Joint UNECE/Eurostat/OECD Work Session on Statistical Metadata (METIS)

Generic Statistical Business Process Model

Version 4.0 – April 2009

Prepared by the UNECE Secretariat [1]

I. Background

1. The Joint UNECE / Eurostat / OECD Work Sessions on Statistical Metadata (METIS) have, over the last few years, been preparing a Common Metadata Framework (CMF)[2]. Part C of this framework is entitled "Metadata and the Statistical Cycle". This part refers to the phases of the statistical business process (also known as the statistical value chain or statistical cycle) and provides generic terms to describe them.

2. During a workshop to progress the development of Part C of the CMF, held in Vienna in July 2007[3], the participants agreed that the model currently used by Statistics New Zealand, with the addition of 'Archive' and 'Evaluate' phases, would provide a good basis for developing a "Generic Statistical Business Process Model" (GSBPM). A first draft of the GSBPM was presented by the UNECE Secretariat at the METIS Work Session in Luxembourg in April 2008[4]. Following two rounds of comments, another workshop was held in Lisbon in March 2009[5] to finalize the model. This current version of the model (version 4.0), was approved by the METIS Steering Group for public release in April 2009. It is considered final at the time of release, however, it is also expected that future updates may be necessary in the coming years, either to reflect experiences from implementing the model in practice, or due to the evolution of the nature of statistical production. The reader is therefore invited to check the website www.unece.org/stats/gsbpm to be sure of having the latest version.

II. The Model

Purpose

3. The original intention was for the GSBPM to provide a basis for statistical organizations to agree on standard terminology to aid their discussions on developing statistical metadata systems and processes. The GSBPM should therefore be seen as a flexible tool to describe and define the set of business processes needed to produce official statistics. The use of this model can also be envisaged in other separate, but often

[1] Prepared by Steven Vale (steven.vale@unece.org), based on previous work by Statistics New Zealand (for the first seven phases) and Statistics Canada (for the Archive phase), with considerable input and feedback from the members of the METIS group.
[2] See: http://www.unece.org/stats/cmf/
[3] The papers from this Workshop are available at: http://www.unece.org/stats/documents/2007.07.metis.htm
[4] See: http://www.unece.org/stats/documents/ece/ces/ge.40/2008/wp.17.e.pdf
[5] See: http://www.unece.org/stats/documents/2009.03.metis.htm

related contexts such as harmonizing statistical computing infrastructures, facilitating the sharing of software components, in the Statistical Data and Metadata eXchange (SDMX) User Guide for explaining the use of SDMX in a statistical organization, and providing a framework for process quality assessment and improvement. These other purposes for which the GSBPM can be used are elaborated further in Section VI.

Applicability

4. 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.

5. Whilst the typical statistical business process includes the collection and processing of raw data to produce statistical outputs, the GSBPM also applies 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. In such cases, it is likely that several sub-processes and possibly some phases (particularly the early ones) would be omitted.

6. As well as being applicable for processes which result in statistics, the GSBPM can also be applied to the development and maintenance of statistical registers, where 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.

7. Some elements of the GSBPM may be more relevant for one type of process than another, which may be influenced by the types of data sources used or the outputs to be produced. Some elements will overlap with each other, sometimes forming iterative loops. The GSBPM should therefore be applied and interpreted flexibly. It 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 business process, and the inter-dependencies between them. Although the presentation follows the logical sequence of steps in most statistical business processes, the elements of the model may occur in different orders in different circumstances. In this way the GSBPM aims to be sufficiently generic to be widely applicable, and to encourage a standard view of the statistical business process, without becoming either too restrictive or too abstract and theoretical.

8. In some cases it may be appropriate to group some of the elements of the model. For example, phases one to three could be considered to correspond to a single planning phase. In other cases, there may be a need to add another, more detailed level to the structure presented below to separately identify different components of the sub-processes. There may also be a requirement for a formal sign-off between phases, where the output from one phase is certified as suitable as input for the next. This sort of formal approval is implicit in the model, but may be implemented in many different ways depending on organizational requirements. The GSBPM should be seen as sufficiently flexible to apply in all of the above scenarios.

Structure

9. The GSBPM comprises four levels:

- Level 0, the statistical business process;
- Level 1, the nine phases of the statistical business process;
- Level 2, the sub-processes within each phase;
- Level 3, a description of those sub-processes.

10. Further levels of detail may be appropriate for certain statistical business processes or in certain organizations, but these are unlikely to be sufficiently generic to be included in this model. A diagram showing the phases (level 1) and sub-processes (level 2) is included in Section IV. The sub-processes are described in detail in Section V.

11. According to process modelling theory, each sub-process should have a number of clearly identified attributes, including:

- Input(s);
- Output(s);
- Purpose (value added);
- Owner;
- Guides (for example manuals and documentation);
- Enablers (people and systems);
- Feedback loops or mechanisms.

However, these attributes are likely to differ, at least to some extent, between statistical business processes, and between organizations. For this reason these attributes are rarely mentioned specifically in this generic model. It is, however, strongly recommended that they are identified when applying the model to any specific statistical business process.

12. The GSBPM also recognizes 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 organization. The first group are considered to be more important in the context of this model, however the second group should also be recognized as they have (often indirect) impacts on several parts of the model.

13. Over-arching statistical processes include the following. The first two are mostly closely related to the model, and are therefore shown in model diagrams and are elaborated further in Section VI.

- Quality management – This process includes quality assessment and control mechanisms. It recognizes the importance of evaluation and feedback throughout the statistical business process;
- Metadata management – Metadata are generated and processed within each phase, there is, therefore, a strong requirement for a metadata management system to ensure that the appropriate metadata retain their links with data throughout the GSBPM;

- Statistical framework management – This includes developing standards, for example methodologies, concepts and classifications that apply across multiple processes;
- Statistical programme management – This includes systematic monitoring and reviewing of emerging information requirements and emerging and changing data sources across all statistical domains. It may result in the definition of new statistical business processes or the redesign of existing ones;
- Knowledge management – This ensures that statistical business processes are repeatable, mainly through the maintenance of process documentation;
- Data management – This includes process-independent considerations such as general data security, custodianship and ownership;
- Process data management – This includes the management of data and metadata generated by and providing information on all parts of the statistical business process.
- Provider management – This includes cross-process burden management, as well as topics such as profiling and management of contact information (and thus has particularly close links with statistical business processes that maintain registers);
- Customer management – This includes general marketing activities, promoting statistical literacy, and dealing with non-specific customer feedback.

14. More general over-arching processes include:

- Human resource management;
- Financial management;
- Project management;
- Legal framework management;
- Organizational framework management;
- Strategic planning.

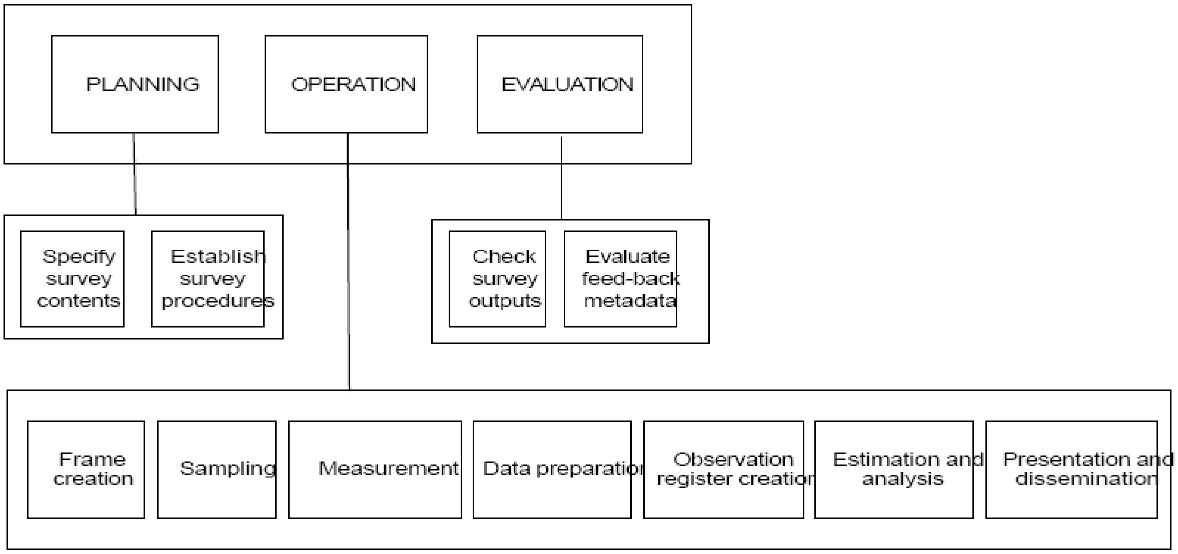
III. Relationships with Other Models and Standards

15. The GSBPM has been developed drawing heavily on the Generic Business Process Model developed by Statistics New Zealand, supplemented by input from Statistics Canada on phase 8 (Archive), and other statistical organizations with experience of statistical process modelling. However, a number of other related models and standards exist for different purposes and in different organizations, both at the national and international level. It would not be practical to give details of all national models here[6], but the main international models and standards are considered below, and related to the GSBPM. A diagram of this relationship is included at the end of this section, and shows that the GSBPM can also be seen as the union of the other models, since it reflects all of their components.

