

GSBPM Revision

Deniz Özkan

Metadata and Standards Department

Quality Group

TurkStat

On behalf of GSBPM Revision Task Team

Why revise?

Users needs

- Recent experiences from users in implementing GSBPM in their organizations and a number of developments in business landscape where statistical organizations operate are requiring changes in GSBPM.

CES

- This model was endorsed by the CES in 2017 on the understanding that it should be updated every 5 years.
 - GSBPM v.5 was released in 2013
 - Revisions only if really needed

Other Models - GAMS0

- The recent developments of GAMS0 requires some updating to GSBPM in order to ensure alignment between the two reference models.

Purpose

The purpose of this project is to revise GSBPM to ensure the model remains relevant and continue serving as the reference framework for statistical organizations.

GSBPM Revision Task Team

Participants are from different NSOs and international organizations such as Eurostat, ILO and UNECE.

Alice Born, Alistair Hamilton, Danny Delcambre, Deniz Ozkan, Edgardo Greising, InKyung Choi, Janusz Dygaszewicz, Johan Eriksson, Juan Munoz, Jūratė Petrauskienė, Laura Lukšaitė-Balakauskienė, Marina Signore, Marta Nagy-Rothengass, Manuel Cuéllar-Rio, Nilgün Dorsan, Steven Vale and Thérèse Lalor

Revision Process

Process

The revision project phased over three major tasks:

- 1) Compilation of user feedback (July 2017 - October 2017)
- 2) Review of feedback and revision (October 2017 - June 2018)
- 3) Public consultation and revision (July 2018 - November 2018)

Summary of issues reported by users/NSOs

Main topics	# of points
<u>General</u>	37
Fundamental issues	17
Issues about GSBPM/GAMSO relationship and harmonization with other standards	14
Other issues	6
<u>Issues on phases of GSBPM</u>	37
Phase 1: Specify Needs	4
Phase 2: Design	7
Phase 3: Build	2
Phase 4: Collect	7
Phase 5: Process	10
Phase 6: Analyse	2
Phase 8: Evaluate	5
Total number of issues	74

General issues	# of issues
Fundamental issues	17
Change text to reflect admin data/commercial big data/geospatial data	5
GSBPM does not reflect the National Accounts statistical process	3
Clarity of sub-processes: Integrated tasks can be classified based on point of view-prone to subjectivity, not clear which tasks to include in which process, Statistical Methods need abstracting	4
GSBPM is a capability model rather than process model. Change representation of GSBPM to look less linear	2
Level of detail (one or more levels below sub-processes)	1
Quality for each stage	2
GSBPM and GAMS0 and other standards	14
Harmonization with other standards (GSIM, GAMS0, QI, SIMS, ESS EA)	6
GSBPM Overarching processes & GAMS0, interference between statistical production and corporate, refer GAMS0 in GSBPM document	6
Management of business register	1
Corporate/legal aspects	1
Other: peripheral issues, not directly related with the GSBPM model	6

Issues Reported on Phases of GSBPM

Issues regarding phases of GSBPM	# of issues
<p>Phase 1 - Specify needs: Text change in some sub-processes, add text to indicate this phase is considered when a new product is created</p>	4
<p>Phase 2 - Design: Add some missing pieces like communication strategy, architecture to the design phase and some more detail when necessary</p>	7
<p>Phase 3 - Build: Rename this phase as perhaps "Assemble", change text to distinguish sub-processes better</p>	2
<p>Phase 4 - Collect: Reflect activities of external data suppliers that are not within NSOs control, non-survey sources, separate collect and process (integrate data), add more detail to sub-processes to reflect collection of non-survey data, add text to set-up collection and run collection</p>	7

Issues Reported on Phases of GSBPM

Issues regarding phases of GSBPM	# of issues
Phase 5 - Process: Cross cutting processes between collect and process such as 5.1 Integrate data, processing completed by data provider, change text to add more detail and clarification	10
Phase 6 - Analyse: Analyse is too much output-oriented, more detail needed	2
Phase 8 - Evaluate: Evaluation needs to be conducted against a benchmark such as quality indicators, add text to phase description and add more detail to sub processes	5

