



---

**Economic Commission for Europe**

**Conference of European Statisticians**

**Group of Experts on Business Registers**

**Twelfth session**

Paris, 14-15 September 2011

Item 6 of the provisional agenda

**The role of business registers in industrializing the production of statistics**

**Applying the Generic Statistical Business Process Model to  
business register maintenance**

**Note by the secretariat**

*Summary*

The Generic Statistical Business Process Model has been adopted as a standard for describing the statistical production process by over thirty national and international statistical organisations worldwide. The documentation accompanying the model states that “the GSBPM can also be applied to the development and maintenance of statistical registers”. This paper examines this claim from the point of view of statistical business registers.

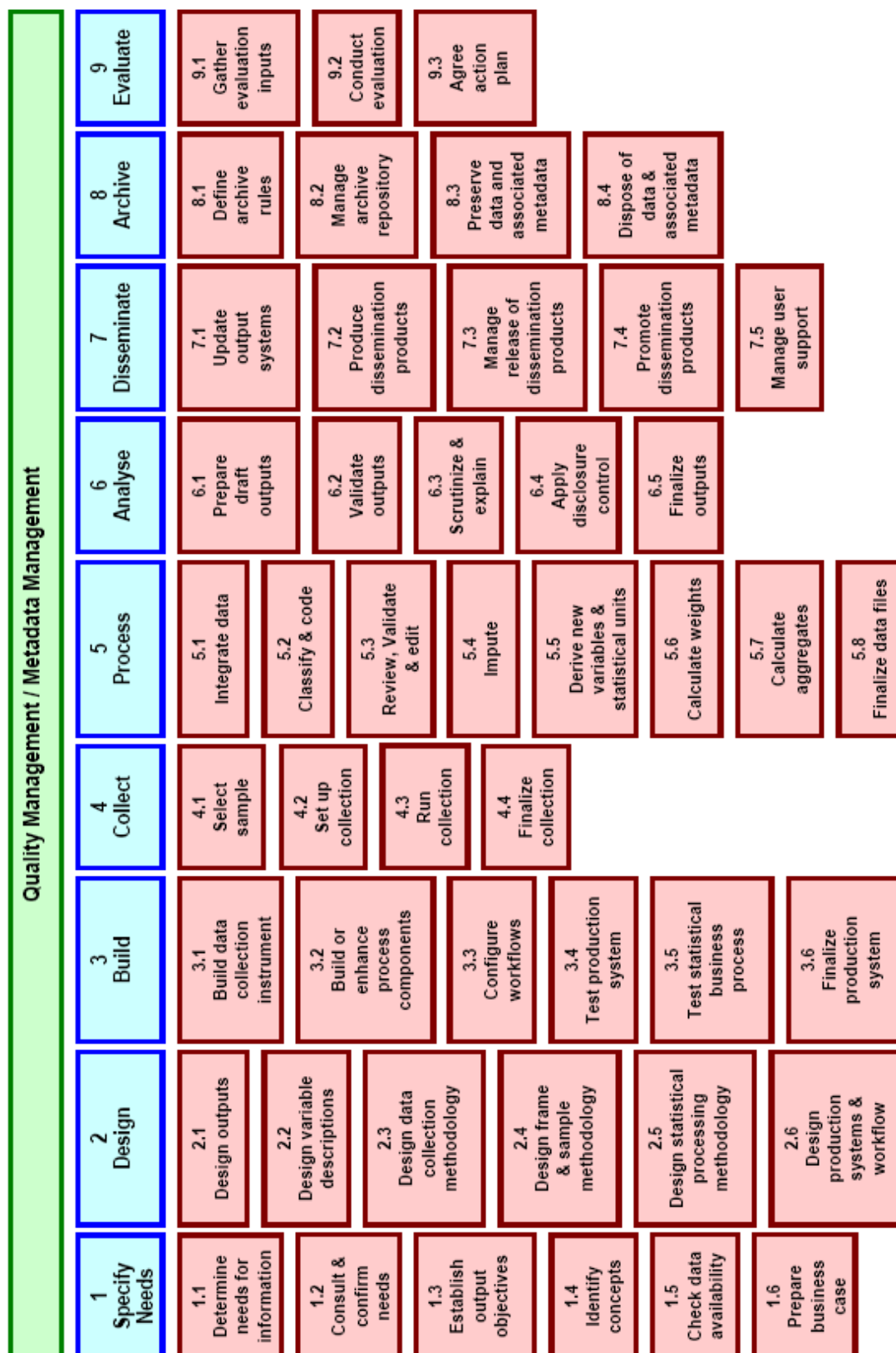
## I. Introducing the GSBPM

1. The "Generic Statistical Business Process Model" (GSBPM)<sup>1</sup> was developed by the UNECE and the Conference of European Statisticians Steering Group on Statistical Metadata (better known as "METIS"). Since its release in April 2009, this model has already been widely adopted by national and international statistical organisations around the world, and is intended to facilitate the convergence of statistical production processes, both within and between organisations.
2. The original aim of the GSBPM was to provide a basis for statistical organisations to agree on standard terminology to aid their discussions on developing statistical metadata systems. It was conceived as a flexible tool to describe and define the set of processes needed to produce official statistics. The GSBPM is, however, increasingly being used in other contexts such as harmonising statistical computing infrastructures, facilitating the sharing of software components, and providing a framework for process quality assessment and improvement.
3. The GSBPM is intended to apply to all activities undertaken by producers of official statistics, at both the national and international levels, which result in data outputs. It is designed to be independent of the data source, so it can be used for the description and quality assessment of processes based on surveys, censuses, administrative records, and other non-statistical or mixed sources.
4. Whilst typical statistical production processes include the collection and treatment of raw data to produce statistical outputs, the GSBPM is also intended to apply to cases where existing data are revised or time-series are re-calculated, either as a result of more or better source data, or a change in methodology. In these cases, the input data are the previously published statistics, which are then processed and analyzed to produce revised outputs.
