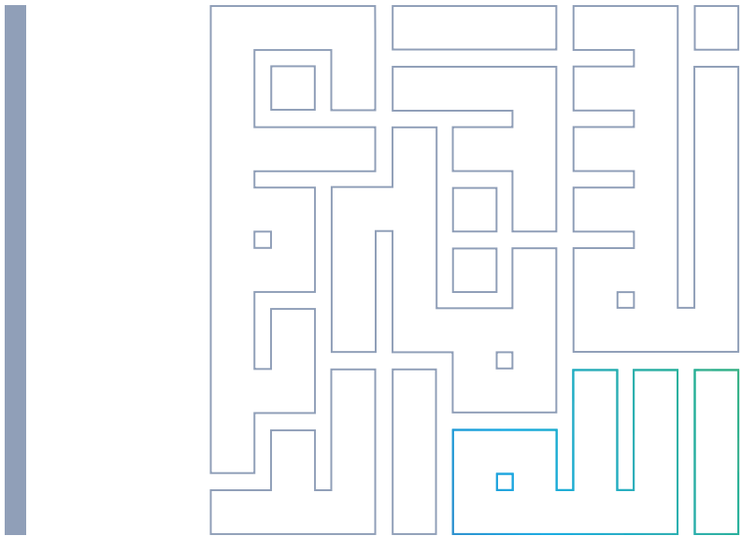
الهيئة العامة للإحصاء
General Authority for Statistics

# Generic Statistical Business Process Model (GSBPM)



**National version - 2026**
Saudi Arabia



رؤيـــة VISION
2030
المملكة العربية السعودية
KINGDOM OF SAUDI ARABIA

www.stats.gov.sa

# Contents


4 First: Introduction
4 Second: Terms and definitions
6 Third: Overview of GSBPM
6 Four: Key changes from the previous version
7 Five: GSBPM objectives and uses
8 Sixth: Understanding the model
9 Seventh: Structure of the model
11 Eighth: Relationship with other models and frameworks
13 Ninth: Applicability
15 Tenth: Overarching activities:

15 1. Quality management
17 2. Metadata management
18 3. Data management
18 4. Process data management
19 5. Knowledge management
19 6. Data supplier management

20 Eleventh: Statistical business process phases and their sub-processes

20 **1. Specify Needs phase**
21 Sub-processes and activities in the Specify Needs phase
21 1.1 Sub-process: Identify Needs
21 1.2 Sub-process: Consult and confirm needs
22 1.3 Sub-process: Establish output objectives
22 1.4 Sub-process: Identify concepts
22 1.5 Sub-process: Check data availability and suitability
23 1.6 Sub-process: Prepare and submit business case

24 **2. Design Phase**
24 Sub-processes and activities in the Design phase
24 2.1 Sub-process: Design outputs and dissemination
25 2.2 Sub-process: Design variables
25 2.3 Sub-process: Design collection
27 2.4 Sub-process: Design frame and sample
27 2.5 Sub-process: Design processing and analysis
28 2.6 Sub-process: Design production systems and workflows

28 **3. Build phase**
28 Sub-processes and activities in the Build phase
28 3.1 Sub-process: Reuse or build collection instruments
30 3.2 Sub-process: Reuse or build processing and analysis components
30 3.3 Sub-process: Reuse or build dissemination components
30 3.4 Sub-process: Configure workflows

31 3.5 Sub-process: Test production systems
31 3.6 Sub-process: Test statistical business process
31 3.7 Sub-process: Finalise production systems

**32 4. Collect phase**

32 Sub-processes and activities in the Collect phase
32 4.1 Sub-process: Create frame and select sample
33 4.2 Sub-process: Set up collection
34 4.3 Sub-process: Run collection
35 4.4 Sub-process: Finalise collection

**37 5. Process phase**

37 Sub-processes and activities in the Process phase
37 5.1 Sub-process: Integrate data
39 5.2 Sub-process: Classify and code
39 5.3 Sub-process: Review and validate
40 5.4 Sub-process: Edit and impute
41 5.5 Sub-process: Derive new variables and units
41 5.6 Sub-process: Calculating weights
42 5.7 Sub-process: Calculate aggregates
42 5.8 Sub-process: Finalise data files

**43 6. Analyse phase**

43 Sub-processes and activities in the Analyse phase
43 6.1 Sub-process: Prepare draft outputs
44 6.2 Sub-process: Validate outputs
45 6.3 Sub-process: Interpret and explain outputs
45 6.4 Sub-process: Apply disclosure control
46 6.5 Sub-process: Finalise outputs

**46 7. Disseminate phase**

47 Sub-processes and activities in the Disseminate phase
47 7.1 Sub-process: Update output systems
47 7.2 Sub-process: Produce dissemination products
48 7.3 Sub-process: Release dissemination products
49 7.4 Sub-process: Promote dissemination products
49 7.5 Sub-process: Provide user support

**50 8. Evaluate phase**

51 8.1 Sub-process: Gather evaluation inputs
51 8.2 Sub-process: Conduct evaluation
53 8.3 Sub-process:Agree an action plan

**54 Twelfth: List of Abbreviations**

## First: Introduction

Building on the General Authority for Statistics' (GASTAT) vision to become one of the world's leading modern statistical organizations. And its mission to provide accurate, comprehensive, and high-quality statistical information and data covering all aspects of life in Saudi Arabia, including social, economic, and environmental dimensions. By applying international best practices in data methodologies and processes. And leading the development of the statistical sector to support decision-making.

GSBPM serves as a key enabler in achieving this mission. It describes and defines a set of statistical activities within the business processes required to produce official statistics, while providing a standard framework that supports the modernization of GASTAT's statistical production processes in line with the latest international practices in this field. Furthermore, the Model serves as one of the operational frameworks for organizing work within the statistical sector and defining its business methodology, in accordance with Cabinet Resolution No. (11), dated 13/1/1437 AH, approving the Organization of the General Authority for Statistics, and specifically Paragraph (2) of Article 4, which stipulates that: "GASTAT is the entity responsible for statistics, the sole official reference for carrying out statistical work, and its technical supervisor and regulator. To achieve its objectives, GASTAT shall have the following mandates", including: «Carry out statistical work in accordance with recognized international standards, including: Define the statistical work methodology».

As part of the international statistical community, GASTAT has taken into account international recommendations and practices when preparing the national version of GSBPM, ensuring its alignment with GASTAT's organizational context.

## Second: Terms and Definitions

**The following terms, wherever they appear in GSBPM, shall have the meanings assigned to each of them:**

1. **GASTAT:** General Authority for Statistics (GASTAT).
2. **Statistical sector:** It comprises GASTAT as the central statistical agency responsible for implementing statistical work and ensuring its comprehensive coverage, along with an integrated system encompassing all statistical departments, units, and sections across all government sectors and institutions.
3. **National model:** Refers to the national version of the Generic Statistical Business Process Model. and may be referred to simply as GSBPM.
4. **International model:** Refers to the Generic Statistical Business Process Model (GSBPM), version 5.2, issued by the United Nations Economic Commission for Europe (UNECE) in May 2025.

5. **Statistical product:** A structured statistical undertaking based on a series of scientific phases, aimed at providing accurate statistical data on specific characteristics of a population, whether through sampling or complete enumeration.
6. **Statistical assets:** Statistical data and information owned by GASTAT.
7. **Databases:** The storage of statistical data using computer software and systems.
8. **Governance:** A set of rules and procedures under which the project is managed and overseen, and the relationships among all relevant stakeholders are regulated.
9. **Authorized decision-maker:** The final decision-maker in accordance with GASTAT's approved governance models.
10. **Users:** Consumers of statistical outputs (such as policymakers, researchers, businesses, and the public).
11. **Respondents:** Entities providing information about themselves or others (primary data) to statistical organizations or other producers of official statistics through forms, surveys, or polls.
12. **Data holders:** Custodians of datasets created or collected for non-statistical purposes, such as government agencies, private companies, or register administrators, who may or may not provide data to national statistical offices (NSOs).
13. **Data suppliers:** Data holders that provide national statistical offices (NSOs) with data, including administrative, geospatial, or other non-survey data.
14. **Stakeholders:** A broad category comprising users, respondents, data suppliers, data holders, and other parties involved in or affected by the statistical process.
15. **Administrative data:** Data generated through the routine administrative processes of public entities, such as tax records, health records, or school enrollment records.
16. **Privately held data:** Data generated through the routine operations of private entities, such as commercial transactions, digital platform records, or social media interactions.
17. **Big data:** High-volume, high-velocity, and high-variety data derived from digital technologies, such as social media activity, mobile device usage, and sensor networks.
18. **Geospatial data:** Data containing geographic or spatial attributes, including satellite imagery, Global Positioning System (GPS) data, and mapping systems, which GSBPM takes into account in a statistical context.
19. **Metadata:** Information that describes data and statistical processes in a standardized manner by providing information on data sources, methods, definitions, and classifications.
20. **Microdata:** Detailed data at the unit level, such as an individual, household, institution, or establishment, representing an entity whose characteristics can be measured and Analysed.
21. **Paradata:** Process metadata, i.e., data generated automatically or operationally about how data are collected, formatted, and processed. They describe how, when, and under what conditions the original statistical data were obtained, without constituting the responses or substantive data themselves.

## Third: Overview of GSBPM

In 2026, GASTAT developed its national version of GSBPM to replace the previous national version issued in 2024. The national version was developed to align with GSBPM version 5.2, issued by UNECE in May 2025.

Further updates to the national version are expected to be issued in the coming years to ensure alignment with new developments and models from the High-Level Group for the Modernisation of Official Statistics (HL-G-MOS), including GSBPM.

## Fourth: Key changes from the previous version

**The key changes to the national version of GSBPM in the current 2026 version compared with the previous 2024 version include the following:**

**1. Improvements across different structural levels:**

- Renaming and repositioning overarching activities: "Overarching processes" was renamed "overarching activities" to align with the Generic Activity Model for Statistical Organizations (GAMSO) and repositioned earlier in the current model to emphasize that these activities apply throughout the lifecycle of the statistical project rather than constituting a subsequent phase.
- Renaming certain sub-processes: Certain sub-processes were renamed to make their meaning and purpose clearer, without adding or removing any sub-processes.

**2. Clarifying the relationship between GSBPM and other models:**

- Relationship with GAMSO.
- Relationship with the Generic Statistical Information Model (GSIM).

**3. Improving or rephrasing descriptions and explanations, including:**

- Additional clarification of the approach to engaging with both respondents and users, emphasizing the need to balance the "response burden" on respondents with the growing needs of users.
- Greater focus on administrative sources and other non-survey data sources.
- Emphasis on the non-sequential nature of GSBPM.
- Additional descriptions of overarching activities, including: Process data management, knowledge management, and data supplier management.
- Clarification of the role of unit types (observation units and statistical units) in distinguishing between concepts and variables.
- Distinction between data as raw or aggregated outputs and products as outputs prepared for end users.

**4. Topics incorporated or given greater emphasis, such as:**

- Multiple clarifications on geospatial considerations.
- Iterative development approaches between the design and build phases, and the importance of considering cybersecurity during these phases.
- Multi-mode data collection.
- Pseudonymisation, also referred to as pseudonymous masking, partial anonymisation, data perturbation, or the removal of direct identifiers.
- Reference to microdata as a potential output in its own right.

Overall, most of the improvements and developments in this national version were aligned with the approach adopted in the international version. As with previous revisions, the model was not significantly changed from the previous version except where there was a strong business case for doing so.

