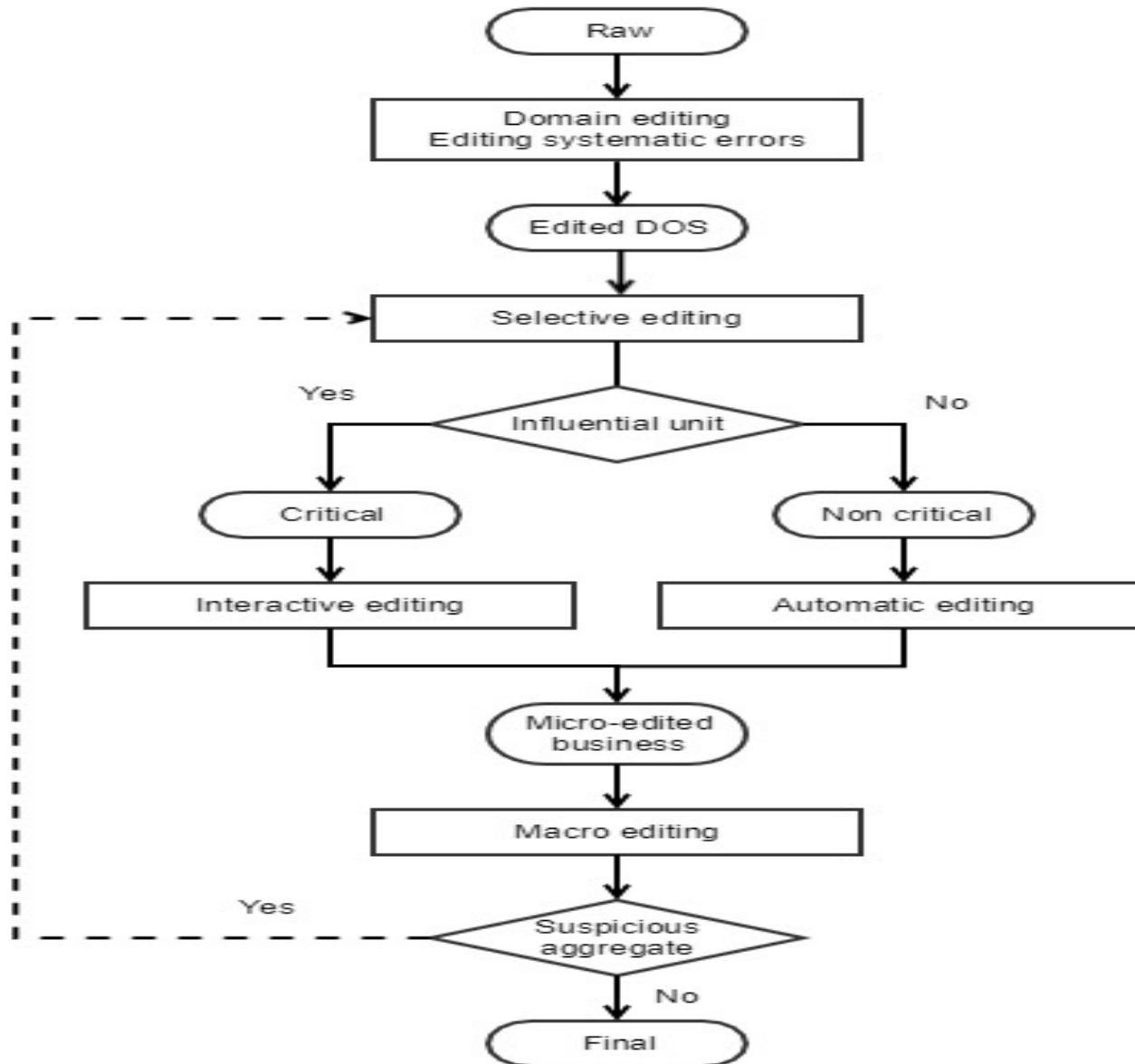
- Information objects describing the SDE process output.
- Other metadata produced by SDE process:
  - Quality information for input and output data;
  - Information how the process has run (paradata).

# 5. SDE process flow models

## Design elements of SDE process flow models

- 1) Structural business statistics
  - cross sectional sample surveys;
  - high number of mostly quantitative variables.
- 2) Short-term business statistics
  - panel surveys with few variables;
  - short production process;
  - Output: indices and variation values, aggregated levels.
- 3) Business census ...
- 4) Household statistics ...
- 5) Statistics through data integration ...

# 5. SDE process flow model - 1) SBS



## 6. Implementation examples (1/3)

- CBS: Intranet article (release of GSDEM 2.0 standard and its use), data cleaning course and R-packages implementing GSDEM based E&I processes applied in various projects.

[www.awesomeofficialstatistics.org](http://www.awesomeofficialstatistics.org)

- INE: Focus on selective editing (conditional mean score/random forest) embedded in modular framework based on GSBPM, GSIM and GSDEM.

Collaboration (methodology-IT-business manager) → stronger/more robust environment to implement innovative initiatives in production.

# 6. Implementation examples (2/3)

- ISTAT:

- Promotion of GSDEM as standard model when E&I processes are redesigned → ongoing continuous implementation.

- UNECE HLG-MOS project “ML for Official Statistics” - design the Editing process of the Statistical Register of economic variable of the public administration.

- UNECE: presentation within the scope of technical cooperation with Armenia.

## 6. Implementation examples (3/3)

- ONS: terminology alignment in training materials of analysts → disseminate of common terminology across the ONS:  
Bookdown document, recorded lectures, a quiz and power-point slides.
- SFSSO: Redesign of the standard processing system aimed to be fully service oriented (ongoing).  
Alignment of training material.
- ...

# GSDEM – Conclusions

- SDE standardization and sharing of tools/knowledge is enhanced through a common framework.
- Support/willingness from management is needed.
- Offices applying GSBPM and GSIM are in a good position to adopt GSDEM.
- Communication: UNECE wiki, Modernstats WS 2019, SDE WS 2020, ...
- Promote, use and enjoy GSDEM!

<https://statswiki.unece.org/display/sde/GSDEM>

Thank you for your attention!