5. The GSBPM is not intended to be a rigid framework in which all steps must be followed in a strict order, but rather a model that identifies the steps in the statistical production process, and the inter-dependencies between them. It aims to be sufficiently generic to be widely applicable, and to encourage a standard view of statistical production, without becoming either too restrictive or too abstract and theoretical. Different statistical production processes will follow different paths through the model, using different sub-processes, in different orders.
6. The GSBPM comprises four levels:
  - (a) Level 0, the statistical business process;
  - (b) Level 1, the nine phases of the statistical business process;
  - (c) Level 2, the sub-processes within each phase;
  - (d) Level 3, a description of those sub-processes.

---

<sup>1</sup> [www.unece.org/stats/gsbpm](http://www.unece.org/stats/gsbpm)

Figure 1  
Levels 1 and 2 of the GSBPM



7. Levels 1 and 2 are illustrated in Figure 1. Information about level 3 can be found in the GSBPM documentation. Further levels of detail may be appropriate for certain statistical business processes or in certain organisations, but these are unlikely to be sufficiently generic to be included in this model.

8. The GSBPM also recognises several over-arching processes that apply throughout the nine phases, and across statistical business processes. These can be grouped into two categories, those that have a statistical component, and those that are more general, and could apply to any sort of organisation. Examples of over-arching statistical processes include:

- (a) Quality management – including quality assessment and control mechanisms;
- (b) Metadata management – ensuring that metadata retain their links with data throughout the business process;
- (c) Statistical framework management – developing standards, methodologies, concepts and classifications that apply across multiple processes;
- (d) Human resource management;
- (e) Financial management;
- (f) Project management.

## **II. Applying the GSBPM to statistical business registers**

9. As well as being applicable for processes which result in statistics, the GSBPM documentation claims that it can also be applied to the development and maintenance of statistical registers. It states that “... the inputs are similar to those for statistical production (though typically with a greater focus on administrative data), and the outputs are typically frames or other data extractions, which are then used as inputs to other processes”.