## Fifth: GSBPM objectives and usesy

1. Improving and harmonizing statistical processes used in the production of statistical work across GASTAT and the statistical sector.
2. Providing a structure for documenting statistical processes, as well as organizing and storing them within GASTAT, thereby enhancing standardization and identifying good practices.
3. Providing a framework for assessing statistical process quality and a mechanism for facilitating the harmonization of process standards as far as possible, thereby fostering a culture of quality, accuracy, comprehensiveness, and transparency to enhance the quality of statistical outputs.
4. Integrating international work on statistical metadata with work on data quality by providing a common framework and common terminology for describing statistical business processes.
5. Enabling methodological standards to be linked to the relevant phase(s) or sub-process(es), which can then be classified and stored within a GSBPM-based structure.
6. Defining the components of statistical processes in a manner that facilitates the sharing of methods and tools across statistical business processes and also facilitates exchanges among different statistical organizations applying GSBPM.
7. Developing a repository of business process models to store process modelling outputs and enable them to be linked to the statistical business process model.
8. Providing a basic model for designing a training framework and a set of standard terms to describe the skills and expertise required in the statistical production process.
9. Providing a tool for capacity building and systematically building technical knowledge by referring to the details of each phase.

10. Measuring the operational costs of statistical processes by estimating the costs of different parts of the processes, thereby helping to identify modernization activities to improve the efficiency of the most costly parts of the processes.
11. Measuring system performance by identifying components that operate inefficiently, unnecessarily duplicate one another, or require replacement, as well as identifying gaps for which new components should be developed.
12. Providing a tool for harmonizing processes associated with non-statistical data suppliers (such as administrative data and geospatial data), facilitating communication between statisticians and experts from other domains, and harmonizing the relevant terminology.

## Sixth: Understanding the model

- A statistical business process comprises a set of interconnected and structured activities and tasks that transform inputs into outputs. In the context of GSBPM, statistical business processes are undertaken to produce official statistics that meet users' needs. The outputs of the process may comprise a combination of physical or digital products that present data and metadata in different ways, such as publications, maps, and/or electronic services, among others.
- GSBPM should be applied and interpreted flexibly. It is not a rigid framework in which all steps must be followed in a strict order. Rather, it identifies the possible activities that may be undertaken as steps in a statistical business process and the interdependencies between them.
- GSBPM is inherently non-sequential, allowing flexibility to revisit phases or sub-processes as needed. While the model is often presented sequentially for clarity, in practice, activities may occur in parallel, be repeated in loops, or be skipped depending on specific requirements or circumstances.
- Typically, the first three phases are considered only when creating a new output or when the process is reviewed as a result of an evaluation. Once the output becomes part of ongoing "routine" activity, these phases are not carried out (for example, it is not necessary to build new collection tools each time Labor Force Survey data are collected).
- GSBPM should be viewed as a matrix with many possible paths through it. In this way, the model aims to be sufficiently generic to be widely applicable and to promote a standard view of statistical production processes, without becoming overly restrictive or excessively abstract and theoretical.
- GSBPM takes into account a number of different entities that should be considered in a statistical business process, such as: Users, respondents, data holders, data suppliers, and stakeholders.

## Seventh: Structure of the model

**The national version of GSBPM consists of four levels, as follows:**

- Level 0: Represents statistical business processes.
- Level 1: Represents the eight phases of statistical business processes.
- Level 2: Represents the sub-processes within each phase.
- Level 3: Represents the activity/activities within each sub-process.

- **Phases, sub-processes and activities have been numbered at three consecutive levels:**

1. 1. Phase (single-level number)
   1.1 Subprocess (two-level number)
   1.1.1 Activity (three-level number)

- The phases and sub-processes are designed to be flexible. Some elements of the model may be more suitable for one type of statistics than another, while in another product, some elements may overlap with one another and be repeated until final approval is obtained.
- The model comprises eight (8) phases and forty-four (44) sub-processes. The model also outlines a number of activities that fall under each sub-process, where these activities describe the sub-processes in detail.

**The following figure illustrates the statistical business processes, their interconnected phases, and the sub-processes that constitute these phases:**

| Statistical Business Processes | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Specify needs** | | **Design** | | **Build** | | **Collect** | | **Process** | | **Analyse** | | **Disseminate** | | **Evaluate** | |
| 1.1 | Identify Needs | 2.1 | Design outputs and dissemination | 3.1 | Reuse or build collection instruments | 4.1 | Create frame and select sample | 5.1 | Integrate data | 6.1 | Prepare draft outputs | 7.1 | Update output systems | 8.1 | Gather evaluation inputs |
| 1.2 | Consult and confirm needs | 2.2 | Design variable | 3.2 | Reuse or build processing and analysis components | 4.2 | Set up collection | 5.2 | Classify and code | 6.2 | Validate outputs | 7.2 | Produce dissemination products | 8.2 | Conduct evaluation |
| 1.3 | Establish output objectives | 2.3 | Design collection | 3.3 | Reuse or build dissemination components | 4.3 | Run collection | 5.3 | Review and validate | 6.3 | Interpret and explain outputs | 7.3 | Release dissemination products | 8.3 | Agree an action plan |
| 1.4 | Identify concepts | 2.4 | Design frame and sample | 3.4 | Configure workflows | 4.4 | Finalise collection | 5.4 | Edit and impute | 6.4 | Apply disclosure control | 7.4 | Promote dissemination products | | |
| 1.5 | Check data availability and suitability | 2.5 | Design processing and analysis | 3.5 | Test production systems | | | 5.5 | Derive new variables and units | 6.5 | Finalise outputs | 7.5 | Provide user Support | | |
| 1.6 | Prepare and submit business case | 2.6 | Design production systems and workflows | 3.6 | Test statistical business process | | | 5.6 | Calculate weights | | | | | | |
| | | | | 3.7 | Finalise production systems | | | 5.7 | Calculate aggregates | | | | | | |
| | | | | | | | | 5.8 | Finalise data files | | | | | | |

**Figure 1**: Phases (Level 1) and sub-processes (Level 2) of GSBPM

In addition to these levels described in Section 11 of this version, Section 10 is dedicated to overarching activities, which are applied throughout the eight phases of statistical processes. The overarching activities include the following:

1) **Quality management:** This overarching activity includes quality assessment and control mechanisms. It recognizes the importance of evaluation and feedback throughout the phases of the statistical process and throughout the statistical business process.

2) **Metadata management:** Metadata are produced and used within each phase, requiring an integrated management system to ensure that the relationship between data and metadata is maintained. This also includes metadata preservation, ownership, quality, and rules for archiving and disposal.

3) **Data management:** This activity covers matters such as data security, ownership, quality, and rules for archiving, retention, and disposal.

4) **Process data management**: This includes activities for recording, classifying, and using data arising from the execution of the statistical business process. Process data can help detect and understand patterns in the collected data, as well as evaluate the execution of the statistical business process as a whole.

5) **Knowledge management:** An ongoing activity that primarily involves maintaining documentation of recurring statistical business processes to ensure their repeatability and replicability.

6) **Data supplier management**: This involves managing response burden across joint processes, as well as topics such as profiling and contact information management (thereby having particularly close links with statistical business processes that manage registers).

**The following figure shows the overarching activities across the phases and sub-processes:**

| Overarching activities | | | | | |
|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 |
| Data supplier management | Knowledge management | Process data management | Data management | Metadata management | Quality management |

**Figure 2:** Overarching activities across the phases and sub-processes in GSBPM

## Eighth: Relationship with other models and frameworks

**GSBPM is closely related to two statistical models:**

**1) Generic Activity Model for Statistical Organizations (GAMSO)**

- GAMSO complements and extends GSBPM by describing and defining activities that occur within a typical statistical organization and that are either not directly related to the production of statistics or are managed at the corporate or strategic level (i.e., activities in the areas of Strategy and Leadership, Capability Development, and Corporate Support).
- These activities include human resource management and quality management activities carried out at the corporate level, such as developing a quality framework, among others. GAMSO provides a general description of these activities. GSBPM sits within GAMSO and constitutes its "Production" activity area, as shown in Figure 3.




**Figure 3:** GAMSO activity areas

- Like GSBPM, GAMSO aims to provide a common vocabulary and framework to support international collaboration activities. GAMSO provides greater value when applied in conjunction with GSBPM.

### 2) Generic Statistical Information Model (GSIM)

- GSIM is a reference framework for statistical information, designed to help modernize official statistics at both national and international levels. It enables a generic description of the definition, management, and use of information (i.e., data and metadata) throughout the statistical production process. It also provides a set of consistently described standardized information classes that represent the inputs and outputs of GSBPM sub-processes, and GSIM helps explain the important relationships among the entities involved in statistical production and can be used to guide the development and use of consistent implementation standards or specifications.
- Like GSBPM, GSIM is one of the key foundations for modernizing official statistics and moving away from working independently by domain (economic, social, and environmental). It defines approximately 130 information classes, examples of which include datasets, variables, statistical classifications, units, and statistical populations, in addition to the rules and standards required to operate production processes (such as data editing rules).
- GSIM and GSBPM are complementary models for the production and management of statistical information. As shown in Figure 4 below, GSIM helps describe GSBPM sub-processes by identifying the information classes that flow between them, are created within them, and are used by them to produce official statistics. Inputs and outputs can be defined in terms of information classes and are formally specified in GSIM.




**Figure 4:** GSIM information classes as inputs and outputs of a GSBPM sub-process

- GSIM will have greater value if applied in conjunction with GSBPM. Likewise, greater value will be obtained from GSBPM if it is applied in conjunction with GSIM. However, it is possible (though not ideal) to apply one without the other.
- Applying GSIM and GSBPM together can facilitate building efficient metadata-driven systems and help align and standardize statistical computing infrastructures.

## Ninth: Applicability

**1.** When a new statistical product is produced for the first time, it goes through all eight phases:

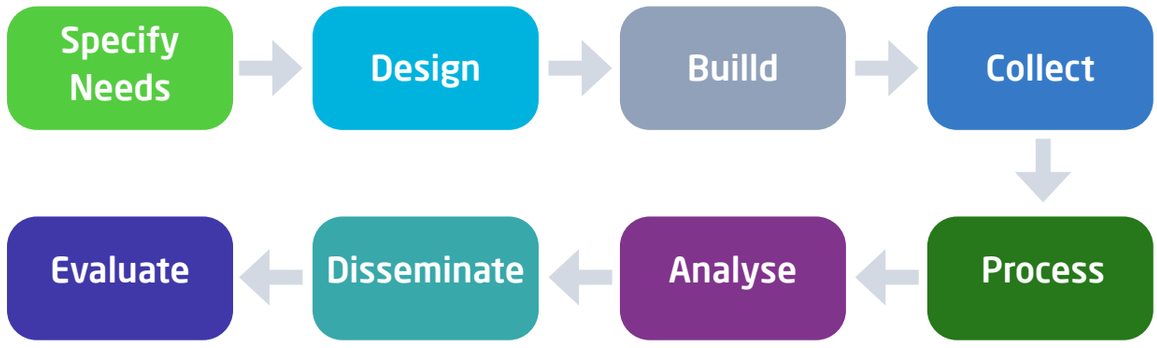



**Figure 5:** Statistical business process phases according to GSBPM

**2.** Four phases are reviewed when the statistical product is reissued:
- Collect phase
- Process phase
- Analyse phase
- Disseminate phase

**3.** Each time a statistical product is produced, the number of times the following phases should be revisited is determined:
- Specify needs phase
- Design phase
- Build phase

**4.** The model is designed to be applicable regardless of the data source, so that it can be used to design, describe, manage, and assess the quality of processes based on surveys, censuses, administrative sources, and other sources that are primarily collected for non-statistical purposes but can be used in statistical production, including (but not limited to) the following types:
- Administrative data
- Privately held data
- Big data
- Geospatial data

**5.** The division used in the model allows phases and their sub-processes to be initiated sequentially or in parallel, depending on the type of statistical product.

