

Generic Statistical Data Editing Models (GSDEMs)

Jeroen Pannekoek, Statistics Netherlands

Workshop on the Modernisation of Official Statistics
The Hague, 24 November 2015

Aim of GSDEMs

Data editing attracts considerable interest in the context of modernising official statistics for several reasons:

- It is one of the most expensive and time-consuming parts of the statistical production process.
- It uses advanced methodology and algorithms with many opportunities for re-use and sharing of methods and tools.
- The use of multiple and new data sources poses new challenges for data editing.

To exchange ideas, experiences and practices for modernisation in data editing, there is a need for common concepts, models and definitions.

The set of GSDEMs are envisaged as standard references for statistical data editing, that build on and are consistent with existing standards (GSBPM and GSIM). They aim to facilitate understanding, communication, practice and development.

The Task Team

- The idea for a Generic Process Framework for Statistical Data Editing was first proposed at the 2014 UNECE Work Session on Data Editing.
- GSDEMs have been developed by an international task team (Finland, France, Italy, Norway, Netherlands, UNECE)
- Under authority of the *High Level Group for the Modernisation of Official Statistics* (HLG-MOS).
- Resulting in a report describing:
 - Data editing business functions (*what*)
 - Methods to perform these functions (*how*)
 - Organisation of the process (*process flow models*)

Data editing functions

Review : *Examining data to identify potential problems.*

Input: *rules and data* → **Output:** *Quality indicators and measures*

- Detection of obvious error
- Assessing the logical consistency of combinations of values
- Measuring the (im)plausibility of values or combinations thereof
- Review and identification of suspicious aggregates
- ...

Selection: *Selection of values or units for specified further treatment.*

Input: *quality measures, data* → **Output:** *selected records or fields*

- Selection of variables for treatment by specific imputation methods
- Selection of influential units values for manual review
- Localizing the variables affected by errors for each unit
- ...

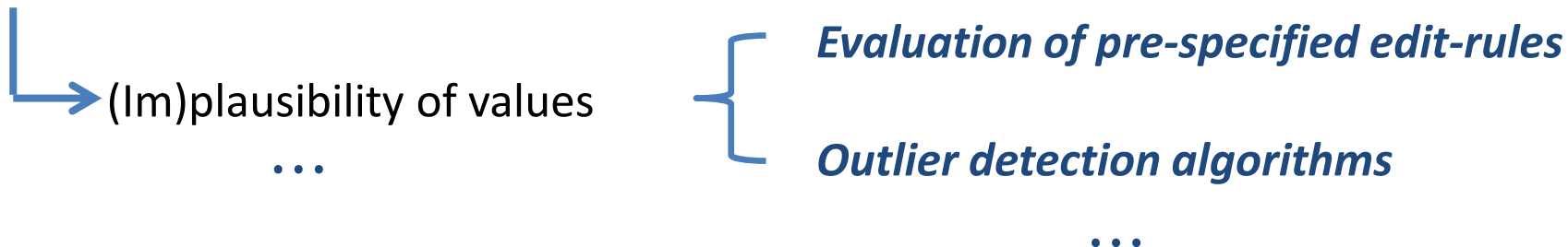
Amendment: *Changing / replacing data values to improve data quality.*

Input: *selected records or fields* → **Output:** *modified data values.*

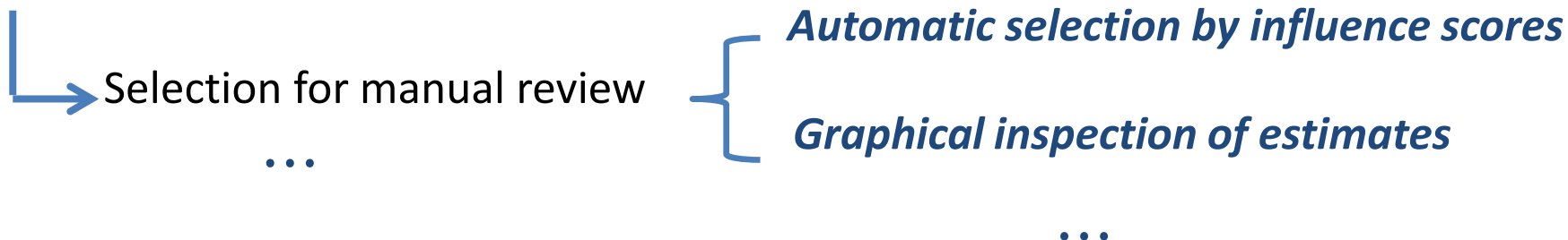
- Correction of systematic errors
- Correction of errors in unit properties
- Imputation of missing or discarded (erroneous) values
- Adjustment for inconsistency
- ...