10. One key difference between the maintenance of statistical business registers and other statistical processes is that register maintenance is usually a continuous activity. In most countries the statistical business register is constantly updated from a mixture of administrative and statistical sources. The process of updating a register therefore does not generally have as clear start and end points as typical survey-based processes.

11. Having said that, statistical business registers have clear outputs. Typically these take two forms: sampling frames (and survey populations), and statistics such as business demography data. In both cases, the outputs are based on “snapshot” views of the register at a particular point in time, which are extracted for further processing. Therefore the register maintenance process between two instances of the creation of such outputs can be seen as conceptually similar to other statistical production processes. As there can be many different register outputs, there will often be multiple processes occurring at any point in time.

12. The first three phases of the GSBPM, “Specify needs”, “Design” and “Build” can be applied to any statistical process, however, they are clearly not intended to be repeated for every iteration of a regular process. For example, there is no need to include these phases each time a monthly survey is carried out. Normally they only apply to the first iteration of a process, and any subsequent reviews of that process to make sure it remains relevant. It is therefore relatively easy to see how these three phases can be applied when statistical registers are being created or reviewed.

13. The “Collect” phase generally involves identifying the units for which data are required, and running a collection process to obtain those data. This phase can apply equally to data from surveys, administrative records, or mixed sources. In the cases of register maintenance, a source of information is identified (e.g. a file of tax records), and a process is followed to obtain those data in a suitable format for integration with the register. Therefore “Collect” should be seen as a synonym for “obtain data”.

14. Sub-process 4.1 “Select sample” is therefore generally not relevant to business register maintenance (in the same way it is not relevant for a census), because typically all relevant records are selected. Sub-process 4.2 “Set up collection” covers the activities necessary to prepare to receive the incoming data, including configuring systems and processes. Sub-process 4.3 “Run collection” refers to the task of actually obtaining the data, and in the case of register maintenance, it typically a routine, scheduled task. Sub-process 4.4 “Finalise collection” covers any activities to prepare the data for loading into the register, for example conversion of file formats.

15. As for many other statistical production processes, the key phase of the GSBPM for register maintenance is phase 5 “Process”. This is the phase where the data obtained from various sources are integrated, cleaned and converted according to the standards followed for the particular register. Sub-processes 5.1 to 5.5 (“Integrate data”, “Classify and code”, “Review, validate and edit”, “Impute”, and “Derive new variables and statistical units”) are clearly relevant to business register maintenance, as there are typical tasks when processing incoming data files.

16. Sub-process 5.6 “Calculate weights” is rarely relevant to business register maintenance, though there may be certain special cases where a unit may be considered to be representative of others, so is given a certain weight. Sub-process 5.7 “Calculate aggregates” applies to the preparation of both sampling frames (e.g. calculation of population totals), and to creating statistics directly from registers (e.g. business demography data such as enterprise births by economic activity). Sub-process 5.8 “Finalize data files” can be related to ensuring the quality of a register just prior to producing a sampling frame or other output.

17. Phase 6 of the GSBPM; “Analyse”, is where outputs are prepared, validated and made ready for release. For statistical business registers this is where sampling frames and direct statistical outputs are prepared, checked, and quality assured for release. Sub-process 6.4 “Apply disclosure control” will not apply for sampling frames, but is very important for outputs such as business demography data.

18. GSBPM phase 7; “Disseminate”, is also clearly more relevant to cases where statistics are produced directly from business registers. However, elements of this phase, particularly sub-process 7.5 “Manage user support” can also be applied to supporting the “users” of sampling frames. For example it could cover managing queries from survey statisticians about changes to the population of businesses or to individual statistical units.