**6.** While typical statistical business processes involve collecting and processing data to produce statistical outputs, GSBPM also applies when existing data are revised or time series are recalculated, whether as a result of improvements to source data or a change in methodology. In these cases, the input data may consist of original microdata and/or additional data, which are then processed and analysed to produce revised outputs. In such cases, many sub-processes and possibly some phases (particularly the early phases) are likely to be omitted. Similarly, GSBPM can be applied to processes such as compiling national accounts and typical processes in international statistical organizations that use secondary data from countries or other organizations.

**7.** In addition to being applicable to processes that result in statistical outputs, GSBPM can also be applied to the development and maintenance of statistical registers, where the inputs are similar to those used for statistical production (although typically with a greater emphasis on administrative data), and the outputs are usually frames or other data extracts, which are then used as inputs to other processes.

**8.** Artificial intelligence (AI) and machine learning (ML) play an increasingly prominent role in statistical production, facilitating the automation of manual tasks and the use of new types of data, and machine learning is referenced where relevant, particularly in the Process and Analyse phases. AI and ML have the potential to be powerful tools in the context of the statistical production process, but they also need to be used appropriately, for example, to ensure the accuracy and validity of any outputs produced. However, GSBPM remains applicable regardless of the specific technologies or methods used to carry out activities in the process.

**9.** GSBPM is sufficiently flexible to be applied to all the scenarios mentioned above.

## Tenth: Overarching activities

GSBPM recognizes many overarching activities that apply throughout the production phases and across statistical business processes.
For example, quality is monitored and assured at different phases of production. Data and metadata are managed throughout the entire process, rather than only within a specific sub-process.
While the production process involves different units and expertise within GASTAT at different phases, it is important that these overarching activities are carried out consistently in accordance with GASTAT's general standards and policies, and in close coordination with a relevant central unit. While these overarching activities are carried out throughout the production phases and across statistical business processes, the activities included in the Generic Activity Model for Statistical Organizations (GAMSO) are carried out at the organizational level to support statistical production.

**Overarching activities include:**

### 01. Quality management:

- The Quality Management overarching activity in GSBPM specifically refers to product and process quality management that takes place within the scope of the production process. More broadly, the concept of quality applies to organizations, products, sources, and processes. Quality at the organizational level (such as adopting a quality policy or a quality assurance framework) or in areas not directly linked to production (such as the quality of an induction programme for new staff) is considered within GAMSO.
- The primary objective of quality management within the statistical production process is to understand and manage the quality of sources, processes, and products. There is general agreement among statistical organizations that quality should be defined according to ISO 9000:2015 as the "degree to which a set of inherent characteristics fulfils requirements". Consequently, quality is a complex, multifaceted concept and is typically defined in terms of several quality dimensions. The quality dimensions considered most important depend on the user's perspective, needs, and priorities, which vary across processes and user groups.
- To improve quality, quality management should be present throughout the business process model. Although it is closely linked to the Evaluate phase, quality management has a deeper and broader scope. In addition to evaluating process iterations, it is also necessary to evaluate separate phases and sub-processes, ideally each time they are applied, or at least according to an agreed schedule. The process metadata generated by the different sub-processes are themselves important as inputs to process quality management. These evaluations can apply within a specific process or across several processes that use common components. Furthermore, the set of quality control measures that should be implemented within sub-processes to prevent and monitor errors and sources of risk plays an essential role in quality management; these measures should be documented and can be used to prepare quality reports.

- Quality management within GASTAT should be based on the defined quality framework approved by senior management. It is not necessary for that framework to fully align with quality frameworks within other statistical organizations; the diversity of such frameworks underscores the importance of benchmarking different results across organizations. While this framework may not be appropriate to apply to every iteration of every part of every statistical business process, it should be used systematically according to a predefined schedule that allows all major parts of the process to be reviewed over a specified period.
- By expanding the scope of the Quality Management overarching activity, the evaluation of groups of statistical business processes can also be considered in order to identify potential duplication or gaps.
- All evaluations lead to feedback, which should be used to improve the relevant process, phase, or sub-process, thereby creating a "quality loop" that reinforces a continuous improvement approach and organizational learning. The quality loop consists of four main phases:
  - **Plan:** The Plan stage primarily refers to the first three phases of GSBPM. The selection or design of the quality framework should be part of the design of the production process, in order to have the necessary means to produce the information required for managing the quality of the process and statistical products.
  - **Run:** When running the process, care should be taken to collect data and metadata for quality management, while updating the data and metadata generated from the execution of each phase of GSBPM.
  - **Evaluate:** An evaluation of quality indicators should be conducted throughout the process to make any required corrections in a timely manner. During the evaluation phase, the summary and analysis of the quality information collected, along with any changes made during the process, will lead to conclusions regarding any improvements that may be required.
  - **Improve:** In response to the quality evaluation, any required improvements can be applied to the process and then reflected and incorporated during the Plan stage of the next cycle.




**Figure 6:** Quality loop

- **Examples of quality management activities include the following:**

  - Establishing quality standards at each phase of the project for use in the process.
  - Defining process quality targets and monitoring compliance.
  - Examining process metadata and quality indicators.
  - Identifying and assessing risks, including: Financial risks, risks related to employee benefits, risks related to customer relations, etc.
  - Implementing risk treatments to ensure fit-for-purpose quality.
  - Monitoring the project work plan and budget.
  - Analysing users' comments and feedback.
  - Reviewing processes and documenting lessons learned.
  - Conducting internal or external audits of the process.
  - Reporting on workflow progress.
  - Approving corrective actions.

## 02. Metadata management:

- Metadata plays an important role and should be managed at an operational level within the statistical production process. This should be done in accordance with the relevant organizational policies and standards. When aspects of metadata management are considered at the organizational or strategic level (for example, where metadata systems exist that affect large parts of the production system), they should be taken into account within the framework of the Generic Activity Model for Statistical Organizations (GAMSO).

- Good metadata management is essential for the efficient operation of statistical business processes. Metadata exist at every phase, where they are either created, updated, carried forward from a previous phase, or reused from another business process. In the context of this model, the focus of the Metadata Management overarching activity is on the design/creation/review, updating, use, and archiving of statistical metadata, although the metadata of the various sub-processes themselves are also of interest as an input to quality management. The main challenge is to ensure that these metadata are captured as early as possible, stored, and transferred from phase to phase alongside the data to which they refer. Consequently, a metadata management strategy and systems are vital to the operation of this model, and this can be facilitated by GSIM.

- GSIM is an information reference framework and enables generic descriptions of the definition, management, and use of data and metadata throughout all phases of the statistical production process. GSIM also supports a consistent approach to metadata, thereby facilitating their essential role, which is for metadata to uniquely and formally define the content of and links between information classes and processes in the statistical information system.

## 03. Data management:

- Data management is essential, as data are produced within many activities of the statistical business process and constitute the main outputs. The main objective of data management is to ensure that data are used appropriately and remain usable throughout their lifecycle, as data management throughout the lifecycle covers activities such as planning and evaluating data management processes, as well as establishing and implementing processes related to data collection, organization, use, protection, preservation, and disposal.
- How data are managed is closely linked to how they are used, which in turn is linked to the statistical business process in which the data are created. Both the data and the processes in which they are created should be well identified and defined to ensure proper data management.
- **Examples of data management activities include the following:**

  - Establishing a governance structure and defining responsibilities for data oversight.
  - Designing data structures and associated datasets, as well as the source and flow of data through the statistical business process.
  - Specifying the database (repositories) for data storage and managing the database.
  - Documenting data (such as logging and inventorying them, and classifying them by content, retention, or any other required classification).
  - Defining data retention periods and archiving, identifying risks that may affect data preservation, and finding solutions to address them, such as:
    - Limited rules for the retention and disposal of data and metadata at the organizational level (for example, deciding which versions need to be retained).
    - Difficulties that may arise in archiving administrative data collected through the statistical business process.
  - Securing data against unauthorized access and use.
  - Protecting data against technological change, deterioration of physical media, and data corruption.
  - Performing data integrity checks (such as periodic checks that provide assurance regarding the accuracy and consistency of data throughout their entire lifecycle).
  - Carrying out data disposal activities once the retention period has ended.

## 04. Process data management:

- The Process Data Management overarching activity within GSBPM specifically refers to managing process data within the scope of the production process. Activities related to process data management at the organizational level should be considered within the framework of GAMSO.

- This overarching activity includes recording, classifying, and using data arising from the execution of the statistical business process. This may include, for example, paradata records generated from the execution of program code, as well as information arising from fieldwork. These data are not quality information in themselves, as they may be highly detailed (such as the length of time taken by a single field researcher to conduct a single interview), but they can be used to derive quality information (such as the total or average time taken to conduct interviews).
- This process data can be reviewed to assess how the statistical business process was carried out, for example, to improve its efficiency or to identify potential problems that occurred and determine their causes. Process data can also be used to identify and understand patterns in the data used to produce statistics.

## 05. Knowledge management:

- The Knowledge Management overarching activity refers to managing knowledge specific to the production process. Activities related to knowledge management at the organizational level should be considered within the framework of GAMSO, such as the "Manage information and knowledge" activity in GAMSO, under the Corporate Support activity area.
- This Knowledge Management overarching activity within GSBPM specifically refers to updating production process documentation (such as handbooks, manuals, and guidance notes already created within sub-process 3.7 (Finalise production systems)) based on knowledge gained about the process. Managing knowledge systematically using this documentation allows the production process to be enriched and improved. It can be taken into account when developing training for producers of statistics and can also inform the evaluation of the production process.
- Knowledge can be derived from a variety of sources, including experience gained by each individual involved in the process, which may relate to algorithm code, software, and scripts used throughout the statistical cycle, good practices, or lessons learned. Reference metadata, paradata analysis, incidents, process logs, data anomalies, and other intermediate outputs from the execution of the statistical process, although not knowledge in themselves, can all lead to useful knowledge about the production process that can be documented for transfer to others.

## 06. Data supplier management:

- The Data Supplier Management overarching activity within GSBPM refers to managing data suppliers within the scope of the production process. Activities related to data supplier management at the organizational level should be considered within the framework of GAMSO, such as the "Manage data suppliers" activity under the Corporate Support activity area.

- Building on the GAMSO Corporate Support activity "Manage data suppliers", which focuses on managing suppliers and ensuring compliance with agreements, this overarching activity within GSBPM expands to include technical-level interactions with data suppliers, including those providing registers.

- **Examples of data supplier management activities may include the following:**
  - Requesting specific information about input data when metadata and/or documentation are scarce.
  - Requesting assistance from the data supplier to overcome technical or operational issues; and/or requesting additional information in the event of unexpected changes in the data received.
  - Managing response burden shared across statistical processes, profiling, and managing contact information (for example, within registers).

## Eleventh: Statistical business process phases and their sub processess

| Phase | Main objective of the phase |
|---|---|
| Specify needs | Understand and confirm statistical needs, and identify possible solutions. |
| Design | Designing the processes and services needed to implement the procedures in the next phases. |
| Build | Build, organize, gather and test production solutions and systems so that they are ready for use in statistical implementation processes. |
| Collect | Collect and validate data. |
| Process | Process data to be converted from raw to usable statistical data. |
| Analyze | Validate and interpret outputs. |
| Disseminate | Prepare the products to be ready for dissemination. |
| Evaluate | Conduct an evaluation of the statistical process. |

## 01 | Specify Needs

**This phase aims to:**

1. Understand the context surrounding the specified need or change.
2. Identify beneficiaries' needs and requirements.
3. Establish clear boundaries and objectives for the statistical product.
4. Determine the scope of solutions to meet the need.