Methods for data editing functions

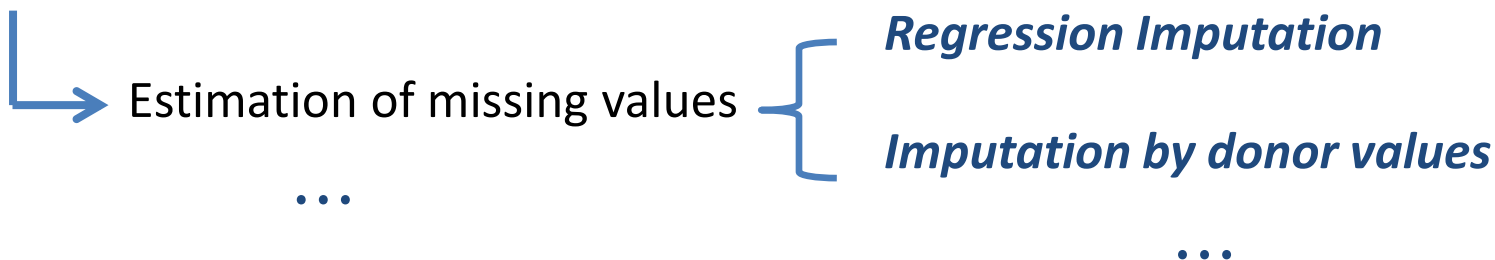
Review



Selection



Amendment



Process flow: Organisation of the data editing process

A data editing process consists of a considerable number of functions with specified methods that are executed in an organised way.

The organisation is described by subdivided it into:

- **Process steps** : Sub-processes each consisting of a number of specified functions. Amendment and/or Review but no Selection.

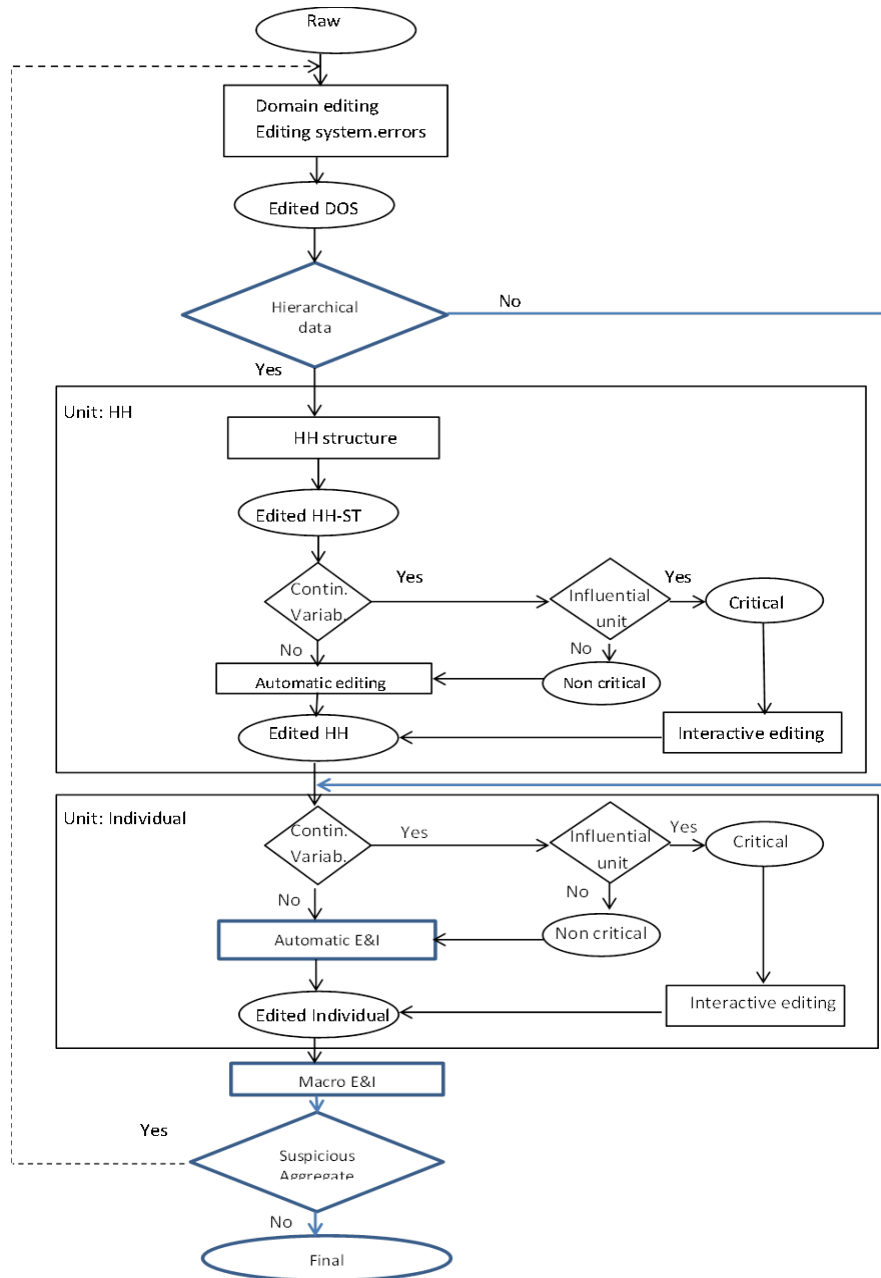
To describe the routing of the data through the process steps:

- **Process controls** : Selection but no amendment functions. Do not modify data values but specify different streams of data through the process flow.

The report contains process-flows for a number of scenario's:

Structural Business Statistics, Short-term statistics, Business census, Household statistics, Statistics based on multiple sources.

Example: SDE flow model for Households statistics



Example: SDE flow model for statistics using data integration