Basic principles

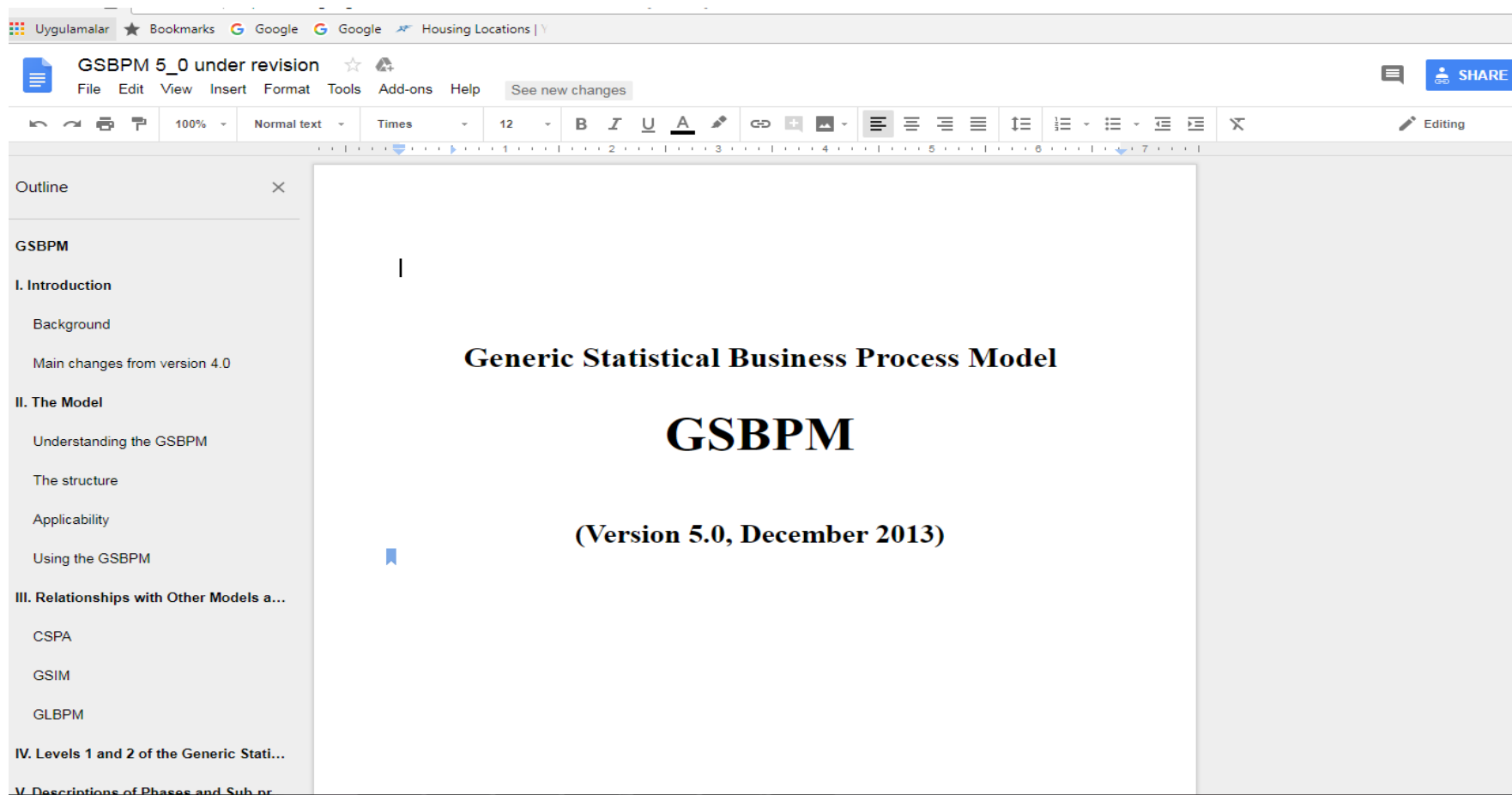
- To the largest extent possible, keep the existing structure unchanged as many organisations have already implemented GSBPM and drastic changes might pose problem to these countries
- So far, the explanatory texts are very much survey-centric; revision needs to encompass other exchange channels and data sources, such as **administrative data, big data and geospatial data**
- Alignment with GSIM revision

Progress

- WG meets every 3 weeks through Webex
- Issues addressed from phase 1. Specify Needs to Phase 4. Collect have been completed
- Currently discussing Phase 5. Process
- Suggestions of the working group are reflected on the shared document
- Agreement on some of the proposals has not been reached by the WG, and will be resolved later

Document under revision

Living document on google doc



The screenshot shows a Google Docs interface. The document title is "GSBPM 5_0 under revision". The main content area displays the following text:

Generic Statistical Business Process Model

GSBPM

(Version 5.0, December 2013)

The left sidebar shows an outline of the document with the following sections:

- GSBPM
 - I. Introduction
 - Background
 - Main changes from version 4.0
 - II. The Model
 - Understanding the GSBPM
 - The structure
 - Applicability
 - Using the GSBPM
 - III. Relationships with Other Models a...
 - CSPA
 - GSIM
 - GLBPM
 - IV. Levels 1 and 2 of the Generic Stati...
 - V. Descriptions of Phases and Sub-pr...