**This phase applies in two cases:**

1. Identifying the need for a new statistical product.
2. Initiating modifications to an existing statistical product that require it to be reviewed from the beginning.

**This phase consists of six (6) sub-processes, which are presented sequentially but can occur in parallel and can be iterative:**

1. Identify needs
2. Consult and confirm needs
3. Establish output objectives
4. Identify concepts
5. Check data availability and suitability
6. Prepare and submit business case

## Sub-processes and activities in the Specify Needs phase

### 1.1 Sub-process: Identify Needs

The work in this sub-process focuses on understanding users' needs for the statistical product and reviewing and examining comments and feedback on existing statistics through a number of procedures in order to determine the actual need for the statistical product and the feasibility of producing it or improving it if the product already exists..

- **Sub-process activities:**

1.1.1 Research the required statistics and identify what is required from those statistics. Understand the reasons for requesting them.

1.1.2 Determine whether to release a new statistical product or develop and improve an existing product, based on a request for new information from users, an organizational change such as a budget reduction, a periodic review, or comments and feedback from respondents.

1.1.3 Examine the available statistical assets that may meet the client's needs, while considering the practices of other statistical organizations (national and international) that produce similar data and the methods used by those organizations.

1.1.4 Consider action plans resulting from evaluations of previous iterations of this process or from other processes, including comments received from respondents, as inputs to this sub-process.

### 1.2 Sub-process: Consult and confirm needs

- **Sub-process activities:**

1.2.1 Communicating and consulting with internal and external stakeholders, and holding the necessary workshops and meetings to reach an in-depth understanding of their needs.

1.2.2 Clarifying the concepts to be measured from the users' perspective, and ensuring their consistency with existing statistical standards.

1.2.3 Determining whether previously identified needs have changed for existing statistical products intended for development.

1.2.4 Confirming and documenting the statistical needs.

## 1.3 Sub-process: Establish output objectives

- **Sub-process activities:**

1.3.1 Identifying the statistical output objectives necessary to meet the user needs identified in consultation with users (sub-process: 1.2 Consult and confirm needs).

1.3.2 Taking into account legal or regulatory constraints and frameworks (such as those related to confidentiality and methodologies), as well as available resources when setting output targets.

1.3.3 Aligning with stakeholders and agreeing on the suitability of the proposed outputs and their quality measures.

## 1.4 Sub-process: Identify concepts

- **Sub-process activities:**

1.4.1 Identifying and defining the concepts to be measured (these concepts are the same concepts identified by the client).

1.4.2 Linking the required concepts to existing statistical standards whenever possible. Aligning the identified concepts that do not conform to existing statistical standards when selecting and defining the concepts and variables to be used in sub-process 2.2 (Design variables).

## 1.5 Sub-process: Check data availability and suitability

- **Sub-process activities:**

1.5.1 Searching for available data and methodologies that assist in developing solutions to meet statistical needs.

1.5.2 Researching and evaluating potential administrative data sources or other non-traditional statistical sources (such as privately held data or "big data") in order to:

- Identify data sources that meet statistical purposes (such as the extent to which administrative concepts align with data requirements, and whether data frequency, quality, security, and continuity of supply are suitable).
- Consider any constraints or conditions regarding available data sources.
- Assess the division of responsibilities between the providers of such data and GASTAT.

- Verifying information and communication technology (ICT) resources, such as data storage and the technology required to handle incoming data and process data, as well as any formal agreements with data providers for accessing and sharing data (such as formats, delivery, accompanying metadata, and quality validation).

1.5.3 Developing a strategy to address any remaining gaps in data requirements. This may include identifying potential partnerships with data owners or conducting a new survey to collect data, if necessary.

1.5.4 Gathering information on legal frameworks that affect the extent to which such data can be collected and used, and accordingly may identify proposals to amend existing legislation or introduce a new legal framework.

## 1.6 Sub-process: Prepare and submit business case

In this sub-process, the findings from the other sub-processes in this phase are documented in the form of a "business case" to obtain approval to implement a statistical business process to produce new or improved statistics. This business case should be aligned with stakeholders' requirements.

- **Sub-process activities:**

1.6.1 Preparing the business case based on the inputs from the previous sub-processes of this phase, and the business case may include the following:

- A description of the current "As-Is" business process (if it already exists), providing information on how the current statistics are produced and highlighting any inefficiencies or issues that need to be addressed.
- Proposed solutions for the "To-Be" business process, detailing how the statistical business process should be developed to produce new or revised statistics or improve existing statistics.
- Assessing costs and benefits against the targets and any external constraints that may prevent or limit the implementation of the statistical process as required.

1.6.2 Describing the available options and providing recommendations through the submitted business case. These may include the benefits, costs, outputs, timeframe, budget, required technical and human resources, risk assessment, and impact on stakeholders for each proposed option. (Sometimes, there may be no suitable solution options that can adequately meet the identified needs.)

1.6.3 Submitting the business case to the authorized decision-maker or the relevant committees at GASTAT to decide whether to "go" and proceed to the next phase or "not go".

# 02 | Design

**This phase applies in two cases:**

1. When designing a new product based on the phase of understanding and confirming statistical needs and identifying possible solutions.
2. When redesigning the product for improvement and development based on the outputs of Phase 8 (Evaluate) from a previous cycle of the product.

**In this phase:**

1. Development and design activities, as well as any associated practical research work necessary to define statistical outputs, variables, methodologies, collection instruments, and operational processes are described, taking into account other requirements such as cybersecurity.
2. All design elements needed to define or improve the statistical products or services specified in the business case submitted during the Specify Needs phase are established.
3. All relevant metadata are identified, in accordance with organizational standards and policies, to be ready for subsequent use in the statistical process.
4. International and national standards are used extensively to reduce the duration and cost of the design process, as well as to enhance interoperability, comparability, and so on.

In some cases, design elements from existing processes can be reused or adapted rather than starting from scratch every time a new statistical product is designed, thereby saving time and effort and significantly reducing financial costs.
It is also important to consider the geospatial aspects of data during design to enhance the usability and value of statistical information.
The outputs of design processes may form the basis for future standards at the organizational, national, or international level.

**This phase consists of six (6) sub-processes, which are presented sequentially but can occur in parallel and can be iterative:**

1. Design outputs and dissemination
2. Design variables
3. Design collection
4. Design frame and sample
5. Design processing and analysis
6. Design production systems and workflows

## Sub-processes and activities in the Design phase

### 2.1 Sub-process: Design outputs and dissemination

This sub-process includes the detailed design of the statistical outputs, products, and services to be produced, including the associated systems, tools, and procedures used in the Disseminate phase.

- **Sub-process activities:**

2.1.1 Designing processes that control access to any confidential outputs.

2.1.2 Designing outputs to follow existing standards whenever possible (including outputs from statistical, administrative, geospatial, and other non-statistical registers and databases). Therefore, inputs to this process may include metadata from similar sets of outputs to those being designed, or from the same outputs in previous cycles, in order to compare results across different releases or to explain methodological differences or other changes, such as changes in classifications. (Examples of output metadata include: A description of the indicator or table, statistical unit, classifications, reference period and methodology, data quality, international standards, and information on practices in other statistical organizations derived from sub-process 1.1 (Identify Needs). … etc.).

2.1.3 Outputs may also be designed in partnership with other organizations that wish to do so, particularly if they are considered joint outputs or will be disseminated by the other organization.

2.1.4 Determining the release timing.

## 2.2 Sub-process: Design variables

Preparing metadata for collected and derived variables, as well as statistical and geospatial classifications, is a necessary prerequisite for the phases subsequent to the Design phase. This sub-process may need to be carried out in parallel with sub-process 2.3 (Design collection), as defining the variables to be collected and selecting collection instruments may be interdependent to some extent. It is also important to follow existing national and international standards when carrying out this sub-process wherever possible.

- **Sub-process activities:**

2.2.1 Identifying the variables to be collected through the collection instrument, as well as any other variables to be derived from them in sub-process 5.5 (Derive new variables and units).

2.2.2 Identifying the statistical or geospatial classifications to be used.

2.2.3 Designing unit types (including observation units and statistical units).

## 2.3 Sub-process: Design collection

This sub-process prepares the data collection strategy.

- **Sub-process activities:**

2.3.1 Identifying appropriate collection instruments and methods, including:
- Identifying instruments and methods based on the type of data collection, the type of collection unit (establishment, individual, or other), and the available sources.

- Identifying the type of collection instrument required, which may include computer-assisted personal or telephone interviews, paper forms, administrative registers, automated and semi-automated data transfer methods (such as application programming interfaces (APIs), system integration, and file transfer), web-scraping techniques, and geospatial data techniques.
- Taking into account considerations when adopting a multi-mode method, such as designing the sequence (such as intelligent chronological sequencing when moving from one mode to another) or designing different instruments for the different modes used.
- Incorporating the direct or indirect use of administrative data to adjust or support survey data when capturing survey information.

2.3.2 Designing collection instruments, questions, and form templates, including:

- Designing collection instruments, questions, and form templates (in conjunction with the variables and statistical classifications designed in sub-process 2.2 (Design variables)).
- Understanding respondents and incorporating users' perspectives (such as decision-makers or researchers) into data collection design by applying the concept of "user experience".
- Using supporting tools (such as question libraries/banks to facilitate their reuse, and form tools from the software and technical systems used to build and assemble forms) to enable the rapid assembly of questions into formats suitable for cognitive testing.

2.3.3 Designing field operations, respondent interaction, and agreement management, including:

- Designing field operations and respondent interaction by managing and designing the complete experience that the respondent goes through from the moment they are contacted or sent a self-completion link until they finish providing the data.
- Confirming existing agreements with data-supplying entities or establishing new agreements (using agreement templates to help standardize terms and conditions, such as service level agreements (SLAs), memoranda of understanding, or data access and supply agreements).
- Designing respondent management systems dedicated to tracking the status of each respondent (individual or establishment) throughout the survey period.

■ **Note:** In the event that a third party performs data collection and processing, the statistical organization is required to:

- Design mechanisms to monitor data and metadata.
- Assess the impacts of any changes made by the third party.

## 2.4 Sub-process: Design frame and sample

This sub-process applies particularly to statistical processes that require sample-based data collection (statistical surveys), but it may also apply to other types of data collection, such as complete enumeration surveys and censuses, for the purpose of identifying and updating the frame. It may also apply to register-based processes to filter records based on the target population.
This sub-process serves as preparation for creating both the sampling frame and the actual sample later in sub-process 4.1 (Create frame and select sample) in the Collect phase.

- **Sub-process activities:**

2.4.1 Identifying the target population and sampling frame sources:
- Identifying and defining the target study population.
- Defining the sampling frame and the register from which it is derived (administrative and statistical registers, censuses, and information derived from other sample surveys).

2.4.2 Selecting the methodology, integrating sources, and analysing the frame:
- Identifying the most appropriate sampling criteria and methodology (which may include complete enumeration).
- Describing how sampling frame sources are integrated (if necessary), with reference to geospatial data and classifications.
- Conducting an analysis to determine the extent to which the frame covers the target population and using spatial analysis techniques to ensure the suitability of the spatial distribution of units.

2.4.3 Developing the sampling plan, which serves as a procedural and implementation guide that translates the scientific and theoretical sampling methodology into practical steps and actual figures..

