

Finer-level of Activities of Generic Statistical Business Process Model

GSBPM Task

(Version 1.0, August 2022)



This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>. If you re-use all or part of this work, please attribute it to the United Nations Economic Commission for Europe (UNECE), on behalf of the international statistical community.

Acknowledgement

This document was developed by the GSBPM “Tasks” Task Team under the Supporting Standards Group of the High-Level Group on the Modernisation of Official Statistics. The Task Team consisted of: Carlo Vaccari (Italy) as Coordinator, Elham Al-Johani, Eyad Alnoaim and Abdulaziz Alsuhaymi (Saudi Arabia), Miodrag Cerovina (Serbia), Natasa Cvetkovic (Serbia), Christophe Dzikowski (France), Edgardo Greising (International Labour Organization), Omurbek Ibraev (Kyrgyz Republic), Jesús Luján (INEGI Mexico), Kevin McCormack (Ireland), Juan Munoz Lopez (INEGI Mexico), Andrea Petres (Hungary), Giorgia Simeoni (Italy), Zoltan Vereczkei (Hungary), Florian Vucko (France), InKyung Choi and Wai Kit Si Tou (United Nations Economic Commission for Europe) as Secretariat.

Table of Contents

I. Introduction.....	5
II. Finer-level of Activities of Generic Statistical Business Process Model	7
Specify Needs Phase.....	7
1.1. Identify needs.....	7
1.2. Consult and confirm needs	7
1.3. Establish output objectives	8
1.4. Identify concepts.....	8
1.5. Check data availability	8
1.6. Prepare and submit business case	9
Design Phase.....	9
2.1. Design outputs	10
2.2. Design variable descriptions.....	10
2.3. Design collection	11
2.4. Design frame and sample.....	11
2.5. Design processing and analysis	12
2.6. Design production systems and workflow.....	12
Build Phase	13
3.1. Reuse or build collection instruments.....	13
3.2. Reuse or build processing and analysis components	14
3.3. Reuse or build dissemination components	14
3.4. Configure workflows	14
3.5. Test production systems	14
3.6. Test statistical business process.....	15
3.7. Finalise production systems.....	15
Collect Phase	16
4.1. Create frame and select sample	16
4.2. Set up collection	16
4.3. Run collection.....	17
4.4. Finalise collection.....	18
Process Phase.....	19
5.1. Integrate data	19
5.2. Classify and code.....	20
5.4. Edit and impute.....	21
5.5. Derive new variables and units.....	21
5.6. Calculate weights.....	22

5.7. Calculate aggregates	22
5.8. Finalise data files	22
Analyse Phase	23
6.1. Prepare draft outputs	23
6.2. Validate outputs	24
6.3. Interpret and explain outputs	24
6.4. Apply disclosure control	25
6.5. Finalise outputs	25
Disseminate Phase	26
7.1. Update output systems	26
7.2. Produce dissemination products	26
7.3. Manage release of dissemination products	27
7.5. Manage user support	27
Evaluate Phase	28
8.1. Gather evaluation inputs	28
8.2. Conduct evaluation	29
III. Suggestions to the GSBPM revision team	30

I. Introduction

1. The Generic Statistical Business Process Model (GSBPM) describes and defines the set of business processes needed to produce official statistics. It provides a standard framework and harmonised terminology to help statistical organisations to modernise their statistical production processes, as well as to share methods and components. The GSBPM can also be used for integrating data and metadata standards, as a template for process documentation, for harmonising statistical computing infrastructures, and to provide a framework for process quality assessment and improvement.
2. Since its release in 2008 (v1.0), GSBPM has gained wide adoption within official statistical community and become a de-facto standard for process model. The GSBPM comprises of eight phases of the process and sub-processes within each phase, which are the building blocks that can be used to design statistical business process. The activities and their descriptions in GSBPM remain at a high-level to keep the model generic, hence applicable for as many organizations as possible. While this generic nature of the model is attributed to its success, several organisations have added activities more detailed than GSBPM sub-processes (called “Task” henceforth) when they adopt GSBPM for local setting. A list of common tasks will provide better clarity on the activities to be conducted in different phases and sub-processes, thereby facilitating the implementation of GSBPM.
3. The purpose of this document is to compile examples from countries that have added lower-level activities to their national version of GSBPM, and to develop a list of common tasks for all GSBPM sub-processes without losing the generic nature of the model under a set of principles (e.g., coding, granularity, minimality, description format) with reference to the current version of the GSBPM (version 5.1).
4. The Task team has collected examples of finer-level activities from 9 countries: Australia, Azerbaijan, Costa Rica, France, Norway, Romania, Saudi Arabia, Serbia, and Spain (see [here](#) for the summary) and developed a proposal for the task level for all GSBPM sub-processes (see section II, highlighted in grey). The Task team has further prepared some suggestions to the GSBPM revision team (see section III).
5. The list of tasks was developed based mainly on the description of each sub-process in the GSBPM, by adding country examples as well as feedback received from the community. Each task uses the format of “Action verb + an object” as well as with hierarchical coding (e.g., 1.1.1, 1.1.2, etc.).
6. It should be noted that the GSBPM is not a rigid framework in which all sub-processes must be followed in a strict order, some sub-processes can be skipped and re-visited. Likewise, tasks in this document are meant to provide a set of building blocks that can be put together to construct a logical sequence for the production process, hence do not need to be followed linearly in the order they are presented, some tasks can be omitted and repeated.
7. During the discussion, many connections have been explored not only with the GSBPM revision team, but also with GSBPM Overarching processes (OPs) that are applied throughout the entire production process. The examples of OPs include management of quality, management of metadata and data management. Given that these activities take place in many moments in the production, the GSBPM does not make direct references to OPs in every sub-process where relevant, but instead, make explicit mention only where they are considered particularly important to highlight. Following the same approach in the GSBPM, this document does not include task-level

activities related to OPs everywhere they could take place, but only in a few sub-processes where they are considered to be particularly relevant to highlight (i.e., essential to the given sub-process from the statistical perspective that exclusion of such tasks would cause more confusion).

8. In the following months, a task team is expected to be formed to analyse the OPs in detail and decide the best way to connect OPs and sub-processes and tasks of the GSBPM.

II. Finer-level of Activities of Generic Statistical Business Process Model

9. With reference to the current version of the GSBPM (version 5.1), this section presents a list of tasks for all GSBPM sub-processes, with notes on example and explanation (highlighted in grey).

