**Quality system in a digitalised and modernised statistical system**

Grete Olsen, Division for Methods, Statistics Norway (SSB), gol@ssb.no

Aslaug Hurlen Foss, Division for Methods, Statistics Norway (SSB), ahf@ssb.no

Ane Seierstad, Division for Methods, Statistics Norway (SSB), sei@ssb.no

**Abstract**

*Statistics Norway has started a program for digitalisation and modernisation of the production process. By 2020 a new system for some statistics will be implemented, and for 2022 the whole production will take part in the new environment*

*Will the modernised production system give us more and better metadata on quality from the production? Do we need new indicators for quality and new systems for quality management? How can we digitalise the production of quality indicators? The dream is to produce quality indicators automatically in the production process and have parts of the quality report ready when publishing the statistics.*

*Today each subject matter unit reports on quality on their statistics. Furthermore, we have a centralised system for following and improving quality in the statistics. Based on the European Statistics Code of Practice (CoP) we have quality reviews on several statistics each year. Statistics Norway do also report to the Ministry of Finance on timeliness, punctuality, response burden and response rates.*

*Statistics Norway try to combine administrative and statistical systems, such as risk management, internal control system, Lean thinking, quality reviews and other quality reports, to find the best way to measure quality and contribute to continuous improvements in a consistent and efficient way.*

*Some of the lean projects has established standardised procedures for different processes with checklists step by step. These checklists may be a base for collecting metadata on quality indicators from each step.*

*In the quality reviews some of the major findings are need for better harmonisation of documentation such as “About the statistics” and harmonisation of metadata. When modernising the production system there will be major changes in the production system. This can give us the opportunity to standardise and harmonise production, metadata, documentation and dissemination.*

**Keywords:** Quality system, indicators, reviews. Standardisation and harmonisation

**1. Background**

In Statistics Norway the program for digitalisation and modernisation (MOD) started in 2016 will contribute to more standardised and automatic production. By 2020, a new system is planned to be completed with basic functionality and implemented for some statistics. After that, statistics will be gradually implemented, along with the expansion of more advanced functionality that the statistics needs. In 2018 we are establishing criteria for quality indicators in the production process. MOD is managed by subprograms based on the UNECE Generic Statistical Business Process Model (GSBPM).

The main approach to the work on quality indicators is based on UNECE, Quality Indicators for the Generic Statistical Business Process Model (GSBPM) -For Statistics derived from Surveys and Administrative Data Sources. In addition, several international documents on the subject are studied, the ESS handbook for quality reports, ONS Guidelines for measuring Statistical output Quality, BLUE/ETS, the European Statististics Code of Practice (CoP), Quality Assurance Framevork of the European Statistical System (QAF) and Quality measures and indicators of frames for social statistics (see 5. References).

Improvement of process and analyse depend on quality indicators from collect. We are now discussing the need for quality criteria’s following data from process 4. Collect in GSBPM to process 5. Process. In fall 2018 we will suggest quality indicators for the processes process and analyse.

**Figure 1. Man-hours used in SSB, estimated distribution on main processes in GSBPM. Percent**

Based on internal reviews, we have estimated use of time in the production distributed on main processes in GSBPM, figure 1. As expected we spend the most time on process, (such as editing), and collect. In the coming years we will meet cut in budget, need for more resources in the modernisation program, and employees expecting exciting and challenging days analysing the statistics. If we want to save time and money, it’s easy to see that we should put the effort in efficiency improvement and automatization of process and collect. Standardised and automatized quality indicators in these processes will help us save time and tell us when the result of the process is good enough.

The internal quality reviews show us that there is room for improvement and harmonisation of how we produce data on quality for different reports such as annual report to the government and “About the statistics” on the website. For “About the statistics” it would also be useful to establish routines for updating the information more frequently.

**2. Principle, argument and consequence of the quality indicators in MOD**

The work on quality indicators is prepared by MOD.

**Principle:** Quality in the production process will be measured by quality indicators.

**Argument:** Quality indicators must be established for each main process in GSBPM. Quality indicators will give the producers of statistics a tool for managing and supervising the processes in the production. Quality indicators is also included in the description of the quality in statistical output (CoP), and should be part of the information in “About the statistics” on ssb.no and in reporting on quality to the Norwegian government and Eurostat.

**Consequence:** The workbench in a modernised system must offer quality indicators throughout all parts of the production process and gather the quality indicators at the end of the production process. The workbench must produce quality indicators and present them for the producers. The workbench must also give the producers possibilities to produce indicators fitting their area of statistics.

*2.1. Generic quality indicators in MOD*

In CoP, quality indicators are divided into three main groups: Institutional environment, Statistical processes and Statistical output. At this stage in MOD we will confine ourselves to Statistical processes and Statistical output.

The suggested quality indicators will be generic, that is, they will be applicable to all statistics or statistics areas such as administrative data or sample surveys.

In addition, it must still be possible to make quality indicators for each statistic out of the statistic’s uniqueness. This must often be done using tailor-made code.

The goal is to make the quality indicators simple and easy to use, but the system should be open for creating complex quality indicators in the future.

In order to create quality indicators, the quantity must often be presented at different stages, and the ratios between the "number of favourable"/"number possible" must be made. For the statisticians, it is important to know this number to estimate workload needed for correcting errors. To get a total picture of how big a problem is, it is important to have ratios, often in the form of percentages.

For some quality indicators, it will be useful to add a report of errors following the quality indicator. This will give the statisticians an opportunity to easily correct errors. One example is the indicator for duplicates and the possibility to remove duplicates from the population.

*2.2. Data states and product quality in the statistical output.*

In the plans for information management in MOD, long-term saving of different data states and quality indicators for each data state are suggested.

One purpose of this storage is the ability to measure the effects of processing performed in the various sub processes, such as editing, and to create reports on quality at different states.

The states defined are: source data –output from collect, processed data – output from process, and statistical data – output from analyse.

The statistical population should be a long-term-stored data condition and associated with quality indicators.

**Figure 2. Process quality and data states**

 

*2.3. Process quality*

For process quality, quality indicators must be attached to business functions and process methods. Similar process methods should have similar quality indicators in different statistics, to ensure that quality indicators can be compared, and the current best method is chosen.

*2.4. Metadata and quality indicators*

Quality indicators are also data, which need metadata to be interpreted. It is important that the corresponding metadata is in place when quality indicators are created. For example, when quality indicators are based on status codes, categories for status codes must be documented.

*2.5 Information model and quality indicators*

Quality indicators must be linked both to data and to processes/business functions and be described in the logical Information model/data model in a generic way.

**3. Quality Indicators for source data**

Before decisions on quality indicators for the later processes in the GSBPM, we need to define what indicators should follow the data from collect. The work on quality indicators for process, analyse and disseminate will be carried out in spring 2018. So far these quality indicators have been presented for collect.

*3.1. General information about the dataset*

* Name:
* Description:
* Identity:
* Date Created:
* Source Name:
* Created by:
* Data URL:
* Data state: Source data /Processes data/Statistical data/temporary data

*3.2. Specific quality indicators for source data*

**Readability**: Yes/No

**Transformed into SSB format**: Yes/No

**Complete metadata:** Yes/no. Number of variables with incomplete metadata and report on which variables that are incomplete.

Definition on the reporting unit and statistical unit will often be necessary. E. g., municipalities report who has been social clients through a calendar year. The reporting unit will then be the municipality, while the statistical unit will be the person. Definition of the units must be considered for each statistics and up against quality indicators.

**Overview of data sets**: Number of units, number of variables and total number of values in the data set (distributed by identifier component, measuring component and attribute component/)

**Completeness of data sets:** The number of values missing in total and distributed by variables (distributed by identifier component, measure component, and attribute component)

**Duplicates**: Number of units with equal identifier component. Duplicates not equal in value in variables must be handled by the responsible statistician in process and analyse.

**Controls:** The number of units stopped by a hard control and the number of units reported in soft controls. The number of hits in controls per unit and in total. Report on hits from controls.

**Comments from respondents or register owner:** Number of units where the data comes with comments.

**Nonresponse:** Number of units who has not reported. ("Number of units in data population or sample” -"number of units who has reported”). Another option is to use response rate. Needs:

* Total and distributed on categories important for the statistics.
* Some surveys register cause for nonresponse. Reasons may be such as: business sold, should not have been in the sample, unknown address, etc. Then units can be removed from the population
* Number of units is reported cumulatively by time during the collection period. This is important for deciding on action like reminders and fines.
* Total for important measurement component such as turnover.

Most of the quality indicators from UNECE concerning the process collection have then been included, except quality indicators for population.

**4. Next steps**

Based on results and experience from the work on quality indicators for the process collect, MOD will carry on defining quality indicators for the other processes in the GSBPM. The indicators will be an important base both for building the new system and for standardising and harmonising quality work and the production process. They will also contribute to improving the process of documenting quality in statistics for international reports, in the risk management system in Statistics Norway, in annual report to the government and for the users of statistics in “About the statistics” on the website ssb.no.

**5. References**

**UNECE**: V2.0. Quality Indicators for the Generic Statistical Business Process Model (GSBPM) -For Statistics derived from Surveys and Administrative Data Sources, [on the UNECE Statistics Wikis](https://statswiki.unece.org/display/GSBPM/Quality%2BIndicators%2BHome)

**EUROSTAT**: [The ESS handbook for quality reports 2014 edition](http://ec.europa.eu/eurostat/web/ess/-/the-ess-handbook-for-quality-reports-2014-edition)

**Storbritannia**: [ONS Guidelines for Measuring Statistical Ouput Quality v4.1](https://www.statisticsauthority.gov.uk/wp-content/uploads/2017/01/Guidelines-for-Measuring-Statistical-Outputs-Quality.pdf)

**BLUE-ETS**: List of quality groups and indicators identified for administrative data Sources:  <https://www.blue-ets.istat.it/fileadmin/deliverables/Deliverable4.1.pdf>

European Statistics Code of Practice (CoP). <http://ec.europa.eu/eurostat/documents/64157/4392716/Revised_CoP_Nov_2017.pdf>

Quality Assurance Framework of the **European Statistical System** <http://ec.europa.eu/eurostat/documents/64157/4392716/ESS-QAF-V1-2final.pdf/bbf5970c-1adf-46c8-afc3-58ce177a0646>

Quality measures and indicators of frames for social statistics: <https://ec.europa.eu/eurostat/cros/system/files/wp2-framequality-finalreport.pdf>