[6] Though examples from Australia and Norway can be found at the following addresses:
http://www1.unece.org/stat/platform/display/metis/2.+Statistical+metadata+systems+and+the+statistical+business+process+(Australia)
http://www.ssb.no/english/subjects/00/90/doc_200817_en/doc_200817_en.pdf

Information Systems Architecture for National and International Statistical Offices

16. This set of guidelines and recommendations was published by the United Nations in 1999. It contains the model of the phases and processes of a survey processing system shown below. Although different in presentation to the GSBPM, the contents are largely the same.



Source: Information Systems Architecture for National and International Statistical Offices – Guidelines and Recommendations, United Nations, 1999, http://www.unece.org/stats/documents/information_systems_architecture/1.e.pdf

The Eurostat "Cycle de Vie des Données" (CVD) model

17. The CVD project ("Cycle de Vie des Données" or "Data Life Cycle") aims to fundamentally revise the way Eurostat treats statistical data, by providing a coherent set of concepts, metadata structures and IT tools to be applied in all statistical domains. It also aims to deliver significant benefits, such as economies of scale for the development and evolution of computing tools and the pursuit of important corporate objectives, such as quality orientation and easier mobility of domain managers. The CVD project centres on metadata as its basic integrating concept, recognizing that metadata have a ubiquitous and overwhelming role in the statistical production process. It also recognizes the GSBPM for modelling statistical business processes. SDMX standards and guidelines play a key role in the whole CVD approach, from data transmission to dissemination, as well as for the exchange of data between the components of the production system.

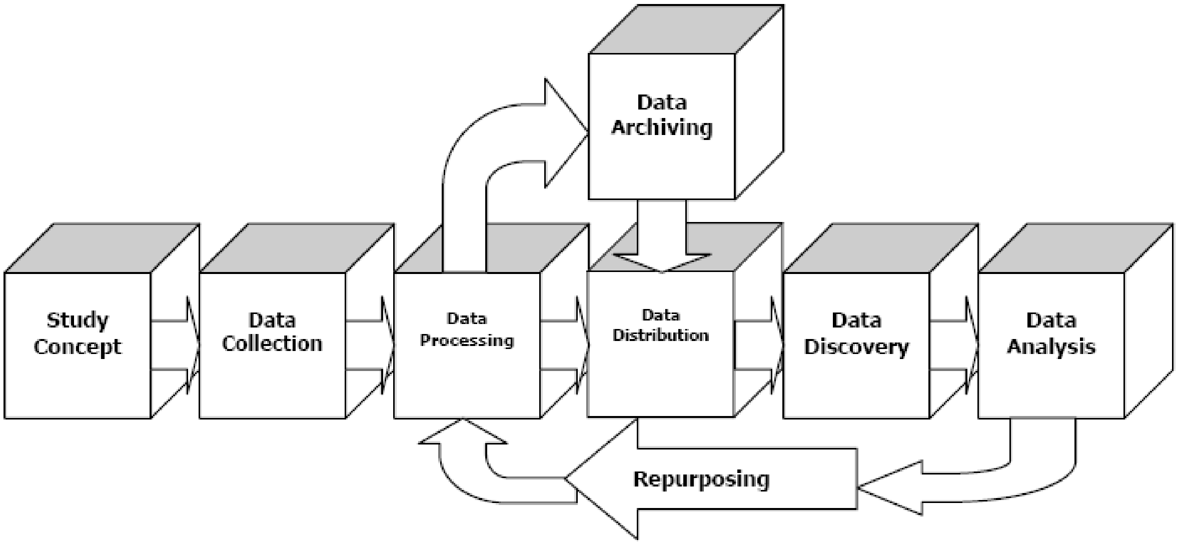
The DDI 3.0 Combined Life Cycle Model

18. This model has been developed within the Data Documentation Initiative (DDI), an international effort to establish a standard for technical documentation describing social science data. The DDI Alliance comprises mainly academic and research institutions, hence the scope of the model below is rather different to the GSBPM, which specifically applies to official statistical organizations. Despite this, the statistical business process

appears to be quite similar between official and non-official statistics producers, as is clear from the high level of consistency between the models.

19. The main differences between the models are:

- The GSBPM places data archiving at the end of the process, after the analysis phase. It may also form the end of processing within a specific organization in the DDI model, but a key difference is that the DDI model is not necessarily limited to processes within one organization. Steps such as “Data analysis” and “Repurposing” may be carried out by different organizations to the one that collected the data.
- The DDI model replaces the dissemination phase with “Data Distribution” which takes place before the analysis phase. This reflects a difference in focus between the research and official statistics communities, with the latter putting a stronger emphasis on disseminating data, rather than research based on data disseminated by others.
- The DDI model contains the process of “Repurposing”, defined as the secondary use of a data set, or the creation of a real or virtual harmonized data set. This generally refers to some re-use of a data-set that was not originally foreseen in the design and collect phases. This is covered in the GSBPM phase 1 (Specify Needs), where there is a sub-process to check the availability of existing data, and use them wherever possible. It is also reflected in the data integration sub-process within phase 5 (Process).
- The DDI model has separate phases for data discovery and data analysis, whereas these functions are combined within phase 6 (Analysis) in the GSBPM. In some cases, elements of the GSBPM analysis phase may also be covered in the DDI “Data Processing” phase, depending on the extent of analytical work prior to the “Data distribution” phase



Source: Data Documentation Initiative (DDI) Technical Specification, Part I: Overview, Version 3.0, April 2008, http://www.ddialliance.org.

SDMX

20. The SDMX (Statistical Data and Metadata eXchange) standards[7] do not provide a model for statistical business processes in the same sense as the three cases above. However they do provide standard terminology for statistical data and metadata, as well as technical standards and content-oriented guidelines for data and metadata transfer, which can also be applied between sub-processes within a statistical organization. The use of commonly agreed data and metadata structures allows exchanged data and metadata to be mapped or translated to and from internal statistical systems. To facilitate this, the SDMX sponsors published a set of cross-domain concepts in January 2009. The use of these common concepts may help to standardise and improve data and metadata transmissions between different organisations, even when models and systems are different. As far as metadata transmission is concerned, the mapping between the metadata concepts used by different international organizations, which is also present in the SDMX content-oriented guidelines package, supports the idea of open exchange and sharing of metadata based on common terminology.

21. The relationship between the model and SDMX was discussed at the April 2008 meeting of the METIS group. The final report of that meeting[8] (paragraph 22) records a suggestion to incorporate the model into the Metadata Common Vocabulary and/or SDMX as a cross-domain concept. SDMX aims, through the Content-oriented Guidelines to harmonize data and metadata terminology and quality, as well as providing transmission standards. The GSBPM, in offering standard terminology for the different phases and sub-processes of the statistical business process, would seem to complement, and fit logically within the SDMX Content-oriented Guidelines.