Specify Needs Phase

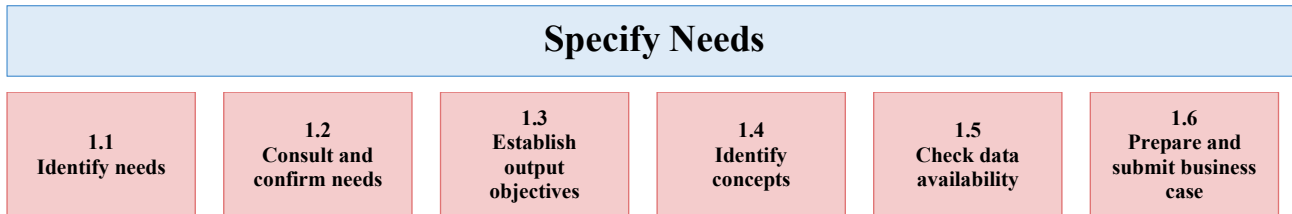


Figure 1. Specify needs phase and its sub-processes

10. This phase is triggered when a need for new statistics is identified or feedback about current statistics initiates a review. It includes all activities associated with engaging stakeholders to identify their detailed statistical needs (current or future), proposing high level solution options and preparing a business case to meet these needs.

11. The "Specify Needs" phase is broken down into six sub-processes (Figure 1), which are generally sequential, from left to right, but can also occur in parallel, and can be iterative. These sub-processes are:

1.1. Identify needs

12. This sub-process includes the initial investigation and identification of what statistics are needed and what is needed of the statistics. It may be triggered by a new information request or an environmental change such as a reduced budget. Action plans from evaluations of previous iterations of the process or from other processes might provide an input to this sub-process. It also includes consideration of practice amongst other (national and international) statistical organisations producing similar data and the methods used by those organisations.

1.1.1 Identify needs for information

1.1.2 Check practice (data and methods) amongst other statistical institutions

1.2. Consult and confirm needs

13. This sub-process focuses on consulting with the internal and external stakeholders and confirming in detail the needs for the statistics. A good understanding of user needs is required so that the statistical organisation knows not only what it is expected to deliver, but also when, how, and, perhaps most importantly, why. For the 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.

1.2.1 Identify relevant internal and external stakeholders for the given needs

1.2.2 Specify the needs in consultation with stakeholders and organizational strategy and priority

1.2.3 Confirm the needs with stakeholders

Note: 1.2.3: e.g., the nature and rationale of the needs, expected deliverables, timeline

1.3. Establish output objectives

14. This sub-process identifies the statistical output objectives that are required to meet the user needs identified in sub-process 1.2 (Consult and confirm needs). It includes agreeing the suitability of the proposed outputs and their quality measures with users. Legal frameworks (e.g. relating to confidentiality), and available resources are likely to be constraints when establishing output objectives.

1.3.1 Check legal framework and resource availability for the proposed outputs

1.3.2 Establish the statistical output objectives, including their quality measures

1.4. Identify concepts

15. This sub-process clarifies the required concepts to be measured from the point of view of the users. At this stage, the concepts identified might not align with existing statistical standards. This alignment, and the choice or definition of the statistical and other concepts and variables to be used, takes place in sub-process 2.2 (Design variable descriptions).

1.4.1 Consult existing statistical and non-statistical concepts

1.4.2 Identify all concepts to be measured by statistics

1.5. Check data availability

16. This sub-process checks whether current sources of data 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 sources of data, to:

- Determine whether they would be suitable for use for statistical purposes (e.g. the extent to which administrative concepts match data requirements, timeliness and quality of the data, security and continuity of data supply);
- Assess the division of responsibilities between data providers and the statistical organisation;
- Check necessary ICT resources (e.g. data storage, technology required to handle incoming data and data processing) as well as any formal agreements with data providers for accessing and sharing the data (e.g. formats, delivery, accompanying metadata and quality check).

17. When existing sources have been assessed, a strategy for filling any remaining gaps in the data requirement is prepared. This may include identifying possible partnerships with data holders. 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.

1.5.1 Determine internal data sources to meet the information needs

1.5.2 Evaluate the internal data sources if they meet the information needs

1.5.3 Determine the availability, suitability and access to external administrative or other non-statistical sources to meet the information needs

1.5.4 Assess legal framework and agreements with data holders for the collection of data from external sources

Note: 1.5.4: There are connections with the GAMS0 Corporate Support activity “Manage Data Suppliers” that covers the relationships with data suppliers. In GAMS0 the data transfer and the data sharing agreements are defined at a higher (corporate) level.

1.6. Prepare and submit business case

18. This sub-process documents the findings of the other sub-processes in this phase in the form of a business case to get approval to implement the new or modified statistical business process. Such a business case would need to conform to the requirements of the approval body, but would typically include elements such as:

- 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.

19. The business case describes options and makes recommendations. It may include the benefits, costs, deliverables, time frame, budget, required technical and human resources, risk assessment and impact on stakeholders for each option.

20. After the business case is prepared, it is submitted for approval to move to the next phase of the business process. At this sub-process, a “go”/“no go” decision is made. Typically, the business case is reviewed and formally approved or disapproved by the appropriate sponsors and governance committees.

1.6.1 Gather findings as a business case

1.6.2 Complete the business case

Note: 1.6.2: e.g., it may include benefits, costs, deliverables, time frame, budget, required technical and human resources, risk assessment and impact on stakeholders for each option and make recommendations.

Design Phase

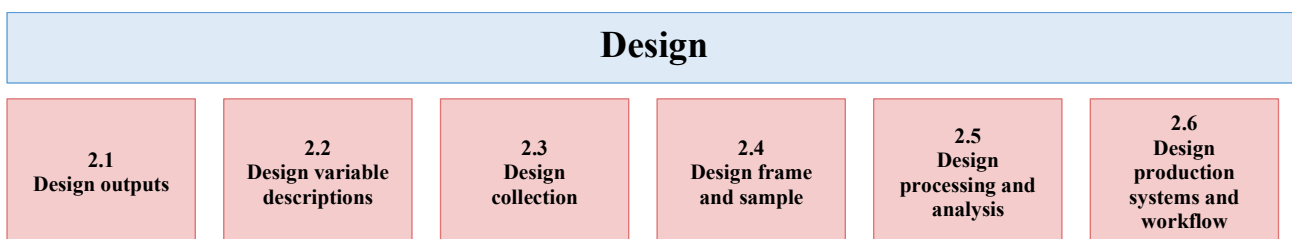


Figure 2. Design phase and its sub-processes

21. 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. It includes all the design elements needed to define or refine the statistical products or services identified in the business case. This phase specifies all relevant

metadata, ready for use later in the business process, as well as quality assurance procedures. For statistical outputs produced on a regular basis, this phase usually occurs for the first iteration and whenever improvement actions are identified in the “Evaluate” phase of a previous iteration.

22. Design activities make substantial use of international and national standards in order to reduce the length and cost of the design process, and enhance the comparability and usability of outputs. Organisations are encouraged to reuse or adapt design elements from existing processes, and to consider geospatial aspects of data in the design to enhance the usability and value of the statistical information. Additionally, outputs of design processes may form the basis for future standards at the organisational, national or international levels.

