

German Federal Statistical Office: Use of GSBPM

Case study: German Federal Statistical Office German Federal Statistical Office: Use of GSBPM German Federal Statistical Office: use of GSIM

Contact person*	Dr. Sven Claußen
Job title	Assistant Head of section "Cross-Cutting IT Processes Relating to Metadata and Data Quality"
Email	metadata@destatis.de
Telephone	

Statistical business process model

The German Statistical Verbund has adopted the GSBPM 4.0 in the two top levels (phases and sub-process) and has also migrated to the current version 5.0. The description of the 44 sub-processes has been customized to fit the needs of the German statistical system. The German adoption is called GMAS (Business Process Model Official Statistics). The current version of GMAS is 2.0 (2004/12/17).

Metadata used/created at each phase

The German Statistical Verbund plans to develop a centralized metadata management system which should support all phases of the GSBPM. However, some phases produce more likely unstructured metadata as drafts, specifications and other textual documents, while some phases deal mainly with structured metadata like e.g. variables, code lists, statistical units and rules for generating variables. The metadata management system will focus on structured metadata but will also include semistructured metadata like reference metadata (e.g. quality reports in certain formats). To prepare the specification and implementation of the metadata management system an analysis of the current infrastructure has been made and each phase of the process model was examined in detail. A brief overview of the examination will be provided here:

Phase 1

In the first phase several important identifiers will be created and determined, e.g. identifier for a new statistics or a new statistical production cycle. The identifiers are needed to prepare the production tools for assigning production resources.

Phase 2

In the second phase based on unstructured metadata produced in phase 1 variables, associated code lists, statistical units, classifications, concepts and metadata used for data editing will be created. The survey database will be prepared to keep all the necessary resources to control the statistical production cycle. For this purpose the identifiers created in phase 1 are reused and a survey definition file will be created and prefilled with important information about the new production cycle.

Phase 3

In the third phase the metadata created in the previous phases (especially the survey definition and the variables used mainly for data editing) will be reused to prepare the collection of the data. Newly produced metadata are e.g. paper / online forms, and program code for data editing. The database for quality reports will also be prepared in this phase.

Phase 4

In the fourth phase metadata is needed to guide the collection of the data properly. Especially the survey identifier is reused. The input control system uses and produces metadata about the reporting units and the ingoing data. Quality related metadata are produced by the production tools during the subprocesses as paradata.

Phase 5

In the fifth phase coding tools reuse code lists and classifications produced in phase 2. For data editing variables and program code are reused. The documentation of these activities will result in additional process-related metadata needed for quality issues. The production of new variables and statistical units needs already existing variables and statistical units as well as concepts, code lists and appropriate rules to derive new metadata.

Phase 6

The sixth phase needs a lot of metadata to prepare the dissemination phase. Many metadata as e.g. variables, concepts, and code lists are useful to finalise the outputs. Quality information is used to develop quality reports. Data files are generated for import into the dissemination data bases.

Phase 7

In the seventh phase all metadata is needed which may be published together with the outputs to enhance the understanding of the data. The data is imported into the dissemination data bases using the import files from phase 6. But also variables, classifications, code lists, concepts, data description files, quality reports will be reused.

Phase 8

The eighth phase is not further examined, since this phase isn't as well defined as the other phases in Germany yet.

Metadata relevant to other business processes

In general, all metadata collected along the core process chain is also relevant to other business processes, albeit often on a more condensed level.

The processes that need more detailed metadata are "management of statistics (statistical activities)", "methodology development" and "quality and metadata management". Apart from the core processes, management and support processes also need metadata, although mostly either in a very general form or very detailed according to specific requests. To deal with this issue, the Database for Statistical Activities will be made available to more users with the possibility to link to budget and accounting systems or other resource planning software. Especially the needs for human resources have been determined for every statistical activity based on the GSBPM. For every sub-process quality policy and objectives has been defined (still in progress). Also knowledge management is structured in compliance with the GSBPM, e.g. internal information for the employees of the statistical offices in Germany about the statistics, the underlying methodologies and important resources as the paper and electronic forms for example.

Lessons learned

So far the German Statistical Verbund has mainly good experiences with the adoption of the GSBPM. Some phases and subprocesses has been renamed though to better fit one of the design principles of a process model (every name should consist of an object and a verb). For example, instead of "Disseminate" the German name of the seventh phase is "Disseminate Outputs" and instead of "Collect" we use "Collect Data" to name the fourth phase.

The GSBPM is used by the German Statistical Verbund in several scopes. As mentioned above the GSBPM is used to model the metadata flow alongside the statistical production processes, which is structured in compliance with the GSBPM as well. Especially the assignment of quality information to the appropriate sub-process of the GSBPM helps to determine the corresponding production tools which should produce the quality information.

Links:

Attachments