[7] See www.sdmx.org

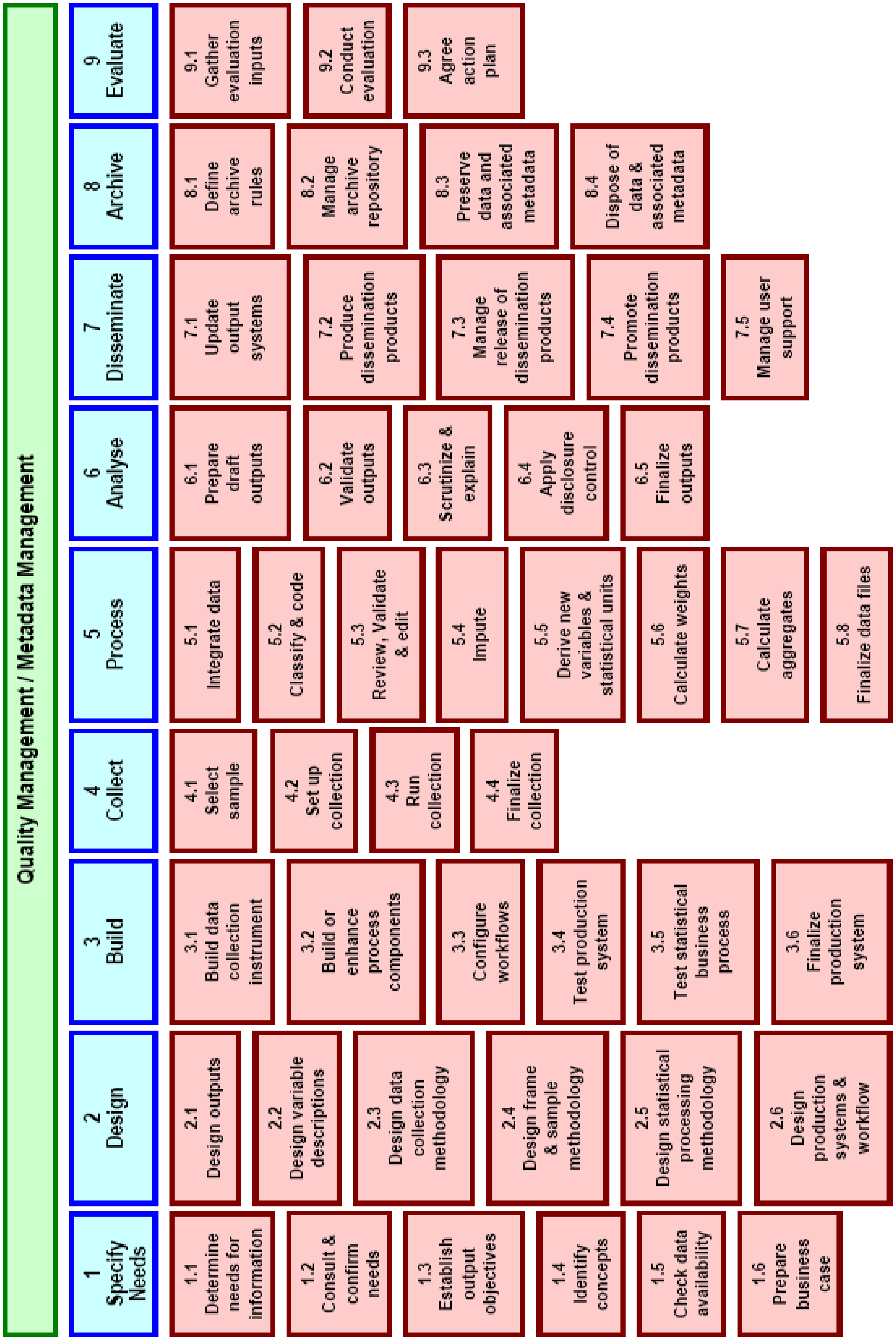
[8] http://www.unece.org/stats/documents/ece/ces/ge.40/2008/zip.9.e.pdf

Relationships between the Different Models

Generic Statistical Business Process Model	Information Systems Architecture Model	DDI 3.0 Combined Life Cycle Model
1 Specify Needs	Planning - Specify survey contents - Establish survey procedures	Study Concept Repurposing (part)
2 Design		
3 Build		
4 Collect	Operation (part) - Frame creation - Sampling - Measurement	Data Collection
5 Process	Operation (part) - Data preparation - Observation register creation	Data Processing (mostly) Repurposing (part)
6 Analyse	Operation (part) - Estimation and analysis Evaluation (part) - Check survey outputs	Data Discovery Data Analysis Data Processing (part)
7 Disseminate	Operation (part) - Presentation and dissemination	Data Distribution
8 Archive		Data Archiving
9 Evaluate	Evaluation (part) - Evaluate feedback metadata	

Quality Management		
Metadata Management		

IV. Levels 1 and 2 of the Generic Statistical Business Process Model



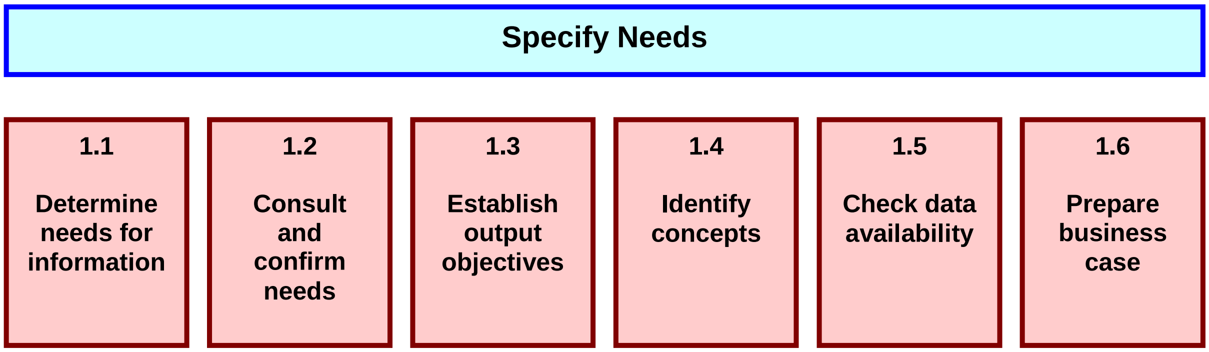
Quality Management / Metadata Management

- 1 Specify Needs
  - 1.1 Determine needs for information
  - 1.2 Consult & confirm needs
  - 1.3 Establish output objectives
  - 1.4 Identify concepts
  - 1.5 Check data availability
  - 1.6 Prepare business case
- 2 Design
  - 2.1 Design outputs
  - 2.2 Design variable descriptions
  - 2.3 Design data collection methodology
  - 2.4 Design frame & sample methodology
  - 2.5 Design statistical processing methodology
  - 2.6 Design production systems & workflow
- 3 Build
  - 3.1 Build data collection instrument
  - 3.2 Build or enhance process components
  - 3.3 Configure workflows
  - 3.4 Test production system
  - 3.5 Test statistical business process
  - 3.6 Finalize production system
- 4 Collect
  - 4.1 Select sample
  - 4.2 Set up collection
  - 4.3 Run collection
  - 4.4 Finalize collection
- 5 Process
  - 5.1 Integrate data
  - 5.2 Classify & code
  - 5.3 Review, Validate & edit
  - 5.4 Impute
  - 5.5 Derive new variables & statistical units
  - 5.6 Calculate weights
  - 5.7 Calculate aggregates
  - 5.8 Finalize data files
- 6 Analyse
  - 6.1 Prepare draft outputs
  - 6.2 Validate outputs
  - 6.3 Scrutinize & explain
  - 6.4 Apply disclosure control
  - 6.5 Finalize outputs
- 7 Disseminate
  - 7.1 Update output systems
  - 7.2 Produce dissemination products
  - 7.3 Manage release of dissemination products
  - 7.4 Promote dissemination products
  - 7.5 Manage user support
- 8 Archive
  - 8.1 Define archive rules
  - 8.2 Manage archive repository
  - 8.3 Preserve data and associated metadata
  - 8.4 Dispose of data & associated metadata
- 9 Evaluate
  - 9.1 Gather evaluation inputs
  - 9.2 Conduct evaluation
  - 9.3 Agree action plan

V. Levels 2 and 3 of the Generic Statistical Business Process Model

22. This section considers each phase in turn, identifying the various sub-processes within that phase, and describing their contents. It therefore covers levels 2 and 3 of the GSBPM.

Phase 1 – Specify Needs


Specify Needs
1.1 Determine needs for information
1.2 Consult and confirm needs
1.3 Establish output objectives
1.4 Identify concepts
1.5 Check data availability
1.6 Prepare business case

23. This phase is triggered when a need for new statistics is identified, or feedback about current statistics initiates a review. It determines whether there is a presently unmet demand, externally and / or internally, for the identified statistics and whether the statistical organization can produce them.

24. In this phase the organization:

- determines the need for the statistics;
- confirms, in more detail, the statistical needs of the stakeholders;
- establishes the high level objectives of the statistical outputs;
- identifies the relevant concepts and variables for which data are required;
- checks if current collections and / or methodologies can meet these needs;
- prepares the business case to get approval to produce the statistics.