1. Specify Needs

This sub-process involves data from non-survey sources. Therefore the text has been amended to include process for non-survey data sources.

1.5. Check data availability

44. This sub-process checks whether current data sources could meet user requirements, and the conditions under which they would be available, including any restrictions on their use. An assessment of possible alternatives would normally include research into potential administrative or other non-statistical data sources, to determine whether they would be suitable for use for statistical purposes. For administrative and non-statistical data sources, this assessment usually covers several items such as identifying the data source, checking to which extent administrative concepts match data requirement, investigating timeliness and quality of the data source, ascertaining security and continuity of data supply, setting up of responsibilities between data provider and NSO, ICT build resources (data storage, technology required to handle incoming data and data processing) and any formal agreements with data suppliers for accessing and sharing the data (e.g. formats, delivery, accompanying metadata and quality checks,...). If data are not available, a contingency plan should be in place. When existing sources have been assessed, a strategy for filling any remaining gaps in the data requirement is prepared. This sub-process also includes a more general assessment of the legal framework in which data would be collected and used, and may therefore identify proposals for changes to existing legislation or the introduction of a new legal framework.

2. Design

Design collection sub-process explanations were updated to include non-survey data sources and other exchange channels

2.3. Design collection

51. This sub-process determines the most appropriate collection method(s) and exchange channel/instrument(s). Data collection methods may depend on the type of survey (exhaustive, sampling or otherwise), the survey unit (enterprise, person, or other) and the availability of various source data. The actual activities in this sub-process will vary according to the type of collection exchange channel/instrument(s) required, which can include computer assisted interviewing, paper questionnaires, administrative data interfaces and - data transfer methods - and data integration techniques. This sub-process includes the design of collection exchange channel/instrument(s), 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 any formal agreements and 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. Direct or indirect use of administrative data source may be introduced in the data collection mode for either controlling survey data or assisting it when capturing survey information.

3. Build

Not so many changes in this phase. Only some wording changes to achieve consistency across the document. E.g. “collection exchange channel” instead of “collection instrument”.

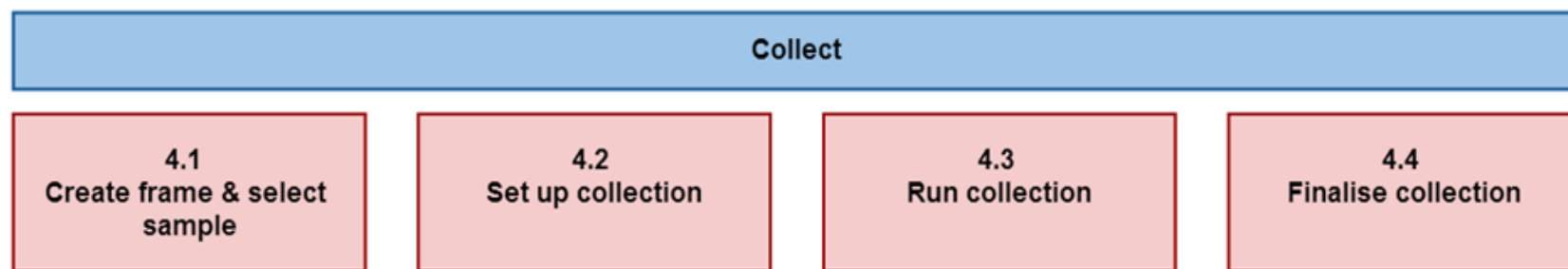
3.1. Build collection exchange channel~~collection instrument~~

58. This sub-process describes the activities to build the collection exchange channel~~collection instruments~~ to be used during the "Collect" phase. The collection exchange channel~~collection instrument~~ is generated or built based on the design specifications created during the "Design" phase. A collection may use one or more modes to receive the data, e.g. personal or telephone interviews; paper, electronic or web questionnaires; SDMX hubs. Collection instruments may also be data extraction routines used to gather data from existing statistical or administrative data sets. This sub-process also includes preparing and testing the contents and functioning of that instrument (e.g. testing the questions in a questionnaire). It is recommended to consider the direct connection of collection exchange channel~~collection instruments~~ to the statistical metadata system, so that metadata can be more easily captured in the collection phase. Connection of metadata and data at the point of capture can save work in later phases. Capturing the metrics of data collection (paradata) is also an important consideration in this sub-process.