## 2.5 Sub-process: Design processing and analysis

This sub-process is concerned with designing statistical processing and analysis plans that will be implemented through the Process and Analyse phases. These plans include:
- Processing and analysis specifications and mechanisms.
- Rules and criteria for data linking, coding, and imputation of missing data.
It may be necessary to improve aspects of these plans during implementation, particularly when they do not perform as expected once the actual data for the statistical product become available.

- **Sub-process activities:**

2.5.1 Designing the statistical methodology to be applied during the Process and Analyse phases, including defining routines and rules for coding, editing, and imputation, while taking into account variations in these rules based on data collection methods and sources.

2.5.2 Designing the data linking plan, including the technical specifications and mechanisms required to combine and integrate data received from multiple and different sources.

2.5.3 Designing data validation and statistical estimation methods through:

- Establishing "smart editing rules" and statistical and subject-matter criteria to ensure the reasonableness of the data after they have been combined and integrated. Examples of such criteria include: Logical consistency that prevents contradictions in the data and checking for outliers.
- Designing statistical estimation methods to generalize the results to represent the population as a whole.

2.5.4 Designing methods and mechanisms for statistical disclosure control to protect confidentiality.

## 2.6 Sub-process: Design production systems and workflows

This sub-process is concerned with designing and structuring the overall workflow of the statistical production process, as well as evaluating existing production solutions and allocating roles and responsibilities.
This is carried out through governance committees and their meetings and, ideally, is enhanced as shared services across the different departments at GASTAT.

- **Activities of the Process:**

2.6.1 Identifying the statistical workflow from the Collect phase through to the Disseminate phase, with an overview of all processes required to ensure that they are efficiently coordinated and integrated without gaps or duplication.

2.6.2 Evaluating existing production solutions and allocating roles & responsibilities, including:

- First examining existing production solutions (such as services, systems, and databases) to determine their fitness for purpose, applying the general principle of reusing processes and technologies across multiple statistical business processes.
- Designing new technical solutions only if any gaps are identified after examining the existing solutions.
- Determining how staff interact with systems and clearly defining roles and responsibilities (who will be responsible for what and when).

# 03 | Build phase

**Build, as a phase of statistical business production, means:** Building, organizing, assembling, and testing production solutions and systems until they are ready for use in statistical implementation processes, where the outputs of the Design phase are assembled and configured in this phase to

create a complete operational environment for running the process. This includes developing and improving the systems used for data collection, processing, analysis, and dissemination. This is done by making use of previous cycles of the statistical process.
This phase is carried out and reviewed each time a new statistical product is produced or when an existing product is developed and improved, rather than in every iteration.
When gaps exist in current systems/services, new systems/services that are widely reusable and compliant with cybersecurity requirements need to be developed or built.

**This phase consists of seven (7) sub-processes, which are presented sequentially but can occur in parallel and can be iterative. Iteration between the sub-processes in the Design and Build phases may also be required.**

The first three sub-processes are concerned with developing and improving the systems used for data collection, processing, analysis, and dissemination. The final four sub-processes focus on the overall end-to-end process. These sub-processes are:

1. Reuse or build collection instruments
2. Reuse or build processing and analysis components
3. Reuse or build dissemination components
4. Configure workflows
5. Test production systems
6. Test statistical business process
7. Finalise production systems

## Sub-processes and activities in the Build phase

### 3.1 Sub-process: Reuse or build collection instruments

This sub-process is concerned with building and reusing the collection instruments that will be used during the Collect phase.

- **Sub-process activities:**

3.1.1 Building collection instruments based on the design specifications developed during sub-process 2.3 (Design collection), whether one or more collection instruments are required. Collection instruments may also consist of routines for extracting data from existing statistical or administrative registers (for example, using existing service interfaces).

3.1.2 Preparing and testing the content and performance of the collection instrument (such as cognitive testing of questionnaire questions).

3.1.3 Directly linking collection instruments to the metadata system, as this procedure enables easier access to metadata during the Collect phase. This procedure enables

work to be saved in later phases. Obtaining data collection measures/metrics (paradata) is an important consideration in this sub-process for calculating and analysing process quality indicators.

## 3.2 Sub-process: Reuse or build processing and analysis components

This sub-process is concerned with reusing existing components or building new components required for the Process and Analyse phases.

- **Sub-process activities:**

3.2.1 Reusing existing components or building new components required for the Process and Analyse phases, as designed in sub-process 2.5 (Design processing and analysis), may include: Tasks and features related to dashboards, information services (the technical mechanism that enables statistical information to flow between different systems or be displayed to users), transformation functions, geospatial data services, workflow management tools, and respondent and metadata management services.

## 3.3 Third sub-process: Reuse or build dissemination components

This sub-process is concerned with building new components or reusing existing components required for disseminating statistical outputs.

- **Sub-process activities:**

3.3.1 Reusing existing components required for disseminating statistical outputs or building new components as designed in sub-process 2.1 (Design outputs and dissemination) of the Design phase, which include all types of dissemination components, ranging from those that produce traditional paper publications to those that provide web services, (linked) open data outputs, geospatial statistics, maps, or access to microdata.

## 3.4 Sub-process: Configure workflows

This sub-process configures the workflows in preparation for testing them in the next sub-process.

- **Sub-process activities:**

3.4.1 Configuring and testing the workflows, systems, and transformations used in the statistical process, from data collection through to dissemination. Based on the design developed in sub-process 2.6 (Design production systems and workflows) of the Design phase.

3.4.2 Modifying workflows for specific purposes that require modification, or combining workflows for different phases together (using an automated workflow/business process management system), and configuring the systems accordingly.

## 3.5 Sub-process: Test production systems

In this sub-process, the workflows configured in the previous sub-process are tested in preparation for their release to the project in the production environment. This sub-process is logically linked to sub-processes 3.1 (Reuse or build collection instruments), 3.2 (Reuse or build processing and analysis components), and 3.3 (Reuse or build dissemination components), with regard to testing the components and services developed in those sub-processes.

- **Sub-process activities:**

3.5.1 Testing the services assembled in the Design phase and the newly configured services in the Build phase as a result of gaps in the existing list of services. This includes technical testing and approval of new software and procedures.

3.5.2 Ensuring that existing routines from other statistical business processes are suitable for use in this case.

3.5.3 Testing the consistency between the assembled and configured services, and ensuring that the production solutions system

3.5.4 operates as a whole in a consistent and coherent manner.

3.5.5 Arranging secure channels for data transfer, and verifying with the data supplier detailed information about the files and metadata and whether the data are suitable for use.

## 3.6 Sub-process: Test statistical business process

- **Sub-process activities:**

3.6.1 Managing and conducting a field or pilot test of the statistical business process. This usually includes collecting data on a small scale to test the collection instruments, followed by processing and analysing the collected data, to ensure that the statistical business process performs as expected. It may be necessary to return to the previous sub-process and make adjustments to the collection instruments, systems, or components of the main statistical business process.

If testing of some of the data being collected is required on a very large scale (for example, a comprehensive trial of the population census process), this testing may be repeated until the process is completed satisfactorily.

## 3.7 Sub-process: Finalise production systems

- **Sub-process activities:**

3.7.1 Putting the assembled and configured processes and services, including modified or newly designed services, into production so that they are ready for use, including:

- Producing documents on the process components, including technical documents and user guides.

- Training system users on how to operate the process.
- Transferring the process components to the production environment and ensuring that they operate as expected in that environment.
- This activity is complementary to the activities of sub-process 3.5 (Test production systems).

■ **Note:** In general, during this phase, the Build phase, the testing processes referred to in the following two sub-processes may be repeated: 3.5 (Test production systems) and 3.6 (Test statistical business process), to configure the statistical workflow procedures before they are ready to be approved for release.

# 04 | Collect phase

In this phase, all required information (data, metadata, and paradata) is collected or acquired using various data collection and acquisition methods (such as traditional and direct collection, data acquisition by obtaining ready-made files or databases from administrative registers, or data extraction by fetching and extracting specific data from technical sources not originally prepared for statistical purposes, such as big data databases, or from web pages (web scraping)). This information is loaded into the appropriate environment in preparation for processing in the next phase, the Process phase.

This phase also includes dataset format validation, with the aim of technically verifying the structure and format of the incoming files and their programmatic validity. However, no derivations or transformations are performed on the collected data during this phase, as they will take place in the Process phase.

This phase is carried out with every iteration of regular statistical products.

**This phase consists of four (4) sub-processes, which are presented sequentially but can occur in parallel and can be iterative:**

1. Create frame and select sample
2. Set up collection
3. Run collection
4. Finalise collection

## Sub-processes and activities in the Collect phase

### 4.1 Sub-process: Create frame and select sample

This sub-process marks the beginning of the operational phase of the statistical product and is concerned with creating the sampling frame and selecting the sample, which take place with each survey cycle (for example: every month or every quarter).

- **Sub-process activities:**

4.1.1 Building the sampling frame and selecting the sample, including:

- Establishing the actual sampling frame and selecting the specific sample for the current collection cycle, based on the scientific parameters and criteria previously established in the Design phase (sub-process 2.4 Design frame and sample).
- Taking into account the nature of the source used, whereby sampling is excluded if the process is based entirely on pre-prepared sources (such as administrative registers or non-survey data). In this case, frames are created directly from the available data and a complete enumeration (census) approach is applied.
- Using variables derived from administrative sources and other non-statistical data as auxiliary variables to support and improve the sampling design and frame classification.

4.1.2 Coordinating samples and managing overlap and response burden, including:

- Coordinating and adjusting the selected samples across cycles and iterations of the same statistical survey (such as managing sample rotation or controlling overlap rates between successive cycles).
- Coordinating and integrating samples across different statistical processes and surveys that share the use of the same frame or central statistical register, with the aim of managing statistical overlap and distributing response burden fairly among establishments or households to prevent repeatedly targeting them.

4.1.3 Governance and quality assurance of the frame and sample, including:

- Conducting quality assurance checks on the final frame to ensure that it is free from coverage errors (such as duplication or omission).
- Reviewing the selected sample, endorsing it, and formally approving it to begin using it in subsequent processes, with complete organizational separation between this activity and the "periodic maintenance and updating of the basic frames/ registers and central repositories" from which sampling frames are created, where maintenance is treated as an ongoing activity independent of the statistical process being implemented.

## 4.2 Sub-process: Set up collection

This sub-process aims to ensure the readiness of data collectors and data providers, as well as the processes and technical systems (such as web-based applications, devices to be used in collection, and communication channels), to collect data and metadata using the methods designed during the Design phase. These activities take place over a period of time, including strategy, planning, and training activities.
The activities of this sub-process are relatively lengthy when the statistical process is implemented once. However, if the statistical process is repeated on a regular basis, some (or all) of these activities may not be required to be implemented in every iteration.

- **Sub-process activities: (For data from survey sources):**

4.2.1 Preparing the collection strategy and plan.
4.2.2 Identifying workforce needs and nominating them for training to meet collection requirements.
4.2.3 Training collection staff.
4.2.4 Establishing a training system that uses supervised machine learning techniques.
4.2.5 Ensuring the availability of collection resources (for example: laptops, collection software, and application programming interfaces).
4.2.6 Agreeing on terms with any intermediary collection entities (for conducting computer-assisted telephone interviews and web service providers).
4.2.7 Configuring collection systems to request and receive data, including location data.
4.2.8 Ensuring the security of the data to be collected.
4.2.9 Preparing data collection instruments (such as printing paper questionnaires, pre-filling them with existing data, loading questionnaires and data onto interviewers' computers, application programming interfaces, and tools for extracting data from websites, "web scraping").
4.2.10 Providing information to respondents (such as drafting letters or brochures explaining the purpose of the survey, and notifying respondents when collection instruments are made available online).
4.2.11 Translating the instruments (for example, into the different languages spoken or used in the country).

