

**UNITED NATIONS
ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS**

**EUROPEAN COMMISSION
STATISTICAL OFFICE OF THE
EUROPEAN UNION (EUROSTAT)**

**ORGANISATION FOR ECONOMIC COOPERATION
AND DEVELOPMENT (OECD)
STATISTICS DIRECTORATE**

Work Session on Statistical Metadata (METIS)

(Geneva, Switzerland, 10-12 March 2010)

Topic (ii) Progress on Part C of the CMF and the Generic Statistical Business Process Model

COOPERATION BASED ON THE GSBPM

Submitted by the Nordic and Netherlands Task Force on Business Architecture¹

ABSTRACT

The general goal of the Nordic / Netherlands Task Force on Business Architecture is to facilitate common development and/or sharing of IT solutions and software tools among participating institutes. In order to do this the group was mandated to 1) compare and translate the statistical business architecture model used by the Netherlands into the process model (based on the GSBPM model) used by some of the Nordic countries and 2) develop a common framework based on the comparison. Based on the previous two tasks the group should also indicate to what extent national Business Architectures agreed with, or differed from, the GSBPM. In particular attention should be paid to possible showstoppers that inhibit closer cooperation on IT solutions. On the other hand, suggestions for good opportunities for further cooperation may also be given. This paper will present the results of the task force including strengths and weaknesses of the GSBPM and the use of the GSBPM.

I. INTRODUCTION

1. The general goal of the task force was to facilitate common development and/or sharing of IT solutions and software tools among participating institutes. The participating institutes were Statistics Denmark, Finland, Norway, Sweden and Netherlands. Each institute had two to three representatives with at least one representative per institute having good working knowledge of business architecture. The ESS Code of Practice (CoP) functioned as the common highest level of business principles in the group.

2. The group started in May 2009 with two-weekly telephone conferences arranged by Statistics Netherlands. Statistics Sweden arranged the Wiki-platform for the group. There was one face-to-face meeting in October in the Netherlands. The work was finalised in November.

¹ Prepared by: Max Booleman (m.booleman@cbs.nl) and Jenny Linnerud (jal@ssb.no)

II. ACTIVITIES

3. As a first understanding of the differences and common ideas between the countries, mappings were made between the national business process models of Finland, Norway, Sweden and the Netherlands and the 2-digit level of GSBPM v4.0. None of the countries adopted fully the GSBPM but Norway and Sweden were closest. The model of Statistics Netherlands is more detailed so also easy to map.

4. A lot of variety could be seen between the different countries, but it was noted that much of the variety could be related to the description level rather than to the content level. For Phase 4 Collect we found the following:

GSBPM Collect	Finland	Netherlands	Norway	Sweden
4.1 Select sample	4.3	4.1+4.2	4.1	4.1+4.2
4.2 Setup collection		4.3+4.4	4.2	4.3
4.3 Run collection		4.5+4.6+4.7	4.3	4.4
4.2+4.3 Set up and run collection	4.1+4.2+4.4	4.3+4.4+4.5+4.6+4.7	4.2+4.3	4.3+4.4
4.4 finalize collection	4.5+4.6	4.8+4.9+4.10+4.11+4.12	4.4	4.5

So for example item 4.1 from GSBPM equals item 4.3 from Finland, 4.1 and 4.2 from Netherlands, 4.1 from Norway and 4.1 and 4.2 from the Swedish model.

5. This mapping exercise was carried out for Collect (4), Process (5), Analyse (6) and Disseminate (7).

6. The group then decided to focus attention on the main statistical production process: Process (5), Analyse (6) and Disseminate (7). In these parts of the 'every-day' process it was foreseen that we could start a co-operation to share and develop software.

7. We used the 2-digit level of GSBPM as a classification to group activities according to their main process. So for each activity one and only one GSBPM 2-digit process was assigned to be its main process. This certainly does not mean these activities could not be performed elsewhere in the process, but it was important for a common understanding to identify the main location of the activity in the overall process.

8. For example, for Process (5) subprocesses 5.1 and 5.2 we had:

GSBPM	Main activities
5.1 Integrate data	
	Create record linkage information
	Join data sources
	Matching
	Prioritising
	Link data sources to establish statistical registers
5.2 Classify and code	
	Classify and code
	Identify

	Recoding: transform a variable using one coding system into a variable with another coding system without losing information
	Classify: create a new variable with the help of other variables losing information
	Coding: create a new variable with closed answers out of open answers

III. TOOLS

9. The next stage was to identify tools that are possible candidates for re-use according to their present user or users. These tools were divided into ‘developed by an NSI’ and ‘developed by others’. For example for 5.3 Review, validate and edit possible candidates for reuse mentioned by the group were: Preselect, SELECT and EDIT from Statistics Sweden, Macroview from Statistics Netherlands, Banff and Canseis from Statistics Canada, ISEE (DynaRev, Pris and Structur) from Statistics Norway. Clearly a more detailed description of each tool would be needed to identify exactly which activities are covered by the tool.

10. At the end promising areas were identified to develop new common software tools. These areas are characterised by a lack of existing tools, by a clear statistical methodology and/or by feeling a need for tools there. These areas were 5.1 Integrate data, 5.2 Classify and code, 5.3 Review, Validate and Edit, 7.1 Update output systems and 7.2 Produce products.

IV. REMARKS ON USING GSBPM

11. To have a common framework to work with was very useful for this task force. No special list of concepts or definitions was needed to communicate within the group because no special terminology was used outside the GSBPM-terminology.

12. In addition, there were two main remarks on the use of GSBPM. The first remark is that not all activities are well located according to a general view on the statistical production process. The second remark on GSBPM is that sometimes the text is formulated in terms of solutions and not in terms of problems, e.g. there may be other ways to estimate population parameters than to calculate weights (5.6) and aggregates (5.7).

13. A final comment regarding the numbering of the METIS model is that we used GSBPM as a classification to group activities. However, if GSBPM is meant as a process model with a logical order of activities, then we note that it may be necessary, in some cases, to begin a ‘later’ subprocess before you can end an ‘earlier’ subprocess. The opposite should not be possible. You cannot end a ‘later’ subprocess before an ‘earlier’ one.

III. CONCLUSIONS

14. There is a common understanding of the statistical production process. This means that at least on the functional level common tools could be developed.

15. In addition to common tools within the GSBPM processes 5 till 7 also other kinds of tools could be developed. We think in this regard of more generic services like metadata tools,

quality tools or archiving tools. An example of a possible metadata tool could be the development of a statistical unit type or a population type server.

16. Any group that has identified a small area to start developing a new tool or who is interested in re-use of tools developed by themselves or other statistical institutes is recommended to make use of the opportunity to cooperate with the CORA-project (ESSnet for Common Reference Architecture). See <http://www.osor.eu/projects/cora>. A more common approach could be possible if more countries are interested.