25. This phase is broken down into six sub-processes. These are generally sequential, from left to right, but can also occur in parallel, and can be iterative. The sub-processes are:

1.1. Determine needs for information - This sub-process includes the initial investigation and identification of what statistics are needed and what is needed of the statistics. It also includes consideration of practice amongst other (national and international) statistical organizations producing similar data, and in particular the methods used by those organizations.

1.2. Consult and confirm needs - This sub-process focuses on consulting with the stakeholders and confirming in detail the needs for the statistics. A good understanding of user needs is required so that the statistical organization knows not only what it is expected to deliver, but also when, how, and, perhaps most importantly, why. For second and subsequent iterations of this phase, the main focus will be on determining whether previously identified needs have changed. This detailed understanding of user needs is the critical part of this sub-process.

1.3. Establish output objectives - This sub-process identifies the statistical outputs that are required to meet the user needs identified in sub-process 1.2 (Consult and confirm need). It includes agreeing the suitability of the proposed outputs and their quality measures with users.

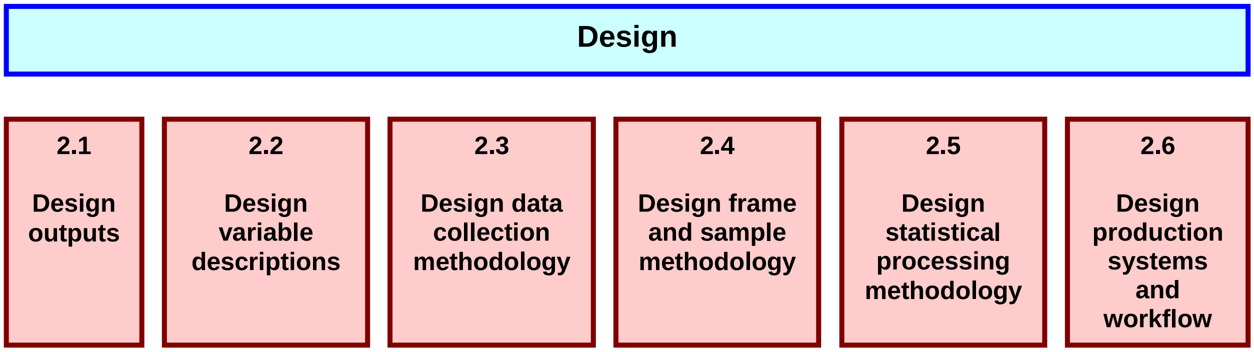
1.4. Identify concepts - This sub-process clarifies the required concepts to be measured by the business process from the point of view of the user. At this stage the concepts identified may not align with existing statistical standards. This alignment, and the choice or definition of the statistical concepts and variables to be used, takes place in sub-process 2.2.

1.5. Check data availability - 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 data sources and their methodologies, to determine whether they would be suitable for use for statistical purposes. 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.

1.6. Prepare business case - This sub-process documents the findings of the other sub-processes in this phase in the form a business case to get approval to implement the new or modified statistical business process. Such a business case would typically also include:

- A description of the “As-Is” business process (if it already exists), with information on how the current statistics are produced, highlighting any inefficiencies and issues to be addressed;
- The proposed “To-Be” solution, detailing how the statistical business process will be developed to produce the new or revised statistics;
- An assessment of costs and benefits, as well as any external constraints.

Phase 2 – Design


Design
2.1
Design outputs
2.2
Design variable descriptions
2.3
Design data collection methodology
2.4
Design frame and sample methodology
2.5
Design statistical processing methodology
2.6
Design production systems and workflow

26. This phase describes the development and design activities, and any associated practical research work needed to define the statistical outputs, concepts, methodologies, collection instruments and operational processes. For statistical outputs produced on a

regular basis, this phase usually occurs for the first iteration, and whenever improvement actions are identified in phase 9 (Evaluate) of a previous iteration.

27. This phase is broken down into six sub-processes, which are generally sequential, from left to right, but can also occur in parallel, and can be iterative. These sub-processes are:

2.1. Design outputs – This sub-process contains the detailed design of the statistical outputs to be produced, including the related development work and preparation of the systems and tools used in phase 7 (Disseminate). Outputs should be designed, wherever possible, to follow existing standards, so inputs to this process may include metadata from similar or previous collections, international standards, and information about practices in other statistical organizations from sub-process 1.1 (Determine need for information).

2.2 Design variable descriptions – This sub-process defines the statistical variables to be collected via the data collection instrument, as well as any other variables that will be derived from them in sub-process 5.5 (Derive new variables and statistical units), and any classifications that will be used. It is expected that existing national and international standards will be followed wherever possible. This sub-process may need to run in parallel with sub-process 2.3 (Design data collection methodology), as the definition of the variables to be collected, and the choice of data collection instrument may be inter-dependent to some degree. Preparation of metadata descriptions of collected and derived variables and classifications is a necessary precondition for subsequent phases.

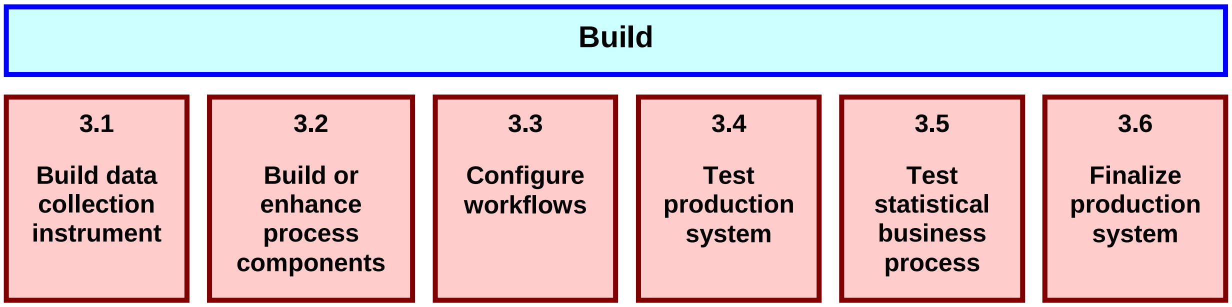
2.3. Design data collection methodology - This sub-process determines the most appropriate data collection method(s) and instrument(s). The actual activities in this sub-process will vary according to the type of collection instruments required, which can include computer assisted interviewing, paper questionnaires, administrative data interfaces and data integration techniques. This sub-process includes the design of questions and response templates (in conjunction with the variables and 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 standardize terms and conditions). This sub-process also includes the design of process-specific provider management systems.

2.4. Design frame and sample methodology - This sub-process identifies and specifies the population of interest, defines a sampling frame (and, where necessary, the register from which it is derived), and determines the most appropriate sampling criteria and methodology (which could include complete enumeration). Common sources are administrative and statistical registers, censuses and sample surveys. This sub-process describes how these sources can be combined if needed. Analysis of whether the frame covers the target population should be performed. A sampling plan should be made: The actual sample is created sub-process 4.1 (Select sample), using the methodology, specified in this sub-process.

2.5. Design statistical processing methodology - This sub-process designs the statistical processing methodology to be applied during phase 5 (Process), and Phase 6 (Analyse). This can include specification of routines for coding, editing, imputing, estimating, integrating, validating and finalising data sets.

2.6. Design production systems and workflow - This sub-process determines the workflow from data collection to archiving, taking an overview of all the processes required within the whole statistical production process, and ensuring that they fit together efficiently with no gaps or redundancies. Various systems and databases are needed throughout the process. A general principle is to reuse processes and technology across many statistical business processes, so existing systems and databases should be examined first, to determine whether they are fit for purpose for this specific process, then, if any gaps are identified, new solutions should be designed. This sub-process also considers how staff will interact with systems, and who will be responsible for what and when.

Phase 3 – Build


Build
3.1 Build data collection instrument
3.2 Build or enhance process components
3.3 Configure workflows
3.4 Test production system
3.5 Test statistical business process
3.6 Finalize production system

28. This phase builds and tests the production systems to the point where they are ready for use in the "live" environment. For statistical outputs produced on a regular basis, this phase usually occurs for the first iteration, and following a review or a change in methodology, rather than for every iteration. It is broken down into six sub-processes, which are generally sequential, from left to right, but can also occur in parallel, and can be iterative. These sub-processes are:

3.1. Build data collection instrument - This sub-process describes the activities to build the collection instruments to be used during the phase 4 (Collect). The collection instrument is generated or built based on the design specifications created during phase 2 (Design). 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 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.