- **Sub-process activities: (For data from non-survey sources):**

This sub-process also ensures that the necessary processes, systems, and confidentiality procedures are in place to receive or extract the required information from the source. This includes:

4.2.12 Assessing requests for access to data.
4.2.13 Initiating communication with data-providing entities and sending an introductory package with details on the data acquisition process (a file that clearly explains the rationale for the request and the approved technical and legal mechanism for data transfer).
4.2.14 Verifying with the data provider the detailed information on the files and metadata and obtaining a sample of the data to assess whether the data are suitable for use.
4.2.15 Arranging secure channels for data transfer.

## 4.3 Sub-process: Run collection

This sub-process aims to carry out the actual collection, using different collection instruments to collect or acquire information, which may include raw microdata or aggregates produced at the source, in addition to any accompanying metadata. This sub-process also manages the relationship between GASTAT and respondents and data providers to ensure its positive continuity.

- **Sub-process activities:**

4.3.1 Data collection (for a survey or census), including:

- Sending notifications to respondents and communicating with them for the purpose of completing the required data and carrying out subsequent follow-up and reminder procedures, while documenting and recording contact details (time, method, and response or interaction status).
- Entering data directly at the point of contact, in accordance with the specified collection methodology and method (field CAPI, telephone CATI, self-administered CAWI).
- Performing real-time georeferencing simultaneously with data completion, using inputs from systems in which location can be determined, such as GPS, or directly plotting and identifying the location on the map, and so on, according to the geographic scope and available technology.
- Building and managing a sustainable positive relationship between GASTAT and respondents throughout the collection period to ensure their cooperation, and receiving, recording, and responding to comments, inquiries, and complaints (as reducing the number of non-respondents significantly contributes to improving the quality of the collected data).

■ **Note:** In practice, automated logical checks are performed at the point of collection (validation during data entry); however, these checks are organizationally part of and an ongoing component of sub-process 5.3 (Review and validate).

4.3.2 Obtaining non-statistical data, such as administrative or geospatial data, including:

- Obtaining administrative, geospatial, or other non-statistical data from data providers and data suppliers based on the agreed timelines, or communicating with them to request that the data be sent and following up with them in the event of delays.
- Enabling direct access to and free use of data published under open data licences and available in machine-readable formats without the need for complex data provision procedures.

■ **Note:** Extensive and ongoing technical interactions with data providers and data suppliers are considered part of the overarching activity of "Data supplier management", as it is a broader organizational governance activity.

4.3.3 Monitoring the data collection process, including:

- Continuously monitoring the statistical data collection process by issuing periodic reports and visualizing the data acquisition process (for example, by converting it into charts or maps that show the current status), and adjusting it immediately while the work is ongoing to ensure its quality and suitability for use.
- Conducting basic checks to verify the structure and integrity of the information received as soon as it flows in (such as verifying that the files are in the correct format and contain the expected fields). This is done before allowing it to enter GASTAT's databases or passing it on to subsequent phases.

- Closing the collection process (technically and administratively) as soon as the required targets are achieved, and issuing the final report on the collection process, which focuses primarily on process data (paradata) and operational quality indicators.

## 4.4 Sub-process: Finalise collection

This sub-process is concerned with loading, archiving, and preparing the collected data and metadata for use in the activities of Phase 5, the Process phase.

- **Sub-process activities:**

4.4.1 Loading data and metadata into databases:

- Loading all collected data and metadata into an appropriate electronic environment specifically prepared for initiating subsequent processes before transferring and loading them into a permanent database (information may be loaded before or after verifying its integrity). This may include reloading new versions of information received after the information has previously been compiled and stored. Data may be obtained as long as they are acceptable and received before the collection process is closed.
- Capturing and converting data from paper questionnaires (if any) and entering them electronically, either manually by office staff or automatically using optical character recognition tools to extract the information.
- Processing data files received from other organizations and entities by converting file formats and encoding or coding the received variables to ensure compatibility with GASTAT's internal systems and the classifications used.
- This may also include analysing metadata and paradata associated with collection, which reflect the extent to which collection activities meet the requirements.

4.4.2 Managing and archiving physical data collection assets and instruments, through:

- Archiving physical data collection instruments (for example: paper questionnaires) and disposing of them after storing them electronically if there is no need to perform further processing on them.
- Managing and archiving software tools, such as an application programming interface (API) or other applications, by taking a complete backup of them and storing it in a secure and closed digital repository for reference when needed.

4.4.3 Anonymising the data at the current phase and before passing them on for any subsequent additional processing.

# 05 | Process phase

- **In this phase, the collected data are processed and transformed into a set of statistical outputs for two purposes:**
  1. Statistical analysis that provides a broader understanding of the data.
  2. Dissemination to users using the various rules or models adopted for dissemination.
- This phase is repeated with each statistical production process. It is repeated as regularly produced statistical outputs are repeated.
- The processes and procedures of this phase can be applied to data derived from statistical and non-statistical sources, with the exception of the sub-process (Calculate weights), which is usually specific to survey data.
- This phase, the "Process" phase, and the following phase, the "Analyse" phase, can begin before the Collect phase is completed. This can enable provisional results to be compiled when timeliness is of critical importance to users and increase the time available for their validation.
- Some variables can be processed before others if they are required for data integration (such as geospatial variables if used for this purpose).
- This phase can be carried out in parallel and iteratively with Phase 6, the Analyse phase. Analysis can provide a broader understanding of the data and, consequently, clarify the extent to which additional processing is needed. In some cases, the estimates being processed may be aggregates that have already been disseminated (in accordance with the revision policy).
- The difference between data validation and editing procedures in the Process and Analyse phases:
  - In the Process phase: Validation is performed on the data entered into the models and used to produce the outputs.
  - In the Analyse phase: Validation is performed on the post-processing outputs.

**This phase consists of eight (8) sub-processes that integrate, classify, validate, clean, and transform the input data. These sub-processes are presented sequentially but can occur in parallel and can be iterative:**

1. Integrate data
2. Classify and code
3. Review and validate
4. Edit and impute
5. Derive new variables and units
6. Calculate weights
7. Calculate aggregates
8. Finalise data files

## Sub-processes and activities in the Process phase

### 5.1 Sub-process: Integrate data

In this sub-process, data from one or more sources are integrated. This is the process through which the results of the sub-processes in the Collect phase are integrated. The input data may be a combination of external or internal sources. They may also come from a variety of collection

instruments, including surveys, administrative data, and other non-statistical data sources. Administrative data or other non-statistical data sources may replace all or some of the variables collected directly from the survey. This sub-process also includes standardizing or creating new identifiers that correspond across data sources. This is done to obtain an integrated dataset.

Datasets may need to be linked several times during this "Process" phase with the aim of improving the quality of the input data.

- **Sub-process activities:**

5.1.1 Integrating data, including:

- Integrating data from multiple sources (such as relevant databases and administrative registers) to produce integrated statistics such as national accounts, or aggregating data to improve the efficiency of monitoring rare phenomena.
- Integrating geospatial data with statistical data or non-statistical data.
- Integrating data through integration followed by reduction or replacement. Data from multiple sources are first linked (integration), then duplicates and redundant fields are removed (reduction), or data from the more accurate and higher-quality source replace data from the less accurate source (replacement).
- Carrying out data structuring activities by transforming collected data from their initial form into structured and organized records that comply with the technical rules of the processing system, in preparation for matching and statistical review steps.

5.1.2 Data matching and record linkage, including:

- Matching data for review and validation purposes by integrating or matching multiple datasets to perform data extraction activities, statistical analyses, and validation of the accuracy and completeness of inputs or outputs.
- Record linkage, with the aim of linking micro-level or aggregate-level data from different sources. This is usually carried out through one of the following two mechanisms:
  - Exact matching: The process of linking records from two different sources based on a common, unique, and standardized identifier. Such as: National ID number, establishment number, or tax number).
  - Statistical matching: The process of linking datasets in the absence of a standardized identifier, based on similar characteristics or attributes (such as: sex, age, geographical area, and educational level).

5.1.3 Prioritizing data, including:

- Prioritizing data sources when two or more sources contain different values for the same variable, by following predefined rules and criteria that ensure accurate selection, the most important of which are:
  - Source quality: Including accuracy, completeness, timeliness, and consistency with other sources or related variables.

- Reliability and analytical objective: Taking into account the suitability of the source for the objective; the most accurate source may not be suitable due to a time lag in its release or the presence of bias.
- Approved methodological policies: Adhering to approved internal policies and regulations (such as giving administrative register data first priority, followed by estimation based on statistical surveys).

5.1.4 Anonymising data, including:

- Removing and deleting all direct identifiers of individuals and establishments (such as: names, addresses, national ID numbers, or registration numbers) from integrated administrative records and statistical surveys immediately after the linkage processes are completed, to ensure compliance with laws and legal regulations that protect data confidentiality and privacy.

■ **Note:** The procedures in this sub-process (5.1) are limited to removing direct identifiers from input datasets prepared for processing. This differs methodologically from sub-process 6.4 (Apply disclosure control); the latter focuses on using advanced technical and statistical methods to reduce the risks of indirect identification in final statistical outputs and publications.

## 5.2 Sub-process: Classify and code

This sub-process is concerned with classifying and coding input data according to predefined statistical classifications to facilitate processing and statistical analysis..

- **Sub-process activities:**

5.2.1 Direct classification and coding, through the use of questionnaires and administrative registers that contain option categories that have been pre-coded during the data collection process, to facilitate their capture and processing as they are received.

5.2.2 Post-coding (automated and manual): Processing responses and text after the collection process and converting them into numerical codes, either through interactive automated processing (using artificial intelligence and machine learning algorithms) or through manual office coding when needed.

## 5.3 Sub-process: Review and validate

This sub-process detects actual or potential deviations, problems, errors, and inconsistencies, such as outliers, item non-response, incorrect coding, and duplicate records. This is done according to predefined editing rules, with questionable data being flagged in preparation for processing, without making any actual corrections or modifications to them at this phase.

**The characteristics of this sub-process include the following:**

- It is repeated with each statistical product to validate the data against predefined rules, which are usually applied in a specific order and sequence.
- While validation is treated as part of the "Process" phase, some elements of validation may, in practice, occur alongside activities in the Collect phase, particularly for computer-assisted collection methods.
- While this sub-process is concerned with detecting and locating actual or potential errors, any corrective activities that actually change the data are carried out in sub-process 5.4 (Edit and impute).

- **Sub-process activities:**

  5.3.1 Monitoring and detecting deviations, errors, and outlying or missing data, including:
  - Examining data derived from various sources (before or after integration, or data imputed in sub-process 5.4 (Edit and impute)) to identify problems such as: Outliers, item non-response, incorrect coding, and duplicate records.
  - Flagging data that require automated or manual examination or modification.
  - Applying predefined editing and review rules, taking into account the order and sequence in which those rules are applied, where applicable.

## 5.4 Sub-process: Edit and impute

This sub-process can insert new values or remove incorrect data. This is done when the input data are considered incorrect, missing, unreliable, outdated, or duplicated. The terms editing and imputation encompass a variety of methods for doing so, and this is often carried out using a rule-based methodology.