23. The “Design” phase is broken down into six sub-processes (Figure 2), 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

24. This sub-process contains the detailed design of the statistical outputs, products and services to be produced, including the related development work and preparation of the systems and tools used in the "Disseminate" phase. Processes governing access to any confidential outputs are also designed here. Outputs should be designed to follow existing standards wherever possible, so inputs to this process may include metadata from similar or previous collections (including extractions from statistical, administrative, geospatial and other non-statistical registers and databases), international standards, and information about practices in other statistical organisations from sub-process 1.1 (Identify needs). Outputs may also be designed in partnership with other interested bodies, particularly if they are considered to be joint outputs, or they will be disseminated by another organisation.

2.1.1 Design statistical concepts

2.1.2 Design statistical outputs, products and services and necessary components

2.1.3 Design dissemination procedure

2.1.4 Design systems and tools for dissemination

Note: 2.1.1: Sub-process 1.4 (Identify concepts) touches on concepts, but it does not seem that it is where we decide on concepts. Design of concepts is crucial for design of output.

Note: 2.1.2: e.g., it could include quality level, dissemination calendar, access to confidential outputs and examples on necessary components are table, chart and metadata.

2.2. Design variable descriptions

25. This sub-process defines the variables to be collected via the collection instrument, as well as any other variables that will be derived from them in sub-process 5.5 (Derive new variables and units), and any statistical or geospatial 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 collection), as the definition of the variables to be collected, and the choice of collection instruments may be inter-dependent to some degree. Preparation of metadata descriptions of collected and derived variables, statistical and geospatial classification is a necessary precondition for subsequent phases.

2.2.1 Design statistical units

2.2.2 Design variables to be collected or derived

2.2.3 Design classifications and code lists to be used

2.3. Design collection

26. This sub-process determines the most appropriate collection instruments and methods which may depend on the type of data collection (census, sample survey, or other), the collection unit type (enterprise, person, or other) and the available sources of data. The actual activities in this sub-process will vary according to the type of collection instrument required, which can include computer assisted interviewing, paper questionnaires, administrative registers (e.g. by using existing service interfaces), data transfer methods, web-scraping technologies as well as technology for geospatial data. Direct or indirect use of administrative data may be introduced in the data collection mode for either controlling survey data or assisting it when capturing survey information.

27. This sub-process includes the design of the collection instruments, questions and response templates (in conjunction with the variables and statistical classifications designed in sub-process 2.2 (Design variable descriptions)). It also includes the confirmation of any formal agreements. 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 provider management systems that are specific to this business process.

28. Where statistical organisations do not collect data directly (i.e. a third party controls the collection and processing of the data), this sub-process may include the design of mechanisms to monitor the data and the metadata to assess impacts of any change made by the third party.

2.3.1 Determine the most appropriate data source, collection mode and related method

2.3.2 Design the collection instruments

2.3.3 Formalise provision agreement

2.3.4 Design systems for data collection management and monitoring

Note: 2.3.1: While in Phase 1 we only identify possible sources, under this task we elaborate them in terms of concrete choice, sources' data and metadata, appropriate methods, etc. It includes the review of existing instruments and methods for reuse and repurpose. Also, data source and collection mode could be direct data collection (e.g., survey, census) and indirect data collection (e.g., web-scraping, administrative data harvest), whereas related methods are PAPI, CAPI, API transfer, file transfer for example.

Note: 2.3.2: e.g., questionnaire, data harvest API

Note: 2.3.4: The task is about the systems needed to manage and monitor the data collection process, like response rate monitoring, sending reminders, management of interviewers' paths, etc.

2.4. Design frame and sample

29. This sub-process only applies to processes which involve data collection based on sampling, such as through statistical surveys. It 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 for a sampling frame are administrative and statistical registers, censuses and information from other sample surveys. It may include geospatial data and classifications. 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 in sub-process 4.1 (Create frame and select sample), using the methodology specified in this sub-process.

2.4.1 Identify and specify the population of interest

2.4.2 Define population frame and how to build it

2.4.3 Design sampling methodology and sampling plan

2.5. Design processing and analysis

30. This sub-process designs the statistical processing methodology to be applied during the "Process" and "Analyse" phases. This can include among others, specification of routines and rules for coding, editing and imputation which may vary based on the mode of data collection and source of data. This sub-process also includes design of specifications for data integration from multiple data sources, validation of data and estimation. Statistical disclosure control methods are also designed here if they are specific to this business process.

2.5.1 Design specifications for data integration

2.5.2 Design specifications for data coding

2.5.3 Design specifications for data validation

2.5.4 Design specifications for data editing and imputation

2.5.5 Design specifications for calculation of weights

2.5.6 Design specifications for calculation of aggregates

2.5.7 Design methodology of data confidentiality protection

2.5.8 Design procedure for production of final outputs

2.5.9 Design systems and tools for data processing and analysis

Note: It is important to design relevant metadata and quality assurance procedures.

2.6. Design production systems and workflow

31. This sub-process determines the workflow from data collection to dissemination, taking an overview of all the processes required within the whole production process and ensuring that they fit together efficiently with no gaps or redundancies. Various systems and databases are needed throughout the process. The GSBPM can be used as the basis of the business architecture layer when a statistical organisation has an existing enterprise architecture in place. The design might be adjusted to fit the organization. A general principle is to reuse processes and technology across many statistical business processes, so existing production solutions (e.g. services, systems and databases) should be examined first, to determine whether they are fit for purpose for this specific production 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.

- 2.6.1 Design workflow from collection to dissemination
- 2.6.2 Design roles, responsibilities and schedule
- 2.6.3 Design systems and tools for workflow management

Note: 2.6.1: If we can assume that workflow within each phase is designed in the previous sub-processes to a certain extent, this task concerns putting those together from phase 4 to phase 7, identifying any gaps, having a final look at the whole workflow.

Build Phase

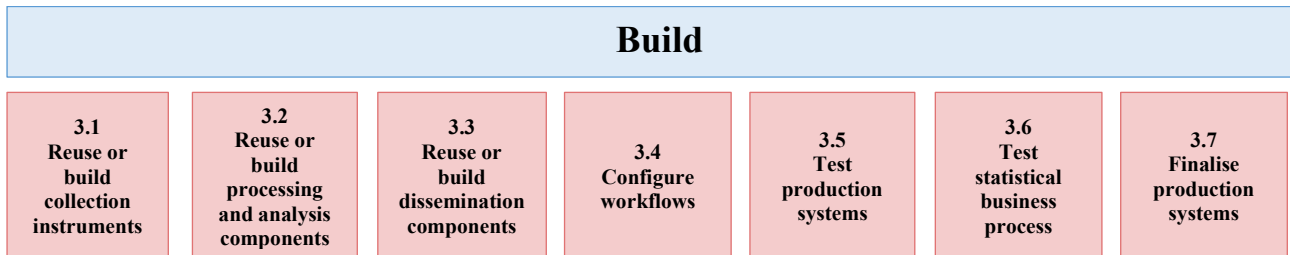


Figure 3. Build phase and its sub-processes

32. This phase builds and tests the production solution to the point where it is ready for use in the "live" environment. The outputs of the "Design" phase are assembled and configured in this phase to create the complete operational environment to run the process. New services are built by exception, created in response to gaps in the existing catalogue of services sourced from within the organisation and externally. These new services are constructed to be broadly reusable in alignment with the business architecture of the organisation where possible.