19. Managing archive “snapshots” of statistical business registers and copies of sample frames is conceptually very similar to managing the archiving of any other statistical data set. Archive rules and policies are required, as well as mechanisms to manage and preserve archived data and metadata. Therefore, GSBPM phase 8 “Archive” can be seen as relevant to the maintenance of statistical business registers.

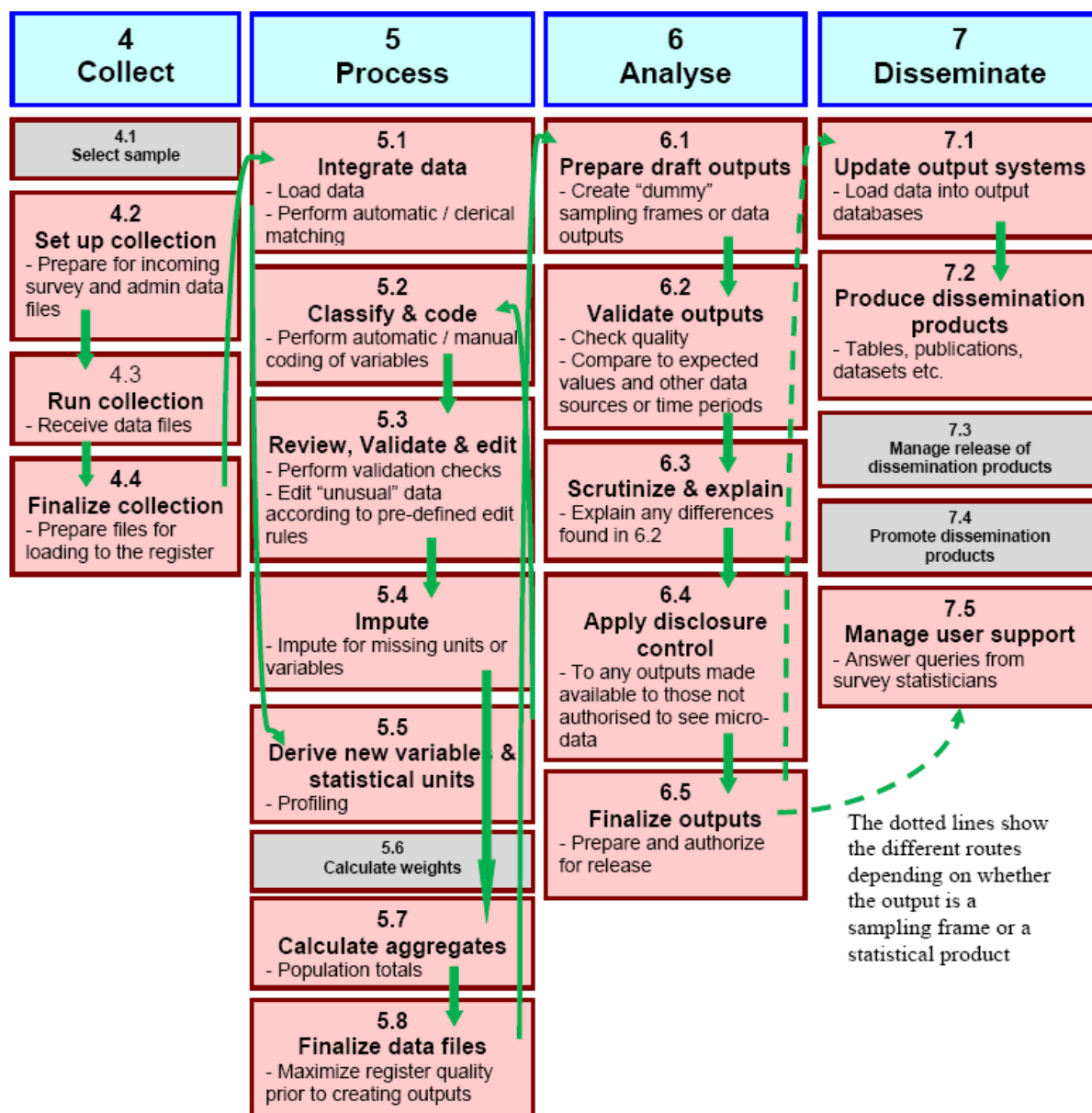
20. The ninth and final phase of the GSBPM is “Evaluate”. Regular evaluation of the quality of the register itself, as well as of outputs, is clearly an important part of register maintenance.