3.2. Build or enhance process components - This sub-process describes the activities to build new and enhance existing software components needed for the business process, as designed in Phase 2 (Design). Components may include dashboard functions and features, data repositories, transformation tools, workflow framework components, provider and metadata management tools.

3.3. Configure workflows - This sub-process configures the workflow, systems and transformations used within the statistical business processes, from data collection, right through to archiving the final statistical outputs. It ensures that the workflow specified in sub-process 2.6 (Processing system and workflow) works in practice.

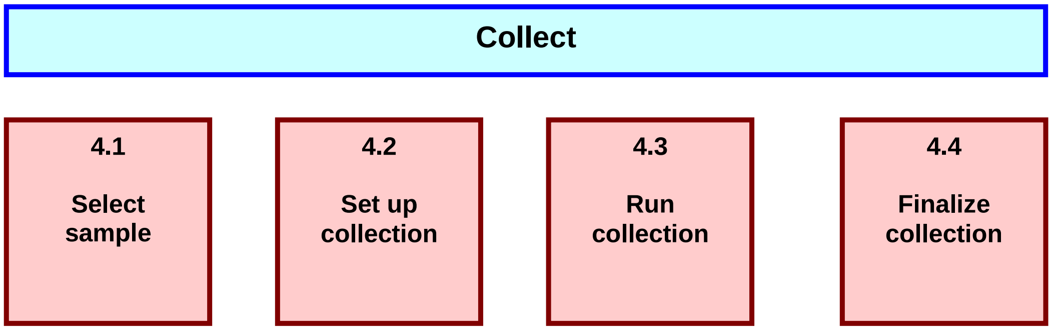
3.4. Test production system – This sub-process is concerned with the testing of computer systems and tools. It includes technical testing and sign-off of new programmes and routines, as well as confirmation that existing routines from other statistical business processes are suitable for use in this case. Whilst part of this activity concerning the testing of individual components could logically be linked with sub-process 3.2 (Build or enhance process components), this sub-process also includes testing of interactions between components, and ensuring that the production system works as a coherent set of components.

3.5. Test statistical business process - This sub-process describes the activities to manage a field test or pilot of the statistical business process. Typically it includes a small-scale data collection, to test collection instruments, followed by processing and analysis of the collected data, to ensure the statistical business process performs as expected. Following the pilot, it may be necessary to go back to a previous step and make adjustments to instruments, systems or components. For a major statistical business process, e.g. a population census, there may be several iterations until the process is working satisfactorily.

3.6. Finalize production systems - This sub-process includes the activities to put the process, including workflow systems, modified and newly-built components into production ready for use by business areas. The activities include:

- producing documentation about the process components, including technical documentation and user manuals
- training the business users on how to operate the process
- moving the process components into the production environment, and ensuring they work as expected in that environment (this activity may also be part of sub-process 3.4 (Test production system)).

Phase 4 – Collect



29. This phase collects all necessary data, using different collection modes (including extractions from administrative and statistical registers and databases), and loads them into the appropriate data environment. It does not include any transformations of collected data, as these are all done in phase 5 (Process). For statistical outputs produced regularly, this phase occurs in each iteration.

30. The Collect phase is broken down into four sub-processes, which are generally sequential, from left to right, but can also occur in parallel, and can be iterative. These sub-processes are:

4.1. Select sample - This sub-process establishes the frame and selects the sample for this iteration of the collection, as specified in sub-process 2.4 (Design frame and sample methodology). It also includes the coordination of samples between instances of the same statistical business process (for example to manage overlap or rotation), and between different processes using a common frame or register (for example to manage overlap or to spread response burden). Quality assurance, approval and maintenance of the frame and the selected sample are also undertaken in this sub-process, though maintenance of underlying registers, from which frames for several statistical business processes are drawn, is treated as a separate business process. The sampling aspect of this sub-process is not usually relevant for processes based entirely on the use of pre-existing data sources (e.g. administrative data) as such processes generally create frames from the available data and then follow a census approach.

4.2. Set up collection - This sub-process ensures that the people, processes and technology are ready to collect data, 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. This sub-process includes:

- preparing a collection strategy
- training collection staff
- ensuring collection resources are available e.g. laptops
- configuring collection systems to request and receive the data;
- ensuring the security of data to be collected;
- preparing collection instruments (e.g. printing questionnaires, pre-filling them with existing data, loading questionnaires and data onto interviewers' computers etc.).

4.3. Run collection - This sub-process is where the collection is implemented, with the different collection instruments being used to collect the data. It includes the initial contact with providers and any subsequent follow-up or reminder actions. It records when and how providers were contacted, and whether they have responded. This sub-process also includes the management of the providers involved in the current collection, ensuring that the relationship between the statistical organization and data providers remains positive, and recording and responding to comments, queries and complaints. For administrative data, this process is brief: the provider is either contacted to send the data, or sends it as scheduled. When the collection meets its targets (usually based on response rates) the collection is closed and a report on the collection is produced.

4.4. Finalize collection - This sub-process includes loading the collected data and metadata into a suitable electronic environment for further processing in phase 5 (Process). It may include automatic data take-on, for example using optical character recognition tools to extract data from paper questionnaires, or converting the formats of data files received from other organizations. In cases where there is a physical data collection instrument, such as a paper questionnaire, which is not needed for further processing, this sub-process manages the archiving of that material in conformance with the principles established in phase 8 (Archive).

Phase 5 – Process

Process

5.1	5.2	5.3	5.4	5.5	5.6	5.7	5.8
Integrate data	Classify and code	Review, validate and edit	Impute	Derive new variables and statistical units	Calculate weights	Calculate aggregates	Finalize data files

31. This phase describes the cleaning of data records and their preparation for analysis. It is made up of sub-processes that check, clean, and transform the collected data, and may be repeated several times. For statistical outputs produced regularly, this phase occurs in each iteration. The sub-processes in this phase can apply to data from both statistical and non-statistical sources (with the possible exception of sub-process 5.6 (Calculate weights), which is usually specific to survey data).

32. The “Process” and “Analyse” phases can be iterative and parallel. Analysis can reveal a broader understanding of the data, which might make it apparent that additional processing is needed. Activities within the “Process” and “Analyse” phases may commence before the “Collect” phase is completed. This enables the compilation of provisional results where timeliness is an important concern for users, and increases the time available for analysis. The key difference between these phases is that “Process” concerns transformations of microdata, whereas “Analyse” concerns the further treatment of statistical aggregates.

33. This phase is broken down into eight sub-processes, which may be sequential, from left to right, but can also occur in parallel, and can be iterative. These sub-processes are:

5.1. Integrate data - This sub-process integrates data from one or more sources. The input data can be from a mixture of external or internal data sources, and a variety of collection modes, including extracts of administrative data. The result is a harmonized data set. Data integration typically includes:

- matching / record linkage routines, with the aim of linking data from different sources, where those data refer to the same unit;
- prioritising, when two or more sources contain data for the same variable (with potentially different values).

Data integration may take place at any point in this phase, before or after any of the other sub-processes. There may also be several instances of data integration in any statistical business process. Following integration, depending on data protection requirements, data may be anonymized, that is stripped of identifiers such as name and address, to help to protect confidentiality.

5.2. Classify and code - This sub-process classifies and codes the input data. For example automatic (or clerical) coding routines may assign numeric codes to text responses according to a pre-determined classification scheme.

5.3. Review, validate and edit - This sub-process applies to collected micro-data, and looks at each record to try to identify (and where necessary correct) potential problems, errors and discrepancies such as outliers, item non-response and miscoding. It can also be referred to as input data validation. It may be run iteratively, validating data against predefined edit rules, usually in a set order. It may apply automatic edits, or raise alerts for manual inspection and correction of the data. Reviewing, validating and editing can apply to unit records both from surveys and administrative sources, before and after integration. In certain cases, imputation (sub-process 5.4) may be used as a form of editing.

5.4. Impute - Where data are missing or unreliable, estimates may be imputed, often using a rule-based approach. Specific steps typically include:

- the identification of potential errors and gaps;
- the selection of data to include or exclude from imputation routines;
- imputation using one or more pre-defined methods e.g. “hot-deck” or “cold-deck”;
- writing the imputed data back to the data set, and flagging them as imputed;
- the production of metadata on the imputation process;

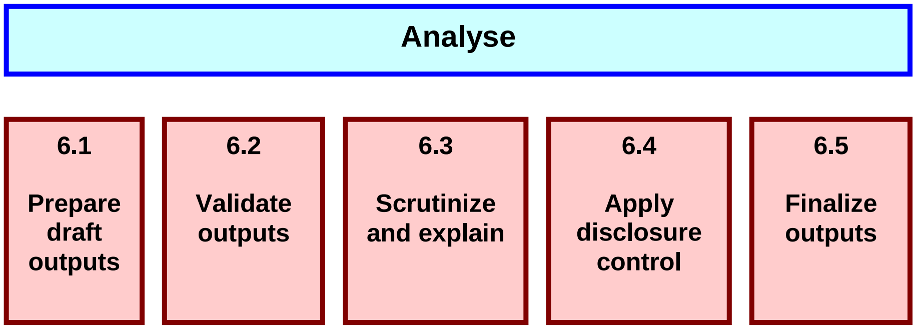
5.5. Derive new variables and statistical units - This sub-process derives (values for) variables and statistical units that are not explicitly provided in the collection, but are needed to deliver the required outputs. It derives new variables by applying arithmetic formulae to one or more of the variables that are already present in the dataset. This may need to be iterative, as some derived variables may themselves be based on other derived variables. It is therefore important to ensure that variables are derived in the correct order. New statistical units may be derived by aggregating or splitting data for collection units, or by various other estimation methods. Examples include deriving households where the collection units are persons, or enterprises where the collection units are legal units.

5.6. Calculate weights - This sub process creates weights for unit data records according to the methodology created in sub-process 2.5 (Design statistical processing methodology). These weights can be used to “gross-up” sample survey results to make them representative of the target population, or to adjust for non-response in total enumerations.

5.7. Calculate aggregates - This sub process creates aggregate data and population totals from micro-data. It includes summing data for records sharing certain characteristics, determining measures of average and dispersion, and applying weights from sub-process 5.6 to sample survey data to derive population totals.

5.8. Finalize data files – This sub-process brings together the results of the other sub-processes in this phase and results in a data file (usually of macro-data), which is used as the input to phase 6 (Analyse). Sometimes this may be an intermediate rather than a final file, particularly for business processes where there are strong time pressures, and a requirement to produce both preliminary and final estimates.

Phase 6 – Analyse


Analyse
6.1 Prepare draft outputs
6.2 Validate outputs
6.3 Scrutinize and explain
6.4 Apply disclosure control
6.5 Finalize outputs

34. In this phase, statistics are produced, examined in detail and made ready for dissemination. This phase includes the sub-processes and activities that enable statistical analysts to understand the statistics produced. For statistical outputs produced regularly, this phase occurs in every iteration. The Analyse phase and sub-processes are generic for all statistical outputs, regardless of how the data were sourced.

35. The Analyse phase is broken down into five sub-processes, which are generally sequential, from left to right, but can also occur in parallel, and can be iterative. The sub-processes are:

6.1. Prepare draft outputs - This sub-process is where the data collected are transformed into statistical outputs. It includes the production of additional measurements such as indices, trends or seasonally adjusted series, as well as the recording of quality characteristics.

6.2. Validate outputs - This sub-process is where statisticians validate the quality of the outputs produced, in accordance with a general quality framework and with expectations. This sub-process also includes activities involved with the gathering of intelligence, with the cumulative effect of building up a body of knowledge about a specific statistical domain. This knowledge is then applied to the current collection, in the current environment, to identify any divergence from expectations and to allow informed analyses. Validation activities can include:

- checking that the population coverage and response rates are as required;
- comparing the statistics with previous cycles (if applicable);
- confronting the statistics against other relevant data (both internal and external);
- investigating inconsistencies in the statistics;
- performing macro editing;
- validating the statistics against expectations and domain intelligence.

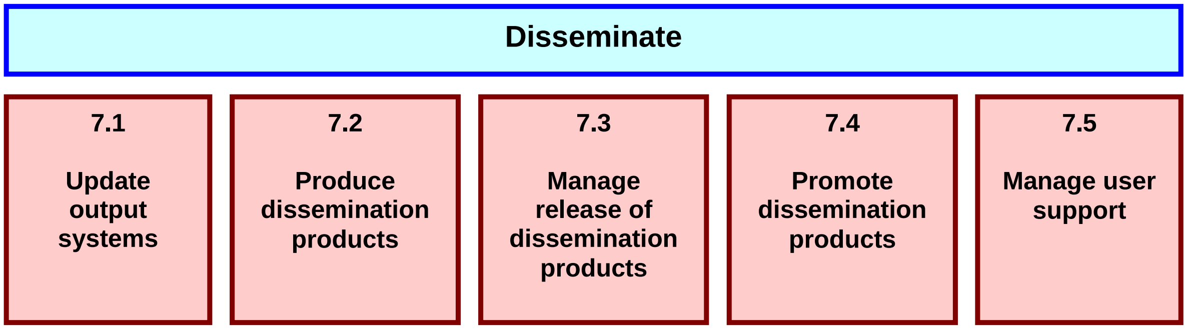
6.3. Scrutinize and explain - This sub-process is where the in-depth understanding of the outputs is gained by statisticians. They use that understanding to scrutinize and explain the statistics produced for this cycle by assessing how well the statistics reflect their initial expectations, viewing the statistics from all perspectives using different tools and media, and carrying out in-depth statistical analyses.

6.4. Apply disclosure control – This sub-process ensures that the data (and metadata) to be disseminated do not breach the appropriate rules on confidentiality. This may include checks for primary and secondary disclosure, as well as the application of data suppression or perturbation techniques.

6.5. Finalize outputs - This sub-process ensures the statistics and associated information are fit for purpose and reach the required quality level, and are thus ready for use. It includes:

- completing consistency checks;
- determining the level of release, and applying caveats;
- collating supporting information, including interpretation, briefings, measures of uncertainty and any other necessary metadata;
- producing the supporting internal documents;
- pre-release discussion with appropriate internal subject matter experts;
- approving the statistical content for release.

Phase 7 – Disseminate


Disseminate
7.1
Update output systems
7.2
Produce dissemination products
7.3
Manage release of dissemination products
7.4
Promote dissemination products
7.5
Manage user support

36. This phase manages the release of the statistical products to customers. For statistical outputs produced regularly, this phase occurs in each iteration. It is made up of five sub-processes, which are generally sequential, from left to right, but can also occur in parallel, and can be iterative. These sub-processes are:

7.1. Update output systems - This sub-process manages the update of systems where data and metadata are stored for dissemination purposes, including:

- formatting data and metadata ready to be put into output databases;
- loading data and metadata into output databases;
- ensuring data are linked to the relevant metadata.

Note: formatting, loading and linking of metadata should preferably mostly take place in earlier phases, but this sub-process includes a check that all of the necessary metadata are in place ready for dissemination.

7.2. Produce dissemination products - This sub-process produces the products, as previously designed (in sub-process 2.1), to meet user needs. The products can take many forms including printed publications, press releases and web sites. Typical steps include:

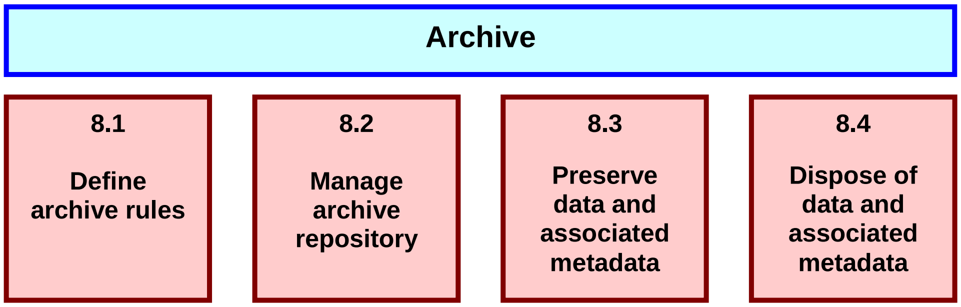
- preparing the product components (explanatory text, tables, charts etc.);
- assembling the components into products;
- editing the products and checking that they meet publication standards.

7.3. Manage release of dissemination products - This sub-process ensures that all elements for the release are in place including managing the timing of the release. It includes briefings for specific groups such as the press or ministers, as well as the arrangements for any pre-release embargoes. It also includes the provision of products to subscribers.

7.4. Promote dissemination products – Whilst marketing in general can be considered to be an over-arching process, this sub-process concerns the active promotion of the statistical products produced in a specific statistical business process, to help them reach the widest possible audience. It includes the use of customer relationship management tools, to better target potential users of the products, as well as the use of tools including web sites, wikis and blogs to facilitate the process of communicating statistical information to users.

7.5. Manage user support - This sub-process ensures that customer queries are recorded, and that responses are provided within agreed deadlines. These queries should be regularly reviewed to provide an input to the over-arching quality management process, as they can indicate new or changing user needs.

Phase 8 – Archive


Archive
8.1
Define archive rules
8.2
Manage archive repository
8.3
Preserve data and associated metadata
8.4
Dispose of data and associated metadata

37. This phase manages the archiving and disposal of statistical data and metadata. Given the reduced costs of data storage, it is possible that the archiving strategy adopted by a statistical organization does not include provision for disposal, so the final sub-process

may not be relevant for all statistical business processes. In other cases, disposal may be limited to intermediate files from previous iterations, rather than disseminated data.

38. For statistical outputs produced regularly, archiving occurs in each iteration, however defining the archiving rules is likely to occur less regularly. This phase is made up of four sub-processes, which are generally sequential, from left to right, but can also occur in parallel, and can be iterative. These sub-processes are:

8.1. Define archive rules – This sub-process is where the archiving rules for the statistical data and metadata resulting from a statistical business process are determined. The requirement to archive intermediate outputs such as the sample file, the raw data from the collect phase, and the results of the various stages of the process and analyse phases should also be considered. The archive rules for a specific statistical business process may be fully or partly dependent on the more general archiving policy of the statistical organization, or, for national organizations, on standards applied across the government sector. The rules should include consideration of the medium and location of the archive, as well as the requirement for keeping duplicate copies. They should also consider the conditions (if any) under which data and metadata should be disposed of. (Note – this sub-process is logically strongly linked to Phase 2 – Design, at least for the first iteration of a statistical business process).

8.2. Manage archive repository – This sub-process concerns the management of one or more archive repositories. These may be databases, or may be physical locations where copies of data or metadata are stored. It includes:

- maintaining catalogues of data and metadata archives, with sufficient information to ensure that individual data or metadata sets can be easily retrieved;
- testing retrieval processes;
- periodic checking of the integrity of archived data and metadata;
- upgrading software-specific archive formats when software changes.

This sub-process may cover a specific statistical business process or a group of processes, depending on the degree of standardization within the organization. Ultimately it may even be considered to be an over-arching process if organization-wide standards are put in place.

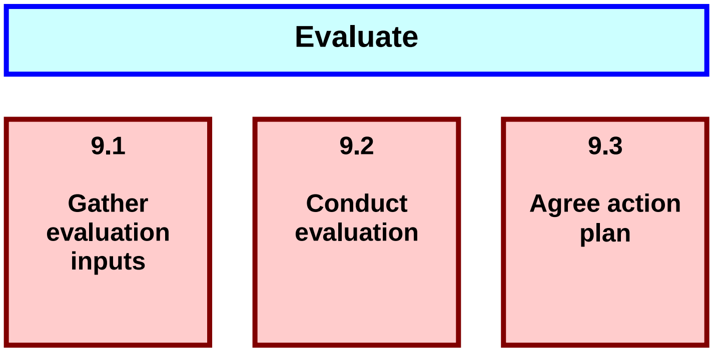
8.3. Preserve data and associated metadata – This sub-process is where the data and metadata from a specific statistical business process are archived. It includes:

- identifying data and metadata for archiving in line with the rules defined in 8.1;
- formatting those data and metadata for the repository;
- loading or transferring data and metadata to the repository;
- cataloguing the archived data and metadata;
- verifying that the data and metadata have been successfully archived.

8.4. Dispose of data and associated metadata – This sub-process is where the data and metadata from a specific statistical business process are disposed of. It includes;

- identifying data and metadata for disposal, in line with the rules defined in 8.1;
- disposal of those data and metadata;
- recording that those data and metadata have been disposed of.

Phase 9 – Evaluate



39. This phase manages the evaluation of a specific instance of a statistical business process, as opposed to the more general over-arching process of statistical quality management described in Section VI. It logically takes place at the end of the instance of the process, but relies on inputs gathered throughout the different phases. For statistical outputs produced regularly, evaluation should, at least in theory occur for each iteration, determining whether future iterations should take place, and if so, whether any improvements should be implemented. However, in some cases, particularly for regular and well established statistical business processes, evaluation may not be formally carried out for each iteration. In such cases, this phase can be seen as providing the decision as to whether the next iteration should start from phase 1 (Specify needs) or from some later phase (often phase 4 (Collect)).

40. This phase is made up of three sub-processes, which are generally sequential, from left to right, but which can overlap to some extent in practice. These sub-processes are:

9.1. Gather evaluation inputs – Evaluation material can be produced in any other phase or sub-process. It may take many forms, including feedback from users, process metadata, system metrics and staff suggestions. Reports of progress against an action plan agreed during a previous iteration may also form an input to evaluations of subsequent iterations. This sub-process gathers all of these inputs, and makes them available for the person or team producing the evaluation.

9.2. Conduct evaluation **–** This sub-process analyzes the evaluation inputs and synthesizes them into an evaluation report. The resulting report should note any quality issues specific to this iteration of the statistical business process, and should make recommendations for changes if appropriate. These recommendations can cover changes to any phase or sub-process for future iterations of the process, or can suggest that the process is not repeated.

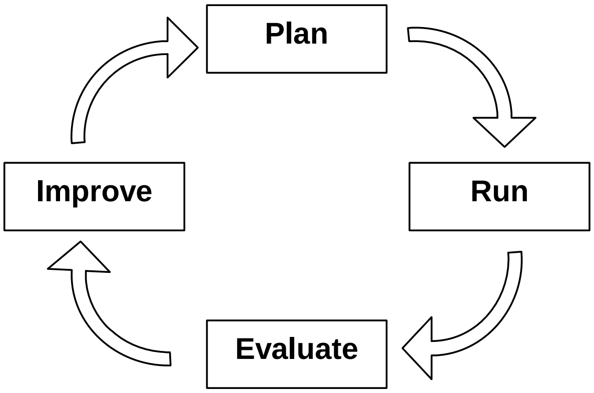
9.3. Agree an action plan **–** This sub-process brings together the necessary decision-making power to form and agree an action plan based on the evaluation report. It should also include consideration of a mechanism for monitoring the impact of those actions, which may, in turn, provide an input to evaluations of future iterations of the process.

VI. Over-arching processes

Quality Management

41. This process is present throughout the model. It is closely linked to Phase 9 (Evaluate), which has the specific role of evaluating individual instances of a statistical business process. The over-arching quality management process, however, has both a deeper and broader scope. As well as evaluating iterations of a process, it is also necessary to evaluate separate phases and sub-processes, ideally each time they are applied, but at least according to an agreed schedule. Metadata generated by the different sub-processes themselves are also of interest as an input for process quality management. These evaluations can apply within a specific process, or across several processes that use common components.

42. Quality management also involves the evaluation of groups of statistical business processes, and can therefore identify potential duplication or gaps. All evaluations should result in feedback, which should be used to improve the relevant process, phase or sub-process, creating a quality loop.



43. Quality management can take several forms, including:

- Seeking and analysing user feedback;
- Reviewing operations and documenting lessons learned;
- Examining process metadata and other system metrics;
- Benchmarking or peer reviewing processes with other organizations.

44. Evaluation will normally take place within an organization-specific quality framework, and may therefore take different forms and deliver different results within different organizations. There is, however, general agreement amongst statistical organizations that quality should be defined according to the ISO 9000-2005 standard: “The degree to which a set of inherent characteristics fulfils requirements."[9]

45. Quality is a therefore multi-faceted, user-driven concept. The dimensions of quality that are considered most important depend on user perspectives, needs and priorities, which vary between processes and across groups of users. Several statistical organizations have developed lists of quality dimensions, which, for international organizations, are being harmonized under the leadership of the Committee for the Coordination of Statistical Activities (CCSA)[10].

[9] ISO 9000:2005, Quality management systems -- Fundamentals and vocabulary. International Organization for Standardization

[10] Current organization-specific quality frameworks, containing lists of dimensions, exist for: UNECE: http://unstats.un.org/unsd/accsub/2007docs-10th/SA-2007-14-Add1-ECERep.pdf, OECD: http://www.oecd.org/dataoecd/26/38/21687665.pdf, and Eurostat: http://epp.eurostat.ec.europa.eu/portal/page/portal/quality/documents/ess%20quality%20definition.pdf

46. The current multiplicity of quality frameworks enhances the importance of the benchmarking and peer review approaches to evaluation, and whilst these approaches are unlikely to be feasible for every iteration of every part of every statistical business process, they should be used in a systematic way according to a pre-determined schedule that allows for the review of all main parts of the process within a specified time period.

Metadata Management

47. Good metadata management is essential for the efficient operation of statistical business processes. Metadata are present in every phase, either created or carried forward from a previous phase. In the context of this model, the emphasis of the over-arching process of metadata management is on the creation and use of statistical metadata, though metadata on the different sub-processes themselves are also of interest, including as an input for quality management. The key challenge is to ensure that these metadata are captured as early as possible, and stored and transferred from phase to phase alongside the data they refer to. Metadata management strategy and systems are therefore vital to the operation of this model.

48. Part A of the Common Metadata Framework[11] identifies the following sixteen core principles for metadata management, all of which are intended to be covered in the over-arching Metadata Management process, and taken into the consideration when preparing the statistical metadata system (SMS) vision and global architecture, and when implementing the SMS. The principles can be presented in the following groups:

Metadata handling

i. ***Statistical Business Process Model***: Manage metadata with a focus on the overall statistical business process model.

ii. ***Active not passive***: Make metadata active to the greatest extent possible. Active metadata are metadata that drive other processes and actions. Treating metadata this way will ensure they are accurate and up-to-date.

iii. ***Reuse***: Reuse metadata where possible for statistical integration as well as efficiency reasons

iv. ***Versions***: Preserve history (old versions) of metadata.

[11] See: http://www.unece.org/stats/cmf/PartA.html

Metadata Authority

i. ***Registration***: Ensure the registration process (workflow) associated with each metadata element is well documented so there is clear identification of ownership, approval status, date of operation, etc.

ii. ***Single source***: Ensure that a single, authoritative source ('registration authority') for each metadata element exists.

iii. ***One entry/update***: Minimize errors by entering once and updating in one place.

iv. ***Standards variations***: Ensure that variations from standards are tightly managed/approved, documented and visible.

Relationship to Statistical Cycle / Processes

i. ***Integrity***: Make metadata-related work an integral part of business processes across the organization.

ii. ***Matching metadata***: Ensure that metadata presented to the end-users match the metadata that drove the business process or were created during the process.

iii. ***Describe flow***: Describe metadata flow with the statistical and business processes (alongside the data flow and business logic).

iv. ***Capture at source***: Capture metadata at their source, preferably automatically as a bi-product of other processes.

v. ***Exchange and use***: Exchange metadata and use them for informing both computer based processes and human interpretation. The infrastructure for exchange of data and associated metadata should be based on loosely coupled components, with a choice of standard exchange languages, such as XML.

Users

i. ***Identify users***: Ensure that users are clearly identified for all metadata processes, and that all metadata capturing will create value for them.

ii. ***Different formats***: The diversity of metadata is recognized and there are different views corresponding to the different uses of the data. Different users require different levels of detail. Metadata appear in different formats depending on the processes and goals for which they are produced and used.

iii. ***Availability***: Ensure that metadata are readily available and useable in the context of the users' information needs (whether an internal or external user).

VII. Other Uses of the GSBPM

As stated in the section on the purpose of the GSBPM, the original aim of the work to develop this model was that it should provide a basis for statistical organizations to agree on standard terminology to aid their discussions on developing statistical metadata systems and processes. However, as the model has developed, it has become increasingly apparent that it can be used for other purposes. This has been confirmed by Statistics New Zealand, who have either applied, or plan to apply their national version of the model in several different areas. The list below aims to highlight potential rather than recommended uses, and to inspire further ideas on how the GSBPM can be used in practice.

1. Harmonizing statistical computing architectures – The GSBPM can be seen as a model for an operational view of statistical computing architecture. It identifies the key components of the statistical business process, promotes standard terminology and standard ways of working across statistical business processes. The potential of the GSBPM as a model for statistical computing architectures will be evaluated further in the proposed European Union “ESSNet” project on a Common Reference Architecture[12] during 2009.

2. Facilitating the sharing of statistical software – Linked to the point above, the GSBPM defines the components of statistical processes in a way that not only encourages the sharing of software tools between statistical business processes, but also facilitates sharing between different statistical organizations that apply the model. It therefore provides an input to the “Sharing Advisory Board”, being created under the auspice of the UNECE / Eurostat / OECD Work Sessions on the Management of Statistical Information Systems[13].

3. Providing a basis for explaining the use of SDMX in a statistical organization in the Statistical Data and Metadata eXchange (SDMX) User Guide[14]. Chapter A2 of this user guide explores how SDMX applies to statistical work in the context of a business process model.

4. Providing a framework for process quality assessment and improvement – If a benchmarking approach to process quality assessment is to be successful, it is necessary to standardize processes as much as possible. The GSBPM provides a mechanism to facilitate this.

5. Better integrating work on statistical metadata and quality – Linked to the previous point, 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 terminology to describe the statistical business process.

6. Providing the underlying model for methodological standards frameworks - Methodological standards can be linked to the phase(s) or sub-process(es) they relate to and can then be classified and stored in a structure based on the GSBPM.

[12] http://circa.europa.eu/Public/irc/dsis/itsteer/library?l=/directors_13-14/proposal_essnetdoc/_EN_1.0_&a=d

[13] As proposed in the report of the MSIS Task Force on Software Sharing: http://www.unece.org/stats/documents/ece/ces/ge.50/2008/crp.2.e.doc

[14] See: http://sdmx.org/index.php?page_id=38, 2009 version

7. Providing a structure for storage of documents – As well as a framework for methodological standards, the GSBPM can also provide a structure for organizing and storing other documents within an organization, in conjunction with document management software tools. It can provide a basic document storage classification that allows clear links between documents and the parts of the statistical business process they relate to.

8. Providing a framework for building organizational capability – The GSBPM can be used to develop a framework assess the knowledge and capability that already exists within an organization, and to identify the gaps that need to be filled to improve operational efficiency.

9. Providing an input to high-level corporate work planning – The national business process model developed by Statistics New Zealand has been used as an input when preparing a high-level survey programme.

10. Developing a business process model repository – Statistics New Zealand has developed a database to store process modelling outputs and allow them to be linked to their statistical business process model. They also plan to develop a Business Process Modelling Community of Practice – i.e. a regular forum to build knowledge of process modelling, to promote the their business process model and increase understanding of it, and to discuss process modelling and models as enablers for process improvement.

11. Measuring operational costs – The GSBPM could conceivably be used as a basis for measuring the costs of different parts of the statistical business process. This, in turn, could help target development work to improve the efficiency of the parts of the process that are most costly.

12. Measuring system performance – Related to the point above on costs, the GSBPM can also be used to identify components that are not performing efficiently, that are duplicating each other unnecessarily, or that require replacing. Similarly it can identify gaps for which new components should be developed.

Annex – Glossary of Terms

Note – this short glossary just covers some of the key terms and abbreviations used in this paper. For a more comprehensive glossary of terms related to the statistical production process see the SDMX Metadata Common Vocabulary - http://sdmx.org/?page_id=11.

CMF - Common Metadata Framework: The need for a common metadata framework emerged from discussions in international forums. The joint UNECE / Eurostat / OECD Group on Statistical Metadata (METIS) is coordinating the work to develop this framework. The aim is to organize the vast pool of information about statistical metadata into a common framework for use by national and international statistical organizations. See http://www.unece.org/stats/cmf/.

Collection / Data collection – A systematic process of gathering data for official statistics. (Source SDMX Metadata Common Vocabulary, 2009)
For the purposes of this model, "collection" therefore includes obtaining data from administrative sources, as well as the more traditional data collection through surveys and censuses.

GSBPM – Generic Statistical Business Process Model: A flexible tool to describe and define the set of business processes needed to produce official statistics.

METIS - The joint UNECE / Eurostat / OECD Group on Statistical Metadata.

Over-arching process – Processes that apply throughout and across statistical business processes. They 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 organization

SDMX – A set of technical standards and content-oriented guidelines, together with an IT architecture and tools, to be used for the efficient exchange and sharing of statistical data and metadata.
(Source SDMX Metadata Common Vocabulary, 2009)

Statistical business process – The complete set of sub-processes needed to support statistical production.
(Source SDMX Metadata Common Vocabulary, 2009)

Statistical metadata system – A data processing system that uses, stores and produces statistical metadata.
(Source SDMX Metadata Common Vocabulary, 2009)