33. For statistical outputs produced on a regular basis, this phase usually occurs for the first iteration, following a review or a change in methodology or technology, rather than for every iteration.

34. The "Build" phase is broken down into seven sub-processes (Figure 3), which are generally sequential, from left to right, but can also occur in parallel, and can be iterative. The first three sub-processes are concerned with the development and improvement of systems used in collection, processing, analysis and dissemination of data. The last four sub-processes focus on the end-to-end process. These sub-processes are:

3.1. Reuse or build collection instruments

35. This sub-process describes the activities to build and reuse the collection instruments to be used during the "Collect" phase. The collection instruments are 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 web services). Collection instruments may also be data extraction routines used to gather data from existing statistical or administrative registers (e.g. by using existing service interfaces). This sub-process also includes preparing and testing the contents and functioning of that collection instrument (e.g. cognitive testing of the questions in a questionnaire). It is recommended to consider the direct connection of collection instruments to a metadata system, so that metadata can be more easily captured in the collection phase. Connecting 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 for calculating and analysing process quality indicators.

3.1.1 Make the necessary adaptations and changes to reuse existing collection instruments to meet the requirements

3.1.2 Build the collection instruments based on the design specifications to capture data, metadata and other relevant information

3.1.3 Test the contents and functioning of collection instruments

3.2. Reuse or build processing and analysis components

36. This sub-process describes the activities to reuse existing components or build new components needed for the “Process” and “Analyse” phases, as designed in the "Design" phase. Services may include dashboard functions and features, information services, transformation functions, geospatial data services, workflow frameworks, provider and metadata management services.

3.2.1 Make the necessary adaptations and changes to reuse and repurpose existing processing and analysis components to meet the requirements

3.2.2 Build the processing and analysis instruments

3.2.3 Test the contents and functioning of processing and analysis instruments

3.3. Reuse or build dissemination components

37. This sub-process describes the activities to build new components or reuse existing components needed for the dissemination of statistical products as designed in sub-process 2.1 (Design outputs). All types of dissemination components are included, from those that produce traditional paper publications to those that provide web services, (linked) open data outputs, geospatial statistics, maps, or access to microdata.

3.3.1 Make the necessary adaptations and changes to reuse and repurpose existing dissemination instruments to fit the latest requirements

3.3.2 Build the dissemination instruments

3.3.3 Test the contents and functioning of dissemination instruments

3.4. Configure workflows

38. This sub-process configures the workflow, systems and transformations used within the business processes, from data collection through to dissemination. In this sub-process, the workflow is configured based on the design created in sub-process 2.6 (Design production systems and workflows). This could include modifying a standardised workflow for a specific purpose, assembling the workflows for the different phases together (possibly with a workflow/business process management system) and configuring systems accordingly.

3.4.1 Assemble production systems

3.4.2 Configure production workflows within and between the phases

3.5. Test production systems

39. This sub-process is concerned with the testing of assembled and configured services and related workflows. 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 and services could logically be linked with sub-process 3.1, 3.2 and 3.3, this sub-process also includes testing of interactions between assembled and configured services, and ensuring that the whole production solution works in a coherent way.

3.5.1 Test service components of the collect phase

3.5.2 Test service components of the process and analysis phases

3.5.3 Test service components of the dissemination phase

3.5.4 Test workflows

3.5.5 Test interactions between service components and services

3.6. Test statistical business process

40. 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 the 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 collection 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.

3.6.1 Prepare testing/piloting of the whole statistical business process

3.6.2 Conduct testing/piloting of the whole statistical business process

3.6.3 Evaluate the testing/piloting of the whole statistical business process

3.6.4 Define actions based on the testing/piloting of the whole statistical business process

Note: This sub-process is basically a small PDCA cycle and there are clear logical connections to Phase 8 (e.g., connections between 3.6.2 and 8.2, 3.6.4 and 8.3).

3.7. Finalise production systems

41. This sub-process includes the activities to put the assembled and configured processes and services, including modified and newly-created services, into production ready for use. The activities include:

- Producing documentation about the process components, including technical documentation and user manuals;
- Training the 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 3.5 (Test production system)).

3.7.1 Produce documentation about the process components, including technical documentation and user manuals

3.7.2 Integrate the components into the production environment and ensure they work as expected

3.7.3 Train users and administrators of the system

Collect Phase

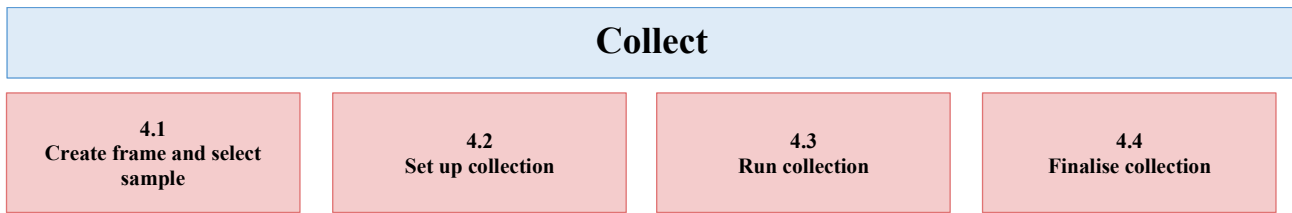


Figure 4. Collect phase and its sub-processes

42. This phase collects or gathers all necessary information (e.g. data, metadata and paradata), using different collection modes (e.g. acquisition, collection, extraction, transfer), 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.

43. The "Collect" phase is broken down into four sub-processes (Figure 4), which are generally sequential, from left to right, but can also occur in parallel, and can be iterative. These sub-processes are:

4.1. Create frame and select sample

44. 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). It also includes the coordination of samples between instances of the same business process (e.g. to manage overlap or rotation), and between different processes using a common frame or register (e.g. to manage overlap or to spread response burden). Quality assurance and approval 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 sources (e.g. administrative registers, web sites) as such processes generally create frames from the available data and then follow a census approach. Variables from administrative and other non-statistical sources of data can be used as auxiliary variables in the construction of sampling design.

4.1.1 Create frame

4.1.2 Select sample

Note: 4.1.1: The "frame" may also be a list of websites to go through in a web-scraping process.

4.2. Set up collection

45. This sub-process ensures that the people, processes and technology (e.g. web-based applications, GPS system) 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;
- Training system 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);
- 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, APIs, web scraping tools);
- Providing information for respondents (e.g. drafting letters or brochures explaining the purpose of the survey, notifying respondents when online reporting instruments will be made available);
- Translating of materials (e.g. into the different languages spoken or used in the country).