21. Figure 2 attempts to summarise the above paragraphs by showing graphically the processes of maintaining statistical business registers to produce sampling frames, and direct statistical outputs. It focuses on the four central phases of the GSBPM, as the other phases are clearly applicable to all statistical processes. The diagram shows which sub-

processes may be used, and in which order. In practice, the exact mapping of register maintenance activities to the GSBPM will vary between different countries and organisations, due to different organisational and operational requirements, different data sources, and different types of outputs.

Figure 2

Applying the phases 4-7 of the GSBPM to register maintenance activities



### III. Why apply the GSBPM to statistical business registers?

22. The previous section showed that it is possible to map the process of maintaining a statistical business register, but two important questions remain. Why is it interesting to do

this? Is this sort of mapping purely an academic exercise, or does it have practical benefits for statistical business register managers?

23. To answer these questions, it is useful to look at experiences of implementing the GSBPM in other domains, because similar questions were asked in those cases, and some clear benefits are starting to emerge.

24. The first main benefit experienced in applying the GSBPM is the standardisation of terminology between processes and organisations. This, combined with a standard framework for describing processes, considerably facilitates the process of benchmarking – i.e. comparing processes to identify good practices from one that could be transferred to another. It also facilitates the identification of synergies between processes, which might not otherwise be apparent. For example, any process that includes an element of coding text responses may be able to use similar tools and techniques, whether the output is health statistics, population census data, or sampling frames for business surveys.

25. This, in turn, leads to the possibility of efficiency savings, either from adopting common tools, or sharing the costs of developing new tools or methods between several processes or organisations. A practical example of this is an inventory of statistical software, developed by national or international statistical organisations, and available for sharing with others<sup>2</sup>. The GSBPM is used as a tool to classify these software tools according to which part of the statistical production process they support.

26. As the GSBPM is still relatively new, the benefits of applying it in practice are still being discovered and evaluated. One example of an emerging benefit is as a tool to manage process quality. International statistical work on quality has traditionally focused on output quality, but the topic of process quality is becoming increasingly important<sup>3</sup>. The GSBPM could provide a useful framework for this. Another potential application of the GSBPM is as a tool for measuring process costs – identifying the most expensive parts of a statistical production process, and hence the priority areas for investment to improve efficiency<sup>4</sup>.

27. Finally, the GSBPM is increasingly being seen as one of the four “pillars” for the industrialisation of statistical production, as outlined in the companion paper from the UNECE Secretariat for this meeting.

## IV. Conclusion

28. This paper argues that the GSBPM can be relatively easily applied to statistical business register maintenance, thus providing enhanced possibilities for benchmarking with other statistical production processes or with other organisations. This approach has clear potential benefits in terms of shared knowledge, methods and tools.

29. However, the value of this approach clearly needs to be demonstrated in practice. This will probably require at least one additional level of detail to that given in the example in Figure 2. However, as registers become increasingly more integral to statistical production, the ability to apply a common business process model has clear advantages. The UNECE would therefore greatly value feedback from any organisation able to take the next step and apply the GSBPM to their register maintenance process.

<sup>2</sup> <http://www1.unece.org/stat/platform/display/msis/Software+Inventory>

<sup>3</sup> For example, Eurostat recently conducted a survey of practices in Member States on this topic

<sup>4</sup> This short note from the Irish Central Statistics Office expands on some of these emerging uses of the GSBPM: <http://www1.unece.org/stat/platform/download/attachments/57835551/Some+ideas+on+communicating+the+GSBP+M.doc?version=1>