This module is part of the

# Memobust Handbook

on Methodology of Modern Business Statistics

26 March 2014

# Theme: Methods and Quality

## Contents

# General section

## 1. Summary

This section describes what the dependencies are between the quality of the *statistical output* and the quality of the *method*. Various quality dimensions of the method are distinguished. Each dimension should be managed by taking the right measures. To attain a structural approach to the theme, the Object-oriented Quality and Risk Management (OQRM) model is used.

## 2. General description

In this section, statistical methods are placed in the context of quality (and risk) management. It will describe which measures or actions are taken in the Memobust project to assure the quality of the method description. In addition, this document assesses which measures should be taken by the users of the handbook in order to manage the quality of the statistical methods.

To attain structure in this document, the OQRM model (Van Nederpelt, 2012) is used. The next section concisely describes the OQRM model. This model is elaborated in the module "General Observations – Quality and Risk Management Models". An important concept of the model is that the quality of the product is dependent on the quality other objects such as methods. An object is everything that can be perceived or conceived. Another concept of the model is, that quality of an object can be decomposed in attributes (also: characteristics or quality dimensions) of that object. For example, an attribute of the object *statistical output* is *accuracy*.

### *2.1 Methods and related objects*

Looking at the object *method*, the following related objects can be distinguished. These objects can be regarded as part of the *method* family and have each their own set of attributes:

- Description of the method
- Values of the parameters of the implemented method
- Software implementation of the method
- Input and output data of the method
- Metadata output of the method

Methods are aimed at processing data in order to produce statistical output. A specific category of methods is the measurement processes for quality indicators and logging. A separate theme module in the handbook is dedicated to logging ("General Observations – Logging").

In the Generic Statistical Business Process Model (GSBPM, 2009) the design of methodology is described at least in sub-process: 2.3 design data collection, 2.4 design frame and sample, and 2.5 design statistical processing. Figure 2 in the module "General Observations – GSBPM: Generic Statistical Business Process Model" shows which sub-processes have high (green), intermediate (light green) or little or no (light yellow) methodological content.

### 2.1.1 Method

The method can be distinguished in the method as such and the method as implemented in a statistical process. Both will be covered by this section. Attributes of a method are:

- Soundness
- Appropriateness
- Applicability, usability and stability (GSIM 0.3, 2012, section 35)
- Feasibility
- Complexity
- Efficiency
- Robustness

**Soundness of a method** can be defined as the extent to which the method(ology) used to compile statistics complies with the relevant international standards (SDMX, 2009).

The CoP (principle 7) states that soundness of methodology "underpins quality statistics". Furthermore, principle seven says that "it requires adequate tools, procedures and expertise". An important criterion is, that methods are accepted by the international community in the statistical domain. This handbook describes, however, also methods that have some drawbacks but that still will be used because better methods are not yet available. The *accuracy of the estimated value* is definitely dependent on the *soundness of the method*.

**Appropriateness of a method** is the degree to which a method can be applied to a specific statistical process. A related attributes is *applicability*. In the template of this handbook this focus area is promoted by the sections: purpose of the method (8), recommended use of the method (9), possible disadvantages of the method (10), logical preconditions (13) and the recommended use of the of the individual variants of the method (11). An inappropriate *method* will affect the *accuracy of the output*.

**Usability of a method** can be defined as the degree to which staff understand the method they use in production. It is required that they understand how they should tune the parameters of the methods not only initially but also in the course of time. This focus area should be managed by the methodologists who redesign a statistical process. A method that is not usable can affect *various quality dimensions of statistical output.*

**Complexity of a method** can be defined as the degree to which capacity is needed to use and implement the method correctly. A method that is too complex can take too much time to customise. A difficult method can be wrongly communicated, more easily be misunderstood, or applied wrongly. If a method should be implemented in custom made software it takes a more capacity to develop that software. So, there is a relationship between the *complexity of the method* and *costs of development* and *correctness of the software* too. Feasibility is related to complexity of the method.

This focus area can be addressed in section possible disadvantages of the method (10). Moreover, it should be managed by the methodologists in the redesign process. Methods that are too complex can affect the *cost* as well as the *timeliness of statistical output*.

**Robustness of a method** is the degree to which a method withstands different input data and produces nevertheless accurate output data. Stability is related to robustness of the method. This focus area can be addressed in section possible disadvantages of the method (10) and should be managed by the methodologists in the redesign process. A method that is not robust can cause inaccurate statistical output.

**Efficiency of a method** is the degree to which a method needs IT-resources or computing power. In case of large datasets, it could take too long to process this dataset using a specific method. The *complexity of the method* will certainly affect the *efficiency of the method*. This focus area could be addressed in the section possible disadvantages of the method (10) and should be managed by the methodologists in the redesign process. A method that is not efficient can affect the *cost* as well as the *timeliness of the statistical output.*

### *2.1.2 Description of the method*

**Completeness, correctness, clarity, unambiguity and consistency of the descriptions of the method.** Relevant attributes of the description of the method are *completeness, correctness, clarity, un-ambiguity and consistency*. *Completeness of the description* is promoted by using a certain structure (template) for the descriptions. Furthermore, all focus areas are assured because the content is written and reviewed by experts. An inappropriate description of the method can cause various problems including problems with the *accuracy and timeliness of statistical output*. It can also lead to a misunderstanding of the actual method used.