46. For non-survey sources, this sub-process ensures that the necessary processes, systems and confidentiality procedures are in place, to receive or extract the necessary information from the source. This includes:

- Evaluating requests to acquire the data and logging the request in a centralised inventory;
- Initiating contacts with organisations providing the data, and sending an introductory package with details on the process of acquiring the data;
- Checking detailed information about files and metadata with the data provider and receiving a test file to assess if data are fit for use;
- Arranging secure channels for the transmission of the data.

4.2.1 Set up the strategy and schedule for data collection

4.2.2 Set up software, physical equipment and system required to collect data

4.2.3 Prepare necessary human resources

Note: 4.2.1: It includes contacting data providers (e.g., administrative registers owners, web sites administrators, etc.)

Note: 4.2.2: e.g., security setting, configuration, printing paper questionnaire, distribution

Note: 4.2.3: Here the training is present because we are talking about people external to NSO (training for internal staff is in "Overarching processes").

4.3. Run collection

47. This sub-process is where the collection is implemented. The different collection instruments are used to collect or gather the information which may include raw microdata or aggregates produced at the source, as well as any associated metadata. It can include the initial contact with providers and any subsequent follow-up or reminder actions. It may include manual data entry at the point of contact, or fieldwork management, depending on the source and collection mode. It records when and how providers were contacted, and whether they have responded. Depending on the

geographical frame and the technology used, geo-coding¹ may need to be done at the same time as collection of the data by using inputs from GPS systems, putting a mark on a map, etc. This sub-process also includes the management of the providers involved in the current collection, ensuring that the relationship between the statistical organisation and data providers remains positive, and recording and responding to comments, queries and complaints. Proper communication with reporting units and minimisation of the number of non-respondents contribute significantly to a higher quality of the collected data.

48. For administrative, geographical or other non-statistical data, the provider is either contacted to send the information or sends it as scheduled. This process may be time consuming and might require follow-ups to ensure that data are provided according to the agreements. In the case where the data are published under an Open Data license and exist in machine-readable form, they may be freely accessed and used.

49. This sub-process may also include the monitoring of data collection and making any necessary changes to improve data quality. This includes generating reports, visualising and adjusting the acquisition process to ensure the data are fit for use. When the collection meets its targets, it is closed and a report on the collection is produced. Some basic checks of the structure and integrity of the information received may take place within this sub-process, (e.g. checking that files are in the right format and contain the expected fields).

4.3.1 Launch data collection

4.3.2 Monitor and manage data collection

4.3.3 Receive data

4.3.4 Check the structure and integrity of the data received

Note: 4.3.1: e.g., core/basic data, contact data, paradata

Note: 4.3.2: e.g., assess measures for data collection, follow up non-responses, fieldwork management, management of data providers.

Note: 4.3.4: e.g., checking that files are in the right format and contain the expected fields

4.4. Finalise collection

50. This sub-process includes loading the collected data and metadata into a suitable electronic environment for further processing. It may include manual or automatic data capture, 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 metadata and paradata associated with collection to ensure the collection activities have met requirements. In cases where there is a physical collection instrument, such as a paper questionnaire, which is not needed for further processing, this sub-process manages the archiving of that material. When the collection instrument uses software such as an API or an app, this sub-process also includes the versioning and archiving of these.

4.4.1 Convert data into a suitable format

4.4.2 Transfer them to the system for processing

¹ In this document, geo-coding is defined as “assigning codes related to geographical places” while geo-referencing is defined more generally as “relating data with where the fact happens”

4.4.3 Analyse metadata and paradata to ensure the collection activities have met requirements

4.4.4 Archive or version the collection instrument

Note: 4.4.2: e.g., basic structural validation performed, like controlled vocabularies (codes) compliance

Process Phase

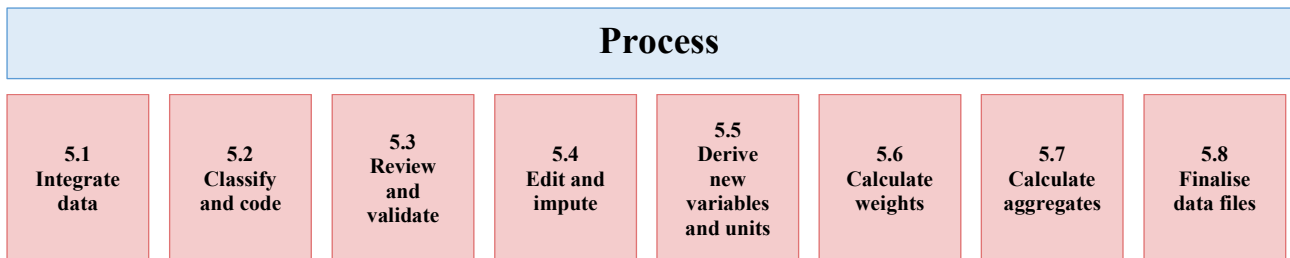


Figure 5. Process phase and its sub-processes

51. This phase describes the processing of input data and their preparation for analysis. It is made up of sub-processes that integrate, classify, check, clean, and transform input data, so that they can be analysed and disseminated as statistical outputs. 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).

52. 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. Sometimes the estimates being processed might be already published aggregates (undertaken according to a Revision Policy).

53. Activities within the "Process" and "Analyse" phases may also 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.

54. The "Process" phase is broken down into eight sub-processes (Figure 5), 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

55. This sub-process integrates data from one or more sources. It is where the results of sub-processes in the "Collect" phase are combined. The input data can be from a mixture of external or internal sources, and a variety of the collection instruments, including extracts of administrative and other non-statistical data sources. Administrative data or other non-statistical sources of data can substitute for all or some of the variables directly collected from survey. This sub-process also includes harmonising or creating new figures that agree between sources of data. The result is a set of linked data. Data integration can include:

- Combining data from multiple sources, as part of the creation of integrated statistics such as national accounts;

- Combining geospatial data and statistical data or other non-statistical data;
- Data pooling, with the aim of increasing the effective number of observations of some phenomena;
- Matching or record linkage routines, with the aim of linking micro or macro data from different sources;
- Data fusion - integration followed by reduction or replacement;
- Prioritising, when two or more sources contain data for the same variable, with potentially different values.

56. 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 de-identified, that is stripped of identifiers such as name and address, to help to protect confidentiality.

5.1.1 Identify and retrieve data sets to be integrated

5.1.2 Select integration methods based on design

5.1.3 Integrate data sets

Note: 5.1.1: It covers traditional data sources, administrative data, and other data sources, e.g., big data

Note: 5.1.2: While the data integration methods were designed in Phase 2, we select the methods and set the parameters for the linking algorithms here.

5.2. Classify and code

57. 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 statistical classification to facilitate data capture and processing. Some questions have coded response categories on the questionnaires or administrative source of data, others are coded after collection using an automated process (which may apply machine learning techniques) or an interactive, manual process.

