Using the GSBPM in Practice

Steven Vale
UNECE
steven.vale@un.org
Introduction

- The original aim of the GSBPM was to standardise terminology for discussions on statistical metadata systems and processes
- But now it has many other uses:
  (Quotes from GSBPM 5.1)
“The GSBPM can provide a structure for organising and storing documentation within an organisation, promoting standardisation and the identification of good practices”
Process quality management

“If a benchmarking approach to process quality assessment is to be successful, it is necessary to standardise processes as much as possible. The GSBPM provides a mechanism to facilitate this”
Integrating metadata and quality

❖ “The common framework provided by the GSBPM can help to integrate international work on statistical metadata with that on data quality by providing a common framework and common to describe the statistical business process”
Mapping Statistical Production Processes to GSBPM

- All processes that result in data outputs can be mapped to the GSBPM
Example 1

- Australian project to improve production of prices data
- Identified activities within GSBPM sub-processes
- Some phases “out of scope”
Example 2

- Denmark – quarterly survey on employment in construction
- Established, regular survey
  - Just phases 4 to 7
# METIS phases for: Employment in construction industries (marked with pink)

<table>
<thead>
<tr>
<th>Phase</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Behov</td>
</tr>
<tr>
<td>2</td>
<td>Design</td>
</tr>
<tr>
<td>3</td>
<td>Udvikl</td>
</tr>
<tr>
<td>4</td>
<td>Indsaml</td>
</tr>
<tr>
<td>5</td>
<td>Behandl</td>
</tr>
<tr>
<td>6</td>
<td>Analysér</td>
</tr>
<tr>
<td>7</td>
<td>Formidl</td>
</tr>
<tr>
<td>8</td>
<td>Arkivér</td>
</tr>
<tr>
<td>9</td>
<td>Evaluer</td>
</tr>
</tbody>
</table>

## Kvalitetsstyring / Håndtering af metadata

<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Udvælg stikprøve</td>
</tr>
<tr>
<td>4.2</td>
<td>Forbered dataindsamling</td>
</tr>
<tr>
<td>4.3</td>
<td>Gennemfor dataindsamling</td>
</tr>
<tr>
<td>4.4</td>
<td>Afslut dataindsamling</td>
</tr>
<tr>
<td>5.1</td>
<td>Integre data</td>
</tr>
<tr>
<td>5.2</td>
<td>Kod data</td>
</tr>
<tr>
<td>5.3</td>
<td>Gennemgå, fejlso og ret data</td>
</tr>
<tr>
<td>5.4</td>
<td>Imputer manglende data</td>
</tr>
<tr>
<td>5.5</td>
<td>Aflæd nye stat. enheder og variable</td>
</tr>
<tr>
<td>5.6</td>
<td>Beregn vægte</td>
</tr>
<tr>
<td>5.7</td>
<td>Beregn aggregater</td>
</tr>
<tr>
<td>5.8</td>
<td>Færdiggør aggregerede dataset</td>
</tr>
<tr>
<td>6.1</td>
<td>Forbered statistikprodukt</td>
</tr>
<tr>
<td>6.2</td>
<td>Kvalitetssikr Statistikprodukt</td>
</tr>
<tr>
<td>6.3</td>
<td>Gransk og forklar</td>
</tr>
<tr>
<td>6.4</td>
<td>Applicér statistisk fortrolighed</td>
</tr>
<tr>
<td>6.5</td>
<td>Afslut analyse</td>
</tr>
<tr>
<td>7.1</td>
<td>Opdater data i formidlingsystemer</td>
</tr>
<tr>
<td>7.2</td>
<td>Udarbejd statistikprodukt</td>
</tr>
<tr>
<td>7.3</td>
<td>Håndter udgivelsen</td>
</tr>
<tr>
<td>7.4</td>
<td>Markedsfor statistikprodukt</td>
</tr>
<tr>
<td>7.5</td>
<td>Håndter bruger-support</td>
</tr>
<tr>
<td>7.6</td>
<td>Håndter bruger-support</td>
</tr>
</tbody>
</table>
Example 3

- Mapping register processes to GSBPM
- Register maintenance is a continuous activity, not a single process
- But statistical registers have:
  - Inputs “collected” from different sources
  - A sequence of processing and analysis
  - Outputs – statistics and sampling frames
- Therefore, register maintenance is similar to other statistical production
The diagram outlines the process of data collection, processing, analysis, and dissemination. It is divided into four main sections: Collect, Process, Analyse, and Disseminate.