- **Sub-process activities:**

  5.4.1 Processing deviations, errors, and outlying or missing data, including:
  - Determining whether data will be added or changed.
  - Determining the appropriate methods for this processing.
  - Applying the processing to data values (addition/change/modification). If deviations result from changing or modifying some data, the traditional form of processing is to leave the data unchanged.
  - Writing the new data values back into the dataset (databases) and flagging them as modified data.
  - Creating and documenting metadata for the editing and imputation process, as this enables the data path to be traced and the reason for the change to be identified. This metadata is also used when calculating statistical product quality indicators and shows the extent to which the results rely on original data versus imputed data.

## 5.5 Sub-process: Derive new variables and units

This sub-process is concerned with deriving values for new and necessary variables and units that were not provided directly during the collection process in order to achieve the required outputs, by applying calculation formulas to one or more variables that already exist in the dataset. Or by applying different model assumptions.

Derivation processes are usually applied several times during the "Process" phase, before activities such as data linking, coding, validation, and obtaining outputs, as some derived variables may themselves be based on other derived variables. Therefore, it is important to ensure that variables are derived in the correct order. New units can be derived by aggregating data or dividing them into collection units, or through other different estimation methods.

- **Sub-process activities:t**

5.5.1 Deriving new categories and units, including:

- Deriving new statistical populations from the collected data (for example: persons under 15 years of age with disabilities).
- Deriving values for new statistical units by aggregating or dividing data into collection or reporting units, or through various other aggregation methods (for example: deriving households when the collection units are persons, and aggregating 7 barrels of oil to equal 1 ton).

5.5.2 Deriving new variables, including:

- Deriving values for new variables necessary to aggregate the required outputs or validate them by applying calculation rules to one or more variables that already exist in the dataset.
- Applying derivation processes to datasets before or after data linking, as some derived variables may themselves be based on other derived variables.
- Ensuring that variables are derived in the correct order, as follows:

1- Ensuring the accuracy and quality of the original variables.
2- Processing outlying or missing values, if any.
3- Deriving new variables.
4- Checking consistency between variables.

Examples of deriving new variables include: Deriving the age variable from the two variables of date of birth and date of interview, deriving monthly income from weekly or annual income, and deriving the number of persons under 10 years of age from a list of persons and their ages.

.

## 5.6 Sub-process: Calculate weights

This sub-process is concerned with creating weights for sampling units according to the methodology established in sub-process 2.5 (Design processing and analysis).

- **Sub-process activities:**

5.6.1 Creating initial weights for sample units to be used to gross up the data and make them representative of the target population (as in sample surveys or scanner data extracts). This is based on the probability of selecting each unit in the sample. (Example: If the probability of selecting a person from the population is 11000/, their initial weight will be 1000).

5.6.2 Adjusting weights for non-response (if any) to compensate for units or categories that did not respond (in a survey or complete enumeration). (Example: If a particular category of the population records a low response rate, the weight is adjusted by increasing the weight of responding individuals from the same category).

5.6.3 Weighting and calibrating weights (if required) by adjusting them so that the overall characteristics of the sample match known and reliable benchmark characteristics of the population (such as the population census for household surveys, or administrative registers such as the number of employees or capital registered in economic or industrial registers for establishment surveys).

## 5.7 Sub-process: Calculate aggregates

This sub-process is concerned with producing aggregate data and population estimates by aggregating microdata (or lower-level aggregates) and applying weights and descriptive statistical measures (such as measures of dispersion and measures of central tendency) to summarize the phenomenon and provide an accurate picture of it before conducting analysis, in addition to calculating sampling errors associated with estimates for sample surveys.

- **Sub-process activities:**

5.7.1 Aggregating microdata and calculating statistical measures, including:
- Collecting and aggregating records that share specific characteristics (such as demographic or geographical classifications) from microdata or lower-level aggregates.
- Calculating descriptive statistical measures for different groups.

5.7.2 Applying the weights calculated in sub-process 5.6 (Calculate weights) to the aggregated data to gross them up and derive the appropriate overall population aggregates and estimates.

5.7.3 Calculating and determining sampling errors and calculating the corresponding confidence intervals for the relevant statistical aggregates and estimates (in the case of outputs based on sample surveys).

## 5.8 Sub-process: Finalise data files

This sub-process is concerned with compiling and linking the results of all previous sub-processes within the Process phase to create a complete dataset (whether microdata or aggregate data).

These files are the direct input to the "Analyse" phase and may be produced as intermediate data files for releasing preliminary estimates when there are time constraints, or as final files for approval and dissemination.

- **Sub-process activities:**

  5.8.1 Compiling and creating data files, by compiling and integrating the outputs of the processing operations and their metadata into a unified dataset (microdata or aggregate data), and storing them in the approved data environment to serve as a direct input to the "Analyse" phase, whether as intermediate files for releasing preliminary estimates or as final files for analysis and preparation of outputs.

# 06 | Analyse phase

This phase is concerned with producing and reviewing statistical outputs in detail and validating them against expectations. The relevant documents are prepared and the content is prepared for dissemination. This phase includes the sub-processes and activities that enable statistical analysts to understand and interpret the resulting outputs.

The outputs of this phase can also be used as inputs to sub-processes in other phases (for example, using the analysis of new sources as an input to the "Design" phase when the statistical process is repeated the next time).

This phase is carried out with each cycle of statistical outputs produced regularly or on a one-time basis, and its outputs include: Statistics, related metadata, and commentary (texts, presentations, and visualizations), whether for internal release or dissemination.

The "Analyse" phase includes all statistical outputs, regardless of how the data were obtained.

**This phase consists of six (5) sub-processes, which are presented sequentially but can occur in parallel and can be iterative:**

1. Prepare draft outputs
2. Validate outputs
3. Interpret and explain outputs
4. Apply disclosure control
5. Finalise outputs

## Sub-processes and activities in the Analyse phase

### 6.1 Sub-process: Prepare draft outputs

This sub-process is concerned with transforming data from sub-processes 5.7 (Calculate aggregates) and 5.8 (Finalise data files) into statistical outputs.

- **Sub-process activities:**

6.1.1 Transforming the data resulting from sub-processes 5.7 (Calculate aggregates) and 5.8 (Finalise data files) into statistical outputs (such as indicators, seasonally adjusted statistics, and microdata).

6.1.2 Preparing maps, geographic information system (GIS) outputs, and statistical geospatial services to maximize the value and ability to analyse statistical information.

6.1.3 Preparing interpretations and methodological notes accompanying the outputs that describe how they were produced based on the methodology designed in the Design phase. In addition, preparing any other interpretations to communicate/explain the necessary information to users when new processing or analysis methods are used (such as methods based on machine learning or artificial intelligence).

## 6.2 Sub-process: Validate outputs

This sub-process is concerned with statisticians validating the produced outputs against the overall quality framework and expectations, comparing them with available supporting information (such as previous cycles and other sources), and applying accumulated experience and knowledge of the statistical domain to identify deviations and check for errors as objectively as possible.

- **Sub-process activities:**

6.2.2 Detecting deviations and checking the consistency of outputs, including:

- Comparing statistics with previous cycles and checking coverage rates and response rates.
- Comparing results with relevant internal and external data and examining inconsistencies (including the geospatial consistency of data).
- Conducting macro editing and using aggregates as a tool to validate microdata.
- Validating statistics and identifying any deviations by comparing them with expectations and specialized knowledge of the statistical domain.
- Checking the availability of metadata, paradata, and quality indicators and their consistency with expectations.

6.2.2 Processing and resolving deviations, including:

- Processing outputs that differ from expectations to ensure their suitability for release, or leaving them unchanged if the deviation can be explained and methodologically justified.
- Applying predefined methods to resolve aggregate-level deviations.
- Documenting all technical decisions related to accepting or resolving deviations in outputs, and building a knowledge record for the statistical domain to serve as a technical reference in future cycles.

## 6.3 Sub-process: Interpret and explain outputs

This sub-process is concerned with interpreting statistical outputs by reviewing statistics from different perspectives and using different tools, and conducting in-depth and econometric analyses to evaluate the results against expectations and understand how they were produced.

- **Sub-process activities:**

6.3.1. In-depth statistical review and analysis, including:

- Reviewing statistics and outputs from multiple perspectives and angles using different tools and presentation media.
- Conducting various types of in-depth analysis, such as time series analysis and econometric analyses, to explore trends and understand how the results were produced.

6.3.2. Evaluating results and preparing interpretations, including:

- Evaluating results and indicators and comparing them against initial expectations to obtain a clear understanding of statistical phenomena.
- Formulating comprehensive and objective interpretations that explain the outputs and preparing them for dissemination and use by decision-makers and beneficiaries.

## 6.4 Sub-process: Apply disclosure control

This sub-process ensures that the data and metadata intended for dissemination or internal storage for future use do not violate confidentiality rules in accordance with GASTAT's dissemination policies or the specific process methodology established in sub-process 2.5 (Design processing and analysis).
It may include primary and secondary disclosure checks, as well as applying data anonymisation or perturbation techniques and checking outputs.

- **Sub-process activities:**

6.4.1 Monitoring and detecting confidential data, including:

- Examining data and metadata (prepared for dissemination or internal storage) to verify that confidentiality rules are not violated in accordance with institutional policies and the methodology designed in sub-process 2.5 (Design processing and analysis). The examination includes geospatial data outputs to protect personal data.
- Conducting primary and secondary disclosure checks using appropriate examination and analysis methods.

6.4.2 Processing confidential data and applying disclosure control, including:

- Using data suppression or perturbation methods to anonymise data when they are disclosed and disseminated.

- Taking into account differences in the degree and method of disclosure control according to the nature of the output (the approach used for microdata datasets for research purposes will differ from that used for published tables or aggregate-level data visualizations).
- Examining the final outputs after applying disclosure control techniques to ensure the effectiveness of protection and the suitability of the data for dissemination.

## 6.5 Sub-process: Finalise outputs

This sub-process ensures that statistics and their accompanying information are suitable for use, fit for purpose, and ready for dissemination in accordance with the required quality standards.

- **Sub-process activities:**

6.5.1 Final verification of outputs, including:
- Conducting final checks on the overall consistency of statistics before their approval.
- Preparing and compiling interpretations, analytical commentary, technical notes, necessary metadata, and measures of uncertainty (Standard Error - SE, Coefficient of Variation - CV, Confidence Intervals – CI, and Margin of Error - MoE).
- Producing reports and internal documents supporting the statistical process and its outputs.
- Translating statistical outputs and accompanying metadata into English.

6.5.2 Approving outputs, including:
- Discussing draft final outputs with relevant internal subject-matter experts.
- Determining the level and controls for release and dissemination, and applying technical reservations and notes to the results (technical and methodological limitations of the data to avoid misinterpretation or inaccurate use).
- Approving the statistical content and approving its release and dissemination.

# 07 | Disseminate phase

This phase is concerned with managing the release of statistical data and products and making them available to users. It covers activities related to assembling and releasing various static and dynamic products through multiple channels, while providing support to users in accessing and using the statistical products released by GASTAT.
This phase relies primarily on indexing and classifying published data and products using relevant metadata standards (particularly international standards as a general rule) to increase their discoverability, accessibility, and ease of use (or benefit from them) internally and externally and by a wide range of users. It also includes linking data to geospatial classifications to

enable the detailed geographical presentation of outputs, while emphasizing the importance of archiving final outputs in this phase (although archiving is also part of the overarching data management and metadata management activities).
This phase is carried out with each cycle of statistical outputs produced regularly or on a one-time basis.

**This phase consists of five (5) sub-processes, which are presented sequentially but can occur in parallel and can be iterative:**

1. Update output systems
2. Produce dissemination products
3. Release dissemination products
4. Promote dissemination products
5. Provide user support

## Sub-processes and activities in the Disseminate phase

### 7.1 Sub-process: Update output systems:

This sub-process manages the updating of systems (such as databases) where data and metadata ready for dissemination purposes are stored.