4. Collect

Proposal to change the name of the “COLLECT” phase as “ACQUISITION”. Detailed text for collection modes to include acquisition of non-survey data.

Collect Aquisition (Collect) Phase



65. This phase collects or gathers all necessary information (structured and unstructured data, and metadata and paradata), using different collection modes (including acquisition, collection, connection, extraction, transfer, transmission or streaming including extractions from statistical, administrative and other non-statistical registers and databases), and loads them into the appropriate environment for further processing. Whilst it can include validation of data set formats, it does not include any transformations of the data themselves, as these are all done in the "Process" phase. For statistical outputs produced regularly, this phase occurs in each iteration.

4. Collect

Added more detail in sub-processes.

The level of detail was not the same for survey and non-survey data.

68. This sub-process ensures that the people, processes and technology (including web-based applications and apps) are ready to collect data and metadata, in all modes as designed. It takes place over a period of time, as it includes the strategy, planning and training activities in preparation for the specific instance of the statistical business process. Where the process is repeated regularly, some (or all) of these activities may not be explicitly required for each iteration. For one-off and new processes, these activities can be lengthy. For survey data, this sub-process includes:

- preparing a collection strategy;
- training collection staff and apps using supervised machine learning techniques;
- ensuring collection resources are available (e.g. laptops, collection apps, APIs);
- agreeing on terms with any intermediate collection bodies, (e.g. sub-contractors for computer assisted telephone interviewing, web services, APIs, data providers)
- configuring collection systems to request and receive the data;
- ensuring the security of data to be collected;
- preparing collection exchange channels ~~collection instruments~~ (e.g. printing questionnaires, pre-filling them with existing data, loading questionnaires and data onto interviewers' computers, APIs, web-scraping tools, etc.).

69. For non-survey sources, this sub-process will include ensuring that the necessary processes, systems and confidentiality procedures are in place, to receive or extract the necessary information from the source. For non-survey data, this sub-process includes:

- Evaluating requests to acquire non-survey data and logging the request in a centralized inventory
- Initiating contacts with organization providing the data, and sending an introductory package with details on the process of acquiring the data
- Negotiating with the data provider detailed information about files and metadata; and receiving a test file to assess data's fitness for use
- Arranging secure channels for the transmission of data

4. Collect

Proposal to create a new sub-process "4.4 Monitor acquisition" in collect phase, but the WG decided to merge the text in 4.3

4.4 Monitor acquisition

This sub-process refers to the monitoring and remediation of the acquisition process towards optimising the quality of data collection during the acquisition phase. This sub-process usually includes generating reports, visualising and adjusting the acquisition process to support subsequent data quality requirements to ensure the data are fit for use.

4.54. Finalise ~~acquisition collection~~

71. This sub-process includes loading the acquired ~~collected~~ data and metadata into a suitable electronic environment for further processing. It may include manual or automatic data capture ~~take-on~~, for example using clerical staff or optical character recognition tools to extract information from paper questionnaires, or converting the formats of files or encoding the variables received from other organisations. It may also include analysis of the process metadata (paradata) associated with acquisition ~~collection~~ to ensure the collection activities have met requirements. In cases where there is a physical collection exchange channel ~~collection instrument~~, such as a paper questionnaire, which is not needed for further processing, this sub-process manages

Current status of revision

- 22 of 74 issues that have been raised on GSBPM discussion forum have been touched
 - ✓ 11 - resolved
 - ✓ 7 - proposal given
 - ✓ 4 – discussing
- Phases from 1.Specify needs to 4. Collect have been reviewed.
- Started to discuss Phase 5.Process
- Remaining issues will be discussed and the document will be opened for public consultation

Issues and challenges

- Semantics, difference between native English speakers' and other members' perception (e.g. Acquisition versus Collect)
- Achieving clarity about the model without creating much clutter in the document
- Determining boundaries of the sub-processes more precisely
- Making room for all data sources without losing generality (geospatial, admin registers, big data, etc.)
- Ensuring consistency with other standards (GSIM, GAMSO, CSPA)
- The need of reviewing and expanding the definition of the overarching processes for "Quality and Metadata Management"
- Difficulty to reflect some users' specific needs in the generic model

Recommendations

- Make sure to keep in GSBPM what has been a success, i.e. a general framework of the data lifecycle for official statistics
- At the same time, the model needs to be updated to reflect the trends and diversity of sources of data ecosystems
- Keep in mind that this is a generic, non-linear and iterative model. Users can adapt and customise the model depending on their specific needs
- Encouraging all countries to actively participate in the GSBPM consultation when the new GSBPM version is sent out

Thank you for your attention.