National guidelines on data editing; the foundation for building a solution for the future

Aslaug Hurlen Foss and Ane Seierstad (Statistics Norway, Norway)

aslaug.hurlen.foss@ssb.no

Abstract

The goal of writing national guidelines on data editing was to establish a common understanding of the data editing process and to promote best practice. The guidelines are based on the general model of the data editing process described in GSDEM. In the national guidelines we have added a process before and after those included in the GSDEM. These are: Control of data before it is submitted to Statistics Norway and control of data before transferring them into the Norwegian StatBank. For each process, quality indicators are suggested. In addition to the guidelines, we have drawn up general principles for data editing. These are based on principles of data validation by Eurostat.

The guidelines are intended to provide a basis for building a data editing solution for the future. The goal is to develop a generic editing code that can cover the editing needs of many statistics. Important functionality is the ability to drill down from macro to micro level. It has been decided that it should be possible to manually edit values, although this should be minimised. Manual editing is also often necessary to train models for automatic correction. All changes to data should be logged. The solution should have access to preapproved functions for selective editing, outlier detection and imputation through the Norwegian Method library. Other important functionality is the use of visualization and quality indicators. The solution will be built using both Python and R.