- **Sub-process activities:**

7.1.1 Preparing and loading outputs into public dissemination databases, including:
- Formatting and preparing data and metadata for placement in public dissemination databases.
- Loading data and metadata into public dissemination databases.
- Ensuring that data are linked to their related metadata.

■ **Note:** It is often preferable for metadata to be formatted, loaded, and linked in the earlier phases, but this activity includes a final check to ensure that all necessary metadata are in place and ready for dissemination.

### 7.2 Sub-process: Produce dissemination products

This sub-process assembles the products previously designed in sub-process 2.1 (Design outputs and dissemination) to meet users' needs. It also prepares the related promotional materials, which may include printed publications, press releases, and websites. The products can take many forms, including interactive charts and tables, maps, microdata datasets for public use, linked open data, and downloadable files. When carrying out this sub-process, additional disclosure control may be applied through sub-process 6.4 (Apply disclosure control).

- **Sub-process activities:**

7.2.1 Preparing and assembling components of regular products:

- Assembling the components of products provided to beneficiaries and released regularly (for example: Explanatory texts and specific analyses, data tables, time series, charts, data visualizations, quality data, etc.).

7.2.2 Preparing and assembling components of custom-designed products:

- Assembling products specifically designed to meet the needs of specific clients (for example: explanatory texts, tables, charts, etc )

■ **Note:** In some cases, custom-designed products are prepared and assembled for release and provided specifically to the beneficiary entity without being disseminated by GASTAT.

7.2.3 Preparing promotional materials:

- Preparing and implementing promotional plans for products, which include studying the target audience and the outreach strategy for the purpose of promoting them.
- Studying the appropriate means for each product according to its level of importance.

7.2.4 Reviewing the products and verifying that they meet dissemination and publication standards, and approving them in their final form for release and dissemination.

## 7.3 Sub-process: Release dissemination products

This sub-process ensures that all elements of the release are ready and complete, including managing the timing of the release. It includes briefings for specific groups, such as the press or relevant decision-makers, and facilitating access to statistical content in public dissemination databases through multiple dissemination channels for different beneficiaries on different devices, in addition to arrangements for any embargoes prior to release. It also includes managing access to microdata by authorized user groups in accordance with GASTAT's access policies. This sub-process also covers handling cases that require the withdrawal of a product (for example, when an error is detected in the outputs or when preliminary estimated data are updated to final data).

- **Sub-process activities:**

7.3.1 Notification and coordination with relevant stakeholders prior to release, including:

- Aligning outputs with stakeholders for lifting the embargo on the dissemination of data and products.
- Coordinating with partner entities in cases requiring the joint release of statistical products.
- Providing briefings to specific groups, such as the press or relevant decision-makers.

7.3.2 Releasing products and services, including:

- Approving and authorizing statistical content in preparation for public dissemination.
- Releasing statistical content through the designated dissemination channels.

- Managing and regulating access by authorized users (such as researchers) to microdata in accordance with the approved controls and user permissions.
- Adhering to the previously announced release date of the statistical product on the statistical calendar on GASTAT's official website.

7.3.3 Archiving and updating published content, including:

- Ensuring that published content is successfully stored and archived and that its continued availability is maintained.
- Handling cases of periodic updates or amendments to data (such as replacing preliminary estimates with final data).
- Implementing procedures for withdrawing statistical products from dissemination channels when errors are detected in the data and correcting them immediately.

## 7.4 Sub-process: Promote dissemination products

This sub-process is concerned with the effective promotion of statistical products and services resulting from the statistical process to ensure that they reach the widest possible range of beneficiaries.

- **Sub-process activities:**

7.4.1 Using customer relationship management (CRM) tools to better target potential users of the products.

7.4.2 Using means such as websites, wikis, and other platforms, such as social media, to facilitate the communication of statistical information to users.

■ **Note:** These activities may be coordinated with partner entities in the case of joint products.

## 7.5 Sub-process: Provide user support

This sub-process includes the processes for collecting and reviewing users' inquiries and requests for data and services, in order to respond to their inquiries and meet their requirements.

- **Sub-process activities:**

7.5.1 Receiving users' inquiries and requests, including:

- Receiving users' inquiries and requests for data and information related to GASTAT's products and services.
- Receiving requests for access to microdata.
- Regularly reviewing these inquiries and requests to provide inputs to the overall quality management process, as they may indicate new or changing user needs.
- Monitoring interaction, measuring customer feedback, and reporting it to the relevant management to make use of it in improvement processes.

7.5.2 Responding to customers' queries, including:
- Responding to users' inquiries shortly after release (for example, questions about released statistics, the availability of additional data, etc.).
- Processing inquiries and providing users with accurate responses and answers promptly or within the agreed timelines.
- Recording inquiries to support the Evaluate phase.
- Handling subsequent inquiries, including requests for custom outputs, products, or services, through customer relationship management and support.

7.5.3 Using responses to questions and requests to feed and update the knowledge base or the "Frequently Asked Questions" page available to users, with the aim of improving response efficiency and reducing the burden of repeated or similar requests.

7.5.4 Providing and managing support for partner entities that may participate in disseminating statistical products.

## 08 | Evaluate phase

This phase is concerned with evaluating a specific instance of a statistical business process and the overall and overarching quality management activity described in the "Overarching activities". It includes evaluating the success of a specific cycle of the statistical business process using a set of quantitative and qualitative inputs collected during the production phases, analysing them, and identifying and prioritizing potential improvements to ensure product quality.

The processes of this phase can be carried out at the end of the Disseminate phase and can also be carried out continuously during the statistical production process.

This phase is carried out with each cycle of statistical outputs produced regularly. However, in the case of regular and fully stable statistical business processes, the evaluation may not be formally conducted in every cycle.

This phase can be viewed, at least theoretically, as providing a decision point on the following after evaluating statistical outputs that are produced regularly:

- Determining whether future iterations of the statistical process should be carried out.
- Determining whether any improvements should be implemented if it is decided to repeat the process.
- Determining whether the next iteration should begin from the "Specify Needs" phase or from another phase (most likely the "Collect" phase).

**This phase consists of three (3) sub-processes, which are presented sequentially but can occur in parallel and can be iterative::**

1. Gather evaluation inputs
2. Conduct evaluation
3. Agree an action plan

## Sub-processes and activities in the Evaluate phase

### 8.1 Sub-process: Gather evaluation inputs

This sub-process aims to compile all forms of evaluation inputs, including: (User feedback and opinions, process metadata, system metrics, staff suggestions, and progress reports on action plans agreed upon during a previous cycle, which may also provide inputs to evaluations of subsequent cycles), and make them available to the team that will conduct the evaluation.
To evaluate specific processes, it may be necessary to carry out specific activities, such as small surveys (such as post-enumeration surveys, re-interview studies, and a survey on the effectiveness of dissemination).

- **Sub-process activities:**

8.1.1 Producing evaluation materials in any other phase or sub-process by collecting quantitative evaluation inputs, through:
- Collecting a set of quantitative information or measurable inputs to enhance the evaluation process. Such as metrics or measures related to the quality of outputs, including: The effects of imputation for missing values, revisions, trends in relative standard errors, etc.).

8.1.2 Producing evaluation materials in any other phase or sub-process by collecting qualitative evaluation inputs, through:
- Collecting qualitative or textual information using a variety of methods (for example: User meetings, internal meetings to discuss outstanding issues, meetings, workshops, etc.).
- This information is collected from:
  a. Production staff. b. Users.
  c. Data providers. d. Service providers.

### 8.2 Sub-process: Conduct evaluation

In this sub-process, evaluation inputs are analysed and compiled into the report prepared for this purpose or a monitoring and control dashboard. The evaluation report includes an explanation of any quality issues specific to this cycle of the statistical business process, any

deviations of performance measures from expected values, as well as recommendations for changes where appropriate. These recommendations may address changes to any process in future iterations or may indicate that the project should not be repeated. The action plan is then presented to the relevant entity to determine priorities and proceed.
The evaluation can also be conducted at the end of the statistical process (post-evaluation) for the entire process or selected activities, or continuously during the implementation of the statistical process, allowing for immediate corrections or continuous improvement.

- **Sub-process activities:**

8.2.1 Analysing evaluation inputs, including:

- Thoroughly analysing the evaluation inputs compiled in sub-process 8.1 (Gather evaluation inputs).
- Comparing the outputs of the analysis with the expected results of the project to determine whether they have been achieved.

8.2.2 Identifying improvements, including:

- Identifying a set of potential improvements or solutions to the problems revealed by the analyses.
- Discussing the identified improvements with the relevant parties before presenting them to the authorized decision-maker.

■ **Note:** Improvements can be made to processes, methodologies and systems, individuals' skills, standards, and statistical frameworks.

8.2.3 Providing instructions on data quality to data providers, including:

- Providing data providers with feedback and information with the aim of improving data quality.

8.2.4 Measuring user satisfaction, including:

- Establishing specific mechanisms for measuring user satisfaction and identifying the elements of satisfaction.
- Coordinating with clients to obtain their views on the product through direct communication or workshops.
- Measuring the level of users' satisfaction with the statistical product.
- Dealing with dissatisfied users and clarifying the situation to them.
- Analysing the results and presenting them to the relevant management for the purpose of improvement.
- Measuring the level of satisfaction of the statistical community.

■ **Note:** A customer satisfaction survey can be conducted for a number of products or for a specific product.

8.2.5 Measuring users' use of the indicator:

- Measuring the usage performance of released statistical indicators and the extent to which these indicators achieve users' objectives and requirements.

8.2.6 Preparing the evaluation report, including:
- Documenting the results of the evaluation analyses.
- Preparing the evaluation report based on the documented analysis results.
- Providing recommendations for improvements.

## 8.3 Sub-process: Agree an action plan

- **Sub-process activities:**

8.3.1 Based on the evaluation and recommendations for improvement, a proposed action plan is presented to the authorized decision-maker for discussion, prioritization, and approval.

- **Note:** It is important to take into account monitoring the impact of these planned changes and improvements on the statistical process and its results, which may provide input to evaluations of future iterations of the statistical process.

## Twelfth: List of Abbreviations

| | | |
|---|---|---|
| GSBPM | Generic Statistical Business Process Model | Generic Statistical Business Process Model: A flexible tool for describing and defining the set of business processes needed to produce official statistics. |
| GAMSO | Generic Activity Model for Statistical Organisations | Generic Activity Model for Statistical Organizations: A reference framework that describes and defines the activities carried out within a typical statistical organization. |
| GSIM | Generic Statistical Information Model | Generic Statistical Information Model: An information reference framework that provides generic descriptions for defining, managing, and using data and metadata throughout all phases of the statistical production process. |
| GPS | Global Positioning System | Global Positioning System |
| HLG-MOS | High-Level Group for the Modernisation of Official Statistics | High-Level Group for the Modernisation of Official Statistics |
| ICT | Information and Communications Technology | Information and Communications Technology |
| UNECE | United Nations Economic Commission for Europe | United Nations Economic Commission for Europe |
| CAPI | Computer-Assisted Personal Interviews | Computer-Assisted Personal Interviews |
| CATI | Computer-Assisted Telephone Interview | Computer-Assisted Telephone Interview |
| CAWI | Computer-Assisted Web Interviewing | Computer-Assisted Web Interviewing |
| AI | Artificial Intelligence | Artificial Intelligence |
| API | Application Programming Interface | Application Programming Interface A protocol that enables different systems and software to communicate and exchange data with each other automatically. |
| ISO | International Organization for Standardization | International Organization for Standardization |

STATS.GOