



Generic Statistical Data Editing Model (GSDEM)

Version 2.0, June 2019.

MWW 27.6.2019

Daniel Kilchmann
(Swiss Federal Statistical Office)



Background

- Need for common reference framework for the staff working in statistical data editing (SDE) area to better share ideas, processes, and services.
- 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... revision → 7 months.
- Latest version, June 2019.

Authors

1. Version, 11 people	Country	2. Version, 12 people	Country
Saara Oinonen, Pauli Ollila, Marjo Pyy-Martikainen	Finland	Alexander Kowarik	Austria
Emmanuel Gros	France	Darren Gray	Canada
Marco Di Zio, Ugo Guarnera, Orietta Luzi	Italy	Katrin Walsdorfer	Germany
Li-Chun Zhang	Norway	Fabiana Rocci, Simona Rosati, Giorgia Simeoni	Italy
Jeroen Pannekoek	Netherlands	Olav ten Bosch, Jeroen Pannekoek (advisor)	Netherlands
Tetyana Kolomiyets, Steven Vale	UNECE	Daniel Kilchmann	Switzerland
		Fern Leather	U.K.
		InKyung Choi, Taeke Gjaltema, Salome Studer	UNECE

GSDM – Chapter description

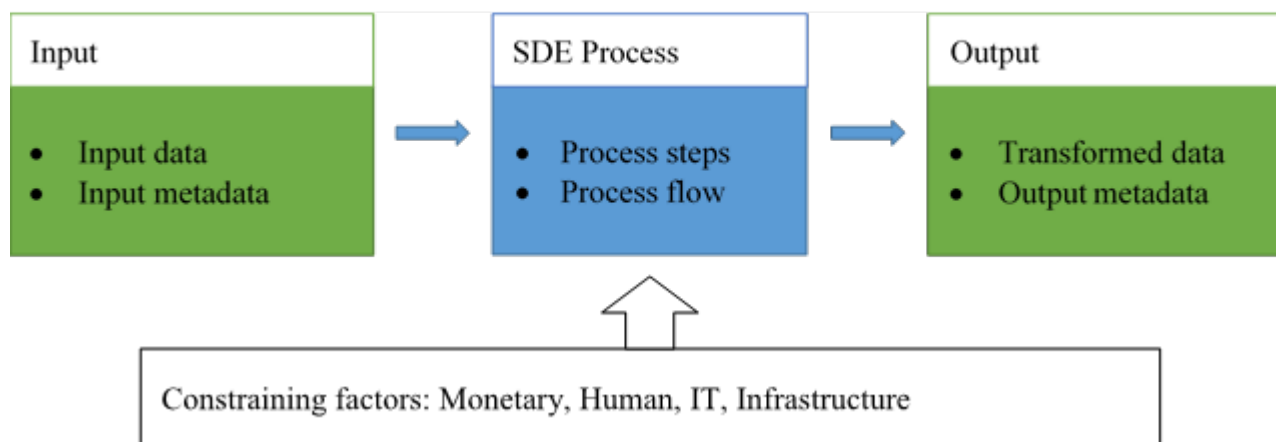
1. Executive summary
2. Background and Introduction of fundamental definitions.
3. Description of most common methods and functions using a generic terminology to explain data editing functions (review, selection, treatment).
4. Description of metadata needed to define and describe the data editing functions → 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.

General proposals of most common flow models with the aim to help users to develop, assess and understand SDE processes.

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 specified 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 (i.e. filling in missing values or 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 metadata: units, variable, value domain, data set, record...).
- 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**

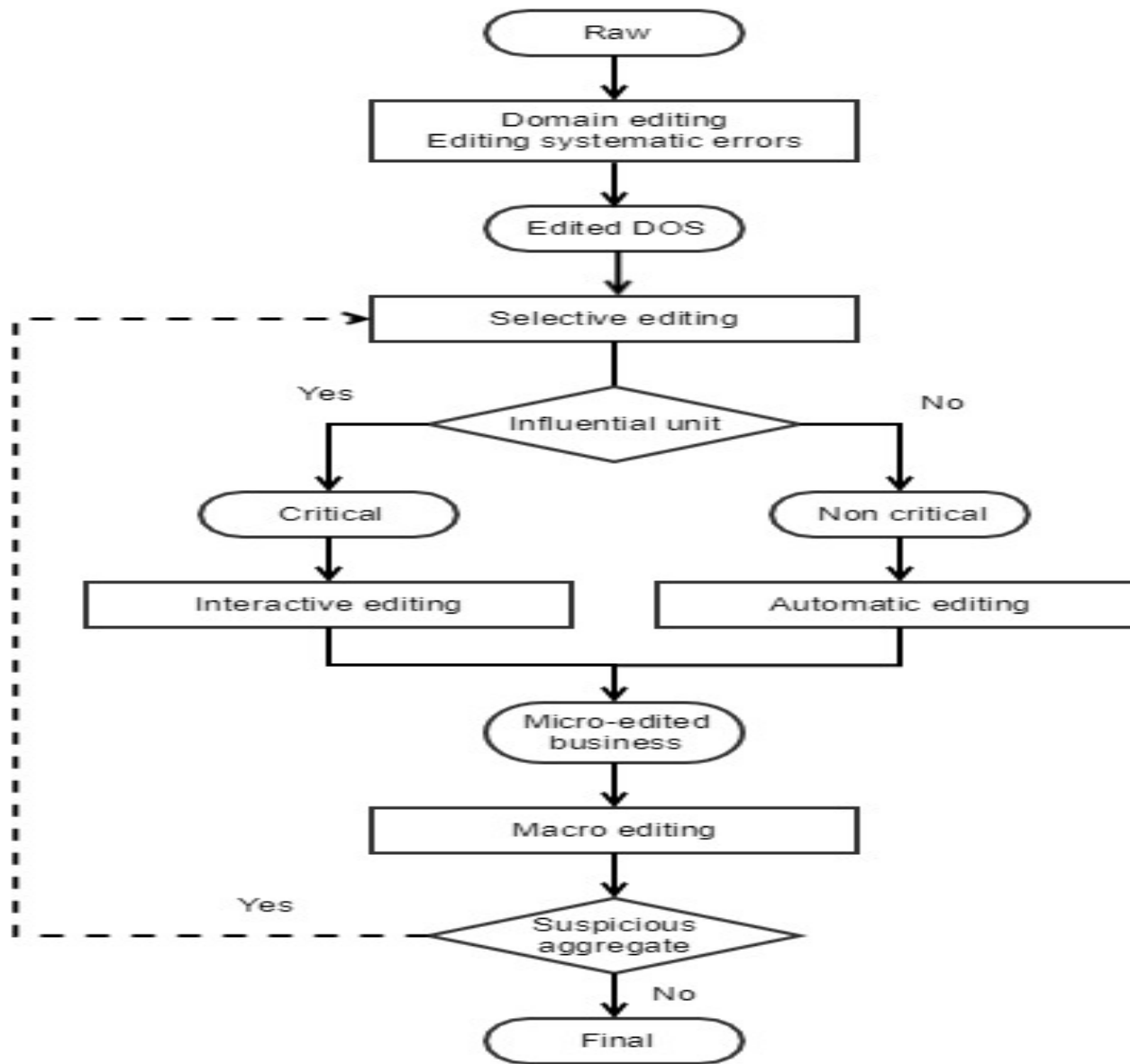
- Information objects describing the SDE process output.
- Other metadata produced by SDE process:
 - Quality information for input and output data.
 - Information how the process has run (not directly related to data quality → paradata).

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 at aggregated levels.
- 3) Business census ...
- 4) Household statistics ...
- 5) Statistics through data integration ...

5. SDE process flow model 1)



GSDEM – Conclusions

- SDE standardization and sharing of tools and of knowledge is enhanced through a common framework.
- Support from management and willingness of the staff is needed.
- Offices applying GSBPM and GSIM are in a good position to adopt GSDEM.
- Communication has to be launched: UNECE wiki, SDE Workshop 2020, etc.
- Promote, use and enjoy GSDEM!
- <https://statswiki.unece.org/display/sde/GSDEM>