**Collect**
- 4.1 Select sample
- 4.2 Set up collection
  - Prepare for incoming survey and admin data files
- 4.3 Run collection
  - Receive data collection files
- 4.4 Finalize collection
  - Prepare files for loading to the register

**Process**
- 5.1 Integrate data
  - Load data
  - Perform automatic/clerical matching
- 5.2 Classify & code
  - Perform automatic/manual coding of variables
- 5.3 Review, Validate & edit
  - Perform validation checks
  - Edit "unusual" data according to pre-defined edit rules
- 5.4 Impute
  - Impute for missing units or variables
- 5.5 Derive new variables & statistical units
  - Profiling
- 5.6 Calculate weights
- 5.7 Calculate aggregates
  - Population totals
- 5.8 Finalize data files
  - Maximize register quality prior to creating outputs

**Analyse**
- 6.1 Prepare draft outputs
  - Create "dummy" sampling frames or data outputs
- 6.2 Validate outputs
  - Check quality
  - Compare to expected values and other data sources or time periods
- 6.3 Scrutinize & explain
  - Explain any differences found in 6.2
- 6.4 Apply disclosure control
  - To any outputs made available to those not authorised to see microdata
- 6.5 Finalize outputs
  - Prepare and authorize for release

**Disseminate**
- 7.1 Update output systems
  - Load data into output databases
- 7.2 Produce dissemination products
  - Tables, publications, datasets etc.
- 7.3 Manage release of dissemination products
- 7.4 Promote dissemination products
- 7.5 Manage user support
  - Answer queries from survey statisticians

The dotted lines show the different routes depending on whether the output is a sampling frame or a statistical product.
Why do this?

There are practical benefits:
- Standardisation of terminology
- Standard framework for benchmarking
- Facilitates use of common tools / methods
- Efficiency savings
- Tool for managing process quality
Documentation

Example: Armenia - 2011 Population Census

Phase 5 – Process

5.1. Integrate data.
On the reason that in Armenia the census data are collected from one source, there is no necessity to integrate data collected from different sources. The data integration is performed after the data coding and data entry. As the result of data entry, the text files are created, each of them corresponds to one enumeration area. After the data entry completion, all the files are combined, and the database is created for further cleaning and analysis. The data integration can be performed also before the data entry completion for the processing software testing. For the data entry as well as integration the software CSPro is used. Duration of the data integration is very few seconds. Responsible for this sub-process is the Division of information technology and programming of the Population census department. The sub-process of data integration is also used in other processes of data production.

5.2. Classify and code
The works on classification and coding were launched in parallel with the information acceptance at the Department of census. Each portfolio corresponds to one
Similar approaches are now widely used in many other countries
Belarus: Using GSBPM 5.0 to describe the existing statistical production processes

- Pilot surveys description
- Labour statistics
- Industry Statistics
Results:

- Identification of gaps in the existing processes
- Lack of necessary documentation
- Existence of unsettled processes
Purpose of documentation

❖ Needs to be agreed before work starts!
❖ Examples:
  • Knowledge management
  • Succession planning
  • Standardisation – understanding the starting point
  • Metadata / quality management
  • International reporting
What to document?

❖ For the whole process:
  • Summary of purpose
  • Sources, outputs and users
  • Links to other processes
  • Costs?
What to document?

❖ For GSBPM sub-processes:
  • Purpose
  • Inputs and outputs (use GSIM?)
  • Tools and methods
  • Quality criteria
    • When to move on to next sub-process
  • Costs?
How to begin?

• Read the GSBPM sub process descriptions:

2.3. Design collection

This sub-process determines the most appropriate collection method(s) and instrument(s). The actual activities in this sub-process will vary according to the type of collection instruments required, which can include computer assisted interviewing, paper questionnaires, administrative data interfaces and data integration techniques. This sub-process includes the design of collection instruments, questions and response templates (in conjunction with the variables and statistical classifications designed in sub-process 2.2 (Design variable descriptions)). It also includes the design of any formal agreements relating to data supply, such as memoranda of understanding, and confirmation of the legal basis for the data collection. This sub-process is enabled by tools such as question libraries (to facilitate the reuse of questions and related attributes), questionnaire tools (to enable the quick and easy compilation of questions into formats suitable for cognitive testing) and agreement templates (to help standardise terms and conditions). This sub-process also includes the design of process-specific provider management systems.
How to begin?