### *2.1.3 Values of the parameters*

**Correctness of the values of the parameters**. Part of the implementation of the method are the *values of the parameters*. Relevant focus area of the values of the parameters is the *correctness of the values of the parameters*. This focus area can be defined as the degree to which parameters of the method are sufficiently adjusted or tuned. In the template for the methods this focus area is addressed in section tuning parameters (14) and recommended use of the individual variants of the method (11).

The *correctness of the values of the parameters* can affect the *accuracy of the estimate* and the *costs of the statistical output*. For example, the sample size is a parameter of the sampling method. The larger the size of the sample, the larger the *costs of the statistical output* will be. The *accuracy of the estimate,* on the other hand, will increase.

### *2.1.4 Software implementation of the method*

**Correctness of the software.** Relevant focus area of the software implementation is the *correctness of the software implementation of the* method. This can be defined as the degree to which the method is correctly implemented in the software. This focus area is not covered by the handbook. However, problems with this focus area will definitely affect the *accuracy of the estimate.* An action to assure the *correctness of the software* is, e.g., testing after development of the software and after each change of existing software. Another measure is to deploy competent and specialised software developers. Incorrect software will affect the *accuracy of the statistical output.*

**Performance of the software**. Another focus area is the performance of the software. This is the time the application needs to process input data. The *performance of the software* is among others dependent on the *complexity of the method*. It is, moreover, dependant on other focus areas like *efficiency of the software, performance of the software tool and performance of the IT infrastructure*. Logging indicators (20) can be used to measure the *performance of the software*. Software that performs badly can affect the *costs and timeliness of statistical output.*

### *2.1.5 Input and output data of the method*

**Quality of input and output data** is a specific field of expertise which will not be addressed here. A separate handbook could be compiled about this subject. However, bad quality of data (input, intermediate results) will certainly the affect the *accuracy of estimates.*

### *2.1.6 Metadata output*

**Quality of metadata output**. A statistical process can next to statistical data produce metadata such as logs and quality reports. Some metadata are related to the method used in the process. Other metadata are related to the input data, the process or the output data. In the template of the method, quality indicators for output quality are defined (21). Statisticians should take corrective actions if the agreed quality is not met or at least report this to the supplier of the input data or the user of the output data. The quality of metadata will not be elaborated here.

## *2.2 Conclusion*

The focus area *accuracy, consistency, timeliness and the costs of the statistical output* are to a substantial degree dependent on the *quality of the methods* as well as to the quality of other objects. This handbook contributes to the assurance of the *quality of methods*.

# 3. Design issues

# 4. Available software tools

# 5. Decision tree of methods

# 6. Glossary

For definitions of terms used in this module, please refer to the separate "Glossary" provided as part of the handbook.

# 7. References

Eurostat (2011), *European Statistics Code of Practice*. Adopted by the European Statistical System Committee, 28th September 2011.

Eurostat (2012), *Eurostat's Concepts and Definitions Database*. Website: http://ec.europa.eu/eurostat/ramon/nomenclatures/index.cfm?TargetUrl=LST_NOM_DTL_GLOSSARY&StrNom=CODED2&StrLanguageCode=EN

GSBPM (2009), Generic Statistical Business Process Model. Version 4.0 – April 2009. Joint UNECE/Eurostat/OECD Work Session on Statistical Metadata (METIS).

GSIM (2012), Generic Statistical Information Model. Version 0.3, March 2012.

Van Nederpelt, P. W. M. (2012), *Object-oriented Quality and Risk Management (OQRM). A practical, scalable and generic method to manage quality and risks*. MicroData, Alphen aan den Rijn, The Netherlands. Website www.oqrm.org/English.

## Interconnections with other modules

### 8. Related themes described in other modules

1. General Observations – Quality and Risk Management Models
2. General Observations – Logging
3. General Observations – GSPBM: Generic Statistical Business Process Model

### 9. Methods explicitly referred to in this module

1. Micro-Fusion – Prorating
2. Micro-Fusion – Minimum Adjustment Methods
3. Micro-Fusion – Generalised Ratio Adjustments
4. Macro-Integration – RAS
5. Macro-Integration – Stone's Method
6. Macro-Integration – Denton's Method

### 10. Mathematical techniques explicitly referred to in this module

1.

### 11. GSBPM phases explicitly referred to in this module

1. 2.3 Design data collection
2. 2.4 Design frame and sample
3. 2.5 Design statistical processing

### 12. Tools explicitly referred to in this module

1.

### 13. Process steps explicitly referred to in this module

1.

# Administrative section

## 14. Module code

General Observations-T-Methods and Quality

## 15. Version history

| Version | Date | Description of changes | Author | Institute |
|---|---|---|---|---|
| 0.1 | 02-04-2012 | first draft | Peter van Nederpelt | Statistics Netherlands |
| 0.2 | 10-06-2012 | comment of Greece processed | Peter van Nederpelt | Statistics Netherlands |
| 0.3 | 26-09-2013 | comment of Editorial Board processed (first part) | Peter van Nederpelt | Statistics Netherlands |
| 0.3.1 | 30-09-2013 | preliminary release | | |
| 0.3.2 | 02-10-2013 | comment of Editorial Board processed (second part: Leon Willenborg). | Peter van Nederpelt | Statistics Netherlands |
| 0.4 | 04-10-2013 | text about quality management moved to theme Quality and Risk Management models | Peter van Nederpelt | Statistics Netherlands |
| 1.0 | 26-03-2014 | final version within the Memobust project | | |
| | | | | |

## 16. Template version and print date

| | |
|---|---|
| Template version used | 1.0 p 4 d.d. 22-11-2012 |
| Print date | 21-3-2014 17:21 |