5.2.1 Select variables to be coded and classifications or other code lists to be used

5.2.2 Select classification methods and coding algorithms based on design

5.2.3 Run classification and coding

Note: 5.2.2: It could be automatic and/or manual coding. In general, automatic coding is usually employed first, followed by manual coding for more specific issues. It is important to manage applying different coding to those that were originally planned, as well as store the mappings to new codes with versioning of code lists.

5.3. Review and validate

58. This sub-process examines data to identify 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 pre-defined edit rules, usually in a set order. It may flag data for automatic or manual inspection or editing. Reviewing and validating can apply to data

from any type of source, before and after integration, as well as imputed data from sub-process 5.4 (Edit and impute). Whilst validation is treated as part of the “Process” phase, in practice, some elements of validation may occur alongside collection activities, particularly for modes such as computer assisted collection. Whilst this sub-process is concerned with detection and localisation of actual or potential errors, any correction activities that actually change the data is done in sub-process 5.4 (Edit and impute).

5.3.1 Select data set to be validated

5.3.2 Select validation methods based on design

5.3.3 Run validation

Note: 5.3.3: The review examines data to identify potential problems, errors and discrepancies. While this sub-process is concerned with detection and localisation of errors, the correction of data is done in sub-process 5.4

5.4. Edit and impute²

59. Where data are considered incorrect, missing, unreliable or outdated, new values may be inserted or outdated data may be removed in this sub-process. The terms editing and imputation cover a variety of methods to do this, often using a rule-based approach. Specific steps typically include:

- Determining whether to add or change data;
- Selecting the method to be used;
- Adding/changing data values;
- Writing the new data values back to the data set, and flagging them as changed;
- Producing metadata on the editing and imputation process.

5.4.1 Determine whether to add or change data

5.4.2 Select editing and imputation methods based on design

5.4.3 Run editing and imputation and update the data set and associated element in data structure

5.5. Derive new variables and units

60. This sub-process derives data for variables and 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, or applying different model assumptions. This activity 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 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.

5.5.1 Select variables and units to be used for derivation

5.5.2 Select derivation methods based on design

5.5.3. Derive new variables and units

² For a more detailed consideration of the different steps and flows involved in data editing, please see UNECE Statistics Wikis - Generic Statistical Data Editing Models (<https://statswiki.unece.org/display/sde/GSDiEMs>)

5.5.4. Produce processing report

5.6. Calculate weights

61. This sub-process creates weights for unit data records according to the methodology developed in sub-process 2.5 (Design processing and analysis). For example, weights can be used to "gross-up" data to make them representative of the target population (e.g. for sample surveys or extracts of scanner data), or to adjust for non-response in total enumerations. In other situations, variables may need weighting for normalisation purposes. It may also include weight correction for benchmarking indicators (e.g. known population totals).

5.6.1 Calculate sampling weights

5.6.2 Correct calculated weights

5.6.3 Validate weights

Note: 5.6.1: Weights for unit data records according to the methodology developed in sub-process 2.5.

Note: 5.6.2: e.g., calibration, re-weighting

Note: 5.6.3: It is necessary to check the quality of the weights. For example, are there any extreme weights? Does the calibration step require excessively large corrections to the weights? If this is the case, it is possible that the methodological steps of point 5.6.1 needs to be improved and that we should return to this stage.

5.7. Calculate aggregates

62. This sub-process creates aggregate data and population totals from microdata or lower-level aggregates. It includes summing data for records sharing certain characteristics (e.g. aggregation of data by demographic or geographic classifications), determining measures of average and dispersion, and applying weights from sub-process 5.6 (Calculate weights) to derive appropriate totals. In the case of statistical outputs which use sample surveys, sampling errors corresponding to relevant aggregates may also be calculated in this sub-process.

5.7.1 Calculate simple aggregations

5.7.2 Calculate more complex statistical functions

5.7.3 Calculate sampling errors if needed

5.8. Finalise data files

63. This sub-process brings together the results of the other sub-processes in this phase in a data file (usually macro-data), which is used as the input to the "Analyse" phase. 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.

5.8.1 Bring the files together and consolidate them

5.8.2 Update data file references (or catalogue) as necessary

5.8.3 Convert files to appropriate format

Note: This finalisation sub-process aims to bring together the results of the other sub-processes and make the files ready for downstream consumption.

Note: 5.8.1: e.g., putting the individual data sets in database, data hub, data lake

Note: 5.8.2: The data catalogue is an inventory of the statistical data that the organisation has in its databases and other files. It provides mappings and other mechanisms to help in the location of the statistical data. A well-structured data catalogue contains structural and reference metadata that helps to satisfy needs of initiatives like open data, and adds elements to evaluate the quality of the information.

Note: 5.8.3 e.g., SDMX

Analyse Phase

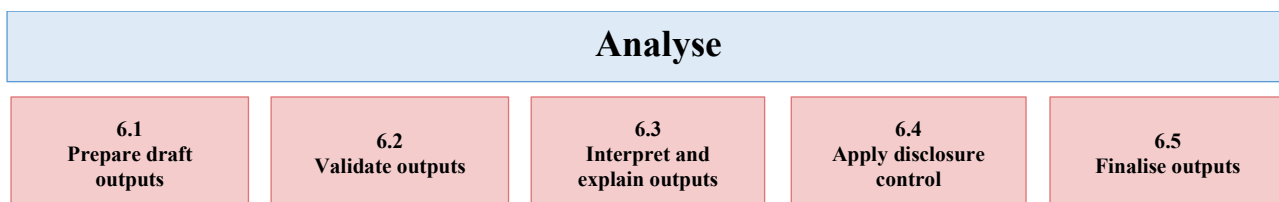


Figure 6. Analyse phase and its sub-processes

64. In this phase, statistical outputs are produced and examined in detail. It includes preparing statistical content (including commentary, technical notes, etc.), and ensuring outputs are “fit for purpose” prior to dissemination to users. This phase also includes the sub-processes and activities that enable statistical analysts to understand the data and the statistics produced. The outputs of this phase could also be used as an input to other sub-processes (e.g. analysis of new sources as input to the “Design” phase). 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.

65. The "Analyse" phase is broken down into five sub-processes (Figure 6), which are generally sequential, from left to right, but can also occur in parallel, and can be iterative. These sub-processes are:

6.1. Prepare draft outputs

66. This sub-process is where the data from sub-processes 5.7 (Calculate aggregates) and 5.8 (Finalise data files) are transformed into statistical outputs such as indexes, seasonally adjusted statistics, e.g. trend, cycle, seasonal and irregular components, accessibility measures, etc., as well as the recording of quality characteristics such as coefficients of variation. The preparation of maps, GIS outputs and geo-statistical services can be included to maximise the value and capacity to analyse the statistical information.

6.1.1 Transform data into draft statistical outputs

6.1.2 Record quality characteristics

Note: 6.1.1: e.g., indexes, seasonally adjusted statistics, such as trend, cycle, seasonal and irregular

components, accessibility measures, etc. A more elaborated process could include the following actions: (i) produce elements that are missing but necessary to transform data into products; (ii) merge the data with the representation elements that will turn the data into products; and (iii) make sure the draft outputs correspond to output expectations in terms of content, representation, format, etc.

Note: 6.1.2: Major quality characteristics of statistical output include relevance, accuracy and reliability, timeliness and punctuality, coherence and comparability, accessibility and clarity, as adopted by the European Statistical System ([reference](#)). A good reference to Quality dimensions can be found at [this StatCan link](#)

6.2. Validate outputs

67. 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 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);
- Checking that the associated metadata, paradata and quality indicators are present and in line with expectations;
- Checking geospatial consistency of the data;
- 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.

6.2.1 Validate internal consistency of the outputs

6.2.2 Validate coherence of outputs with previous cycles, other sources and domain experts

6.2.3 Validate the associated metadata, paradata, and quality indicators are present and in line with expectations

Note: 6.2.1: In accordance with quality framework and expectations expressed. Also perform macro editing.

6.3. Interpret and explain outputs

68. This sub-process is where the in-depth understanding of the outputs is gained by statisticians. They use that understanding to interpret and explain the statistics 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 such as time-series analysis, consistency and comparability analysis, revision analysis (analysis of the differences between preliminary and revised estimates), analysis of asymmetries (discrepancies in mirror statistics), etc.

6.3.1 Carry out in-depth statistical analysis

6.3.2 Interpret and explain outputs by assessing how well they reflect their initial expectations

Note: 6.3.1: e.g., time-series analysis, consistency and comparability analysis, revision analysis, analysis of asymmetries, etc

Note: 6.3.2: How the outputs were produced (i.e., method / technique used) should also be explained.

6.4. Apply disclosure control

69. This sub-process ensures that the data (and metadata) to be disseminated do not breach the appropriate rules on confidentiality according to either organisation policies and rules, or to the process-specific methodology created in sub-process 2.5 (Design processing and analysis). This may include checks for primary and secondary disclosure, as well as the application of data suppression or perturbation techniques and output checking. The degree and method of statistical disclosure control may vary for different types of outputs. For example, the approach used for microdata sets for research purposes will be different to that for published tables, finalised outputs of geospatial statistics or visualisations on maps.

6.4.1 Assess the fulfilment of confidentiality rules

6.4.2 Apply statistical disclosure control methods and produce safe outputs

6.4.3 Undertake output checking activity

Note: 6.4.2: The degree and method of statistical disclosure control may vary for different types of outputs. For example, the approach used for microdata sets for research purposes will be different to that for published tables, finalised outputs of geospatial statistics or visualisations on maps.

6.5. Finalise outputs

70. 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, commentary, technical notes, briefings, measures of uncertainty and any other necessary metadata;
- Producing the supporting internal documents;
- Conducting pre-release discussion with appropriate internal subject matter experts;
- Translating the statistical outputs in countries with multilingual dissemination;
- Approving the statistical content for release.

6.5.1 Complete consistency checks

6.5.2 Determine the level of dissemination and draft any caveats that should be issued

6.5.3 Ensure that the output produced can be disseminated and complies with commitments

6.5.4 Collate additional information such as interpretations of results, comments, technical notes, guidance notes related to disseminated products

6.5.5 Approve content of the publication

6.5.6 Finalise the description of the required metadata

Disseminate Phase

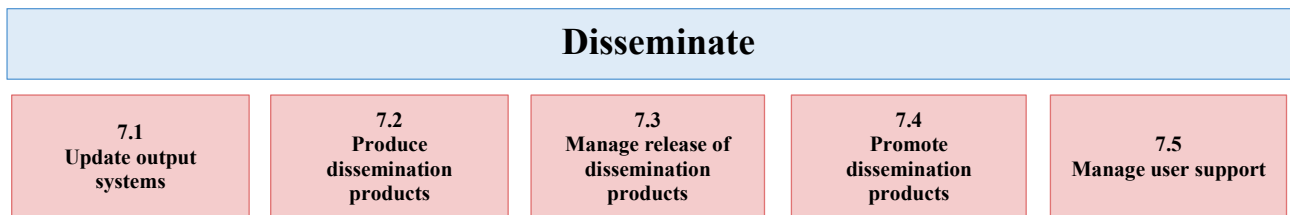


Figure 7. Disseminate phase and its sub-processes

71. This phase manages the release of the statistical products to users. It includes all activities associated with assembling and releasing a range of static and dynamic products via a range of channels. These activities support users to access and use the products released by the statistical organisation. For statistical products produced regularly, this phase occurs in each iteration.

72. The “Disseminate” phase is broken down into five sub-processes (Figure 7), 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

73. This sub-process manages the update of systems (e.g. databases) where data and metadata are stored ready for dissemination purposes, including:

- Formatting data and metadata ready to be put into output systems;
- Loading data and metadata into output systems;
- Ensuring data are linked to the relevant metadata.

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

7.1.1 Format data and metadata ready to be put into dissemination systems

7.1.2 Load data and metadata into output systems

7.1.3 Link data to the relevant metadata

Note: 7.1.1: It could also include other information exchange systems.

7.2. Produce dissemination products

75. This sub-process produces the dissemination products, as previously designed in sub-process 2.1 (Design outputs), to meet user needs. They could include printed publications, press releases and websites. The products can take many forms including interactive graphics, tables, maps, public-use microdata sets, linked open data and downloadable files. Typical steps include:

- Preparing the product components (explanatory texts, tables, charts, maps, quality statements etc.);
- Assembling the components into products;
- Editing the products and checking that they meet publication standards.

7.2.1 Produce texts, graphics, files, and other materials for dissemination

7.2.2 Assemble all elements for dissemination

7.2.3 Edit the products to ensure that they meet publication standards

7.3. Manage release of dissemination products

76. 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, and managing access to confidential data by authorised user groups, such as researchers. Sometimes an organisation may need to retract a product, for example, if an error is discovered. This is also included in this sub-process.

7.3.1 Collect release notices, publications, notes, etc. to be released

7.3.2 Manage release of products to different user groups access to data

7.3.3 Retract a product if there is any error

7.4. Promote dissemination products

77. Whilst marketing in general can be considered to be an overarching 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 websites, wikis and blogs to facilitate the process of communicating statistical information to users.

7.4.1 Prepare promotional materials

7.4.2 Target potential users and promote the products via different channels

Note: 7.4.2 The use of customer relationship management tools could help target potential users of the products. Depending on the target audience, different channels (e.g., websites, wikis, blogs, etc.) could facilitate the process of communicating statistical information to users.

7.5. Manage user support

78. This sub-process ensures that user queries and requests for services such as microdata access are recorded, and that responses are provided within agreed deadlines. These queries and requests should be regularly reviewed to provide an input to the overarching quality management process, as they can indicate new or changing user needs. Replies to user requests can also be used to populate a knowledge database or a “Frequently Asked Questions” page, that is made publicly available, thus reducing the burden of replying to repeated and/or similar requests from external users. This sub-process also includes managing support to any partner organisations involved in disseminating the products.

7.5.1 Respond to user queries

7.5.2 Respond to user requests for services

7.5.3 Store information on the requests received in a knowledge database

Note: 7.5.1: e.g., to clarify the details of the data, where to find the data, etc.

Note: 7.5.2: Including microdata and metadata access.

Note: 7.5.3: e.g., an internal or external “Frequently Asked Questions” page

Evaluate Phase

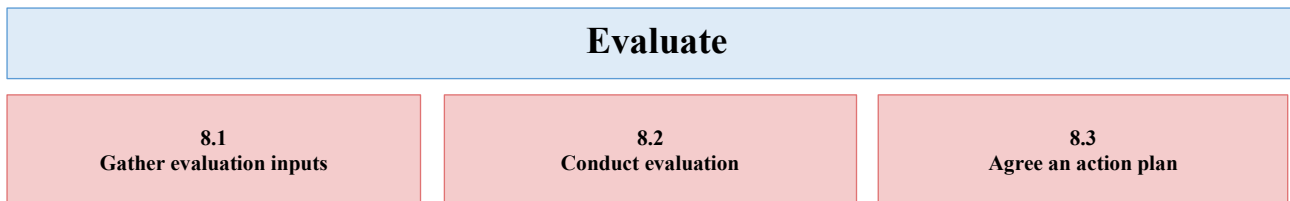


Figure 8. Evaluate phase and its sub-processes

79. This phase manages the evaluation of a specific instance of a statistical business process, as opposed to the more general overarching process of statistical quality management described in Section VI (Overarching Processes). It can take place at the end of the instance of the process, but can also be done on an ongoing basis during the statistical production process. It relies on inputs gathered throughout the different phases. It includes evaluating the success of a specific instance of the statistical business process, drawing on a range of quantitative and qualitative inputs, and identifying and prioritising potential improvements.

80. 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 might 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 the “Specify Needs” phase, or from some later phase (often the “Collect” phase).

81. The “Evaluate” phase is broken down into three sub-processes (Figure 8), which are generally sequential, from left to right, but can also occur in parallel, and can be iterative. These sub-processes are:

8.1. Gather evaluation inputs

82. Evaluation material can be produced in any other phase or sub-process. It may take many forms, including feedback from users, process metadata (paradata), 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, compiles quality indicators and makes them available for the person or team producing the evaluation. The collection of some of these evaluation materials can be automated and take place in a continuous way throughout the whole process, as defined by the quality framework (see Quality Management in Section VI). On the other hand, for the evaluation of certain processes it can be necessary to perform specific activities such as small surveys, (e.g. post-enumeration surveys, re-interview studies, survey on effectiveness of dissemination).

8.1.1 Gather all available material for evaluation

8.1.2 Produce additional evaluation material

Note: 8.1.1: e.g., feedback from users, process metadata, paradata, staff suggestions, reports of progress against an action plan agreed during a previous iteration

Note: 8.1.2: For evaluation materials that do not exist, some additional activities may be needed to collect the necessary information. The activities include small surveys, e.g., post-enumeration surveys, re-interview studies, survey on effectiveness of dissemination.

8.2. Conduct evaluation

83. This sub-process analyses the evaluation inputs, compares them to the expected/target benchmarking results (when available), and synthesises them into an evaluation report or control dashboard. The evaluation can take place at the end of the whole process (ex-post evaluation) for selected activities, during its execution in a continuous way, or throughout the process, thus allowing for quick fixes or continuous improvement. The resulting report should note any quality issues specific to this iteration of the statistical business process as well as highlight any deviation of performance metrics from expected values, 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.

8.2.1 Analyse evaluation inputs

8.2.2 Compare evaluation inputs to the expected/target benchmarking results

8.2.3 Prepare evaluation report

Note: 8.2.1: A standard checklist (e.g., SDMX) could be used for design as a part of evaluation.

8.3. Agree an action plan

84. 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.

8.3.1 Prepare action plan

8.3.2 Prepare mechanism to monitor its implementation

Note: This is the end of the process but it should feed as input for the next wave of business processes (e.g., Phase 1).

III. Suggestions to the GSBPM revision team

85. The Task team has further prepared some suggestions to the GSBPM revision team (also available at this [wiki page](#)):

Sub-process 1.3:

We are talking more about "accessing data" instead of "transmitting data". This change of perspective should be reflected in the text of the GSBPM (maybe also generate additional tasks somewhere?)

Sub-process 1.6:

How to handle decision points in the statistical business process model, such as approval of business cases of action plans (1.6 & 8.3 and possibly even more in the process)?

Sub-process 2.1:

While Phase 7 (Disseminate) has many activities about release and dissemination, nothing about this is explicitly designed anywhere in Phase 2. It should be included as a task in sub-process 2.1, also in GSBPM explicitly.

Sub-process 2.2:

"Statistical unit" is not discussed anywhere, but this is very important when we design variables, this needs to be in GSBPM explicitly. Also, this task comprises design of delineation of "statistical units" (observation and analytical units) from so called "units of collection" (administrative, legal etc. units).

Sub-process 2.4:

The wording "population frame" is used to also cover non-sampling scenarios. Although GSBPM sub-process 2.4 says "This sub-process only applies to processes which involve data collection based on sampling, such as through statistical surveys.", even if we use administrative registers (non-direct collection), we think there could be some kind of "filtering" of the entire register depending on the target population, also, for census, we will need some frame for the enumeration.

Sub-process 4.4:

The GSBPM needs to systematically analyse and define what is happening in the "finalise" sub-processes (Phases: 3, 4, 5, 6).

Sub-processes 5.3 & 5.4:

There are some modern approaches/tools that could detect errors and edit them at the same time, which blur the distinction between sub-processes 5.3 and 5.4.

Sub-process 6.4:

GSBPM does not mention ethical considerations here, even though there are. The GSBPM revision team should analyse how the ethical considerations are affecting the statistical work and how those perspectives can be built into the model.

Sub-process 7.1:

The word "system" is a bit confusing. Components or modules may be more appropriate.

Sub-process 7.2:

It is unclear when dissemination actually takes place as 7.2 is about producing dissemination products and 7.3 is on managing release of dissemination products.

Sub-process 7.3:

A suggestion from MWW2022 is that a version number could be added to the product.

It is important to explain and amend the error after dissemination, instead of just retracting a product.

It is noted that GSBPM does not have much discussion on dissemination planning.

Sub-process 7.4:

Doesn't the core process of data production end with the release of the products? Would product promotion and user support fit better as a corporate level activity (or under Manage Consumers in GAMSO)?