• Bring together all colleagues who are involved in the process
• Agree and describe the steps needed to complete the process
• Document them!
• Benefits include increased transparency and identifying areas for improvement
• Discussions often bring new ideas
Resource planning

Indicative Current and Future Effort Distribution

[Bar chart showing effort distribution across stages: Specify Need, Design, Build, Collect, Process, Analyse, Disseminate, Archive. Diagram indicates a comparison between 'As Is' and 'Future' effort with upward and downward trends.]
GSBPM quality indicators

- Developed by task team
  - Canada, Hungary, Italy, Turkey, Eurostat, UNECE

- Generic quality indicators for each GSBPM sub-process
  - Processes based on surveys: 2016
  - Expanded to include administrative data: 2017

- Consistent with existing frameworks:
  - UN NQAF, ESS Code of Practice / QAF
Quality Indicators for the Generic Statistical Business Process Model (GSBPM) - For Statistics derived from Surveys and Administrative Data Sources

(Version 2.0, October 2017)

https://statswiki.unece.org/display/GSBPM/Quality+Indicators
Uses of the Quality Indicators

• To provide a standard framework / common terminology to support a process-oriented approach to Quality Management

• To rationalise quality work within an NSO

• To define a mid-term quality policy
  • Set quality targets for a 3-5 year period
### Example: 4.4 Finalise Collection

<table>
<thead>
<tr>
<th>Quality Dimension</th>
<th>Indicator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost-effectiveness</td>
<td>Discrepancy between planned versus actual collection costs. Percentage of collection activities that met requirements (assessed through analysis of paradata)</td>
</tr>
<tr>
<td>Accuracy and reliability</td>
<td>Outgoing error rates; estimate of non-sampling error</td>
</tr>
<tr>
<td>Accuracy and reliability</td>
<td>The rate of over-coverage: The proportion of units accessible via the frame that do not belong to the target population (are out-of-scope).</td>
</tr>
</tbody>
</table>
Quality and Metadata Management

- Needed at many different levels:
  - Process / sub-process level – GSBPM
  - Organisation level – GAMSO
GSBPM implementation information
## Implementations and case studies

### GSBPM Resources Repository

*Created by Inkyung Choi, last modified on 08 Sep, 2021*

<table>
<thead>
<tr>
<th>Title</th>
<th>Resource type</th>
<th>Date</th>
<th>Relevant GSBPM version</th>
<th>Language</th>
<th>Author name</th>
<th>Author organisation</th>
<th>Category</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mapping data production processes to GSBPM</td>
<td>Presentation</td>
<td>2014-09</td>
<td>5.0</td>
<td>English</td>
<td>Steven Vale</td>
<td>UNECE</td>
<td></td>
<td>GSBPM Workshop, NSO Malta, September 2014</td>
</tr>
<tr>
<td>Communicating the GSBPM – How GSBPM’s other uses can play a role</td>
<td>Paper</td>
<td></td>
<td></td>
<td>English</td>
<td>Joe Treacy</td>
<td>Central Statistics Office, Ireland</td>
<td></td>
<td>USE CASE COMMUNICATION</td>
</tr>
<tr>
<td>Example using Statistics Norway's Business Process Model v1.1</td>
<td>Excel spreadsheet</td>
<td></td>
<td></td>
<td>English</td>
<td></td>
<td>Statistics Norway</td>
<td></td>
<td>USE CASE IMPLEMENTATION</td>
</tr>
<tr>
<td>A five-stage data quality compliance framework</td>
<td>Presentation</td>
<td>2019-06</td>
<td>5.0</td>
<td>English</td>
<td>Eduardo Jallath</td>
<td>INEGI, Mexico</td>
<td></td>
<td>Presented at ModernStats World Workshop 2019, 2019-06</td>
</tr>
</tbody>
</table>
Comments or questions about the GSBPM? Ideas for revision? Experiences to share?

Use this discussion forum to share your thoughts.

You need to be logged in to participate in the forum. To request a username and password contact support.stat@un.org

Search for Topics...  New Topic  84 topics found

Show Advanced Filters

<table>
<thead>
<tr>
<th>Status</th>
<th>Topic</th>
<th>Author</th>
<th># of Replies</th>
<th>Last Activity Date</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Issue #5: Mapping GSBPM Overarching processes to GAMSQO (from Franck Cotton)</td>
<td>Chris Jones</td>
<td>0</td>
<td>Sep 08, 2021</td>
</tr>
<tr>
<td></td>
<td>Draft mapping produced during the last GAMSQO revision</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Other papers on the web

About 64,900 results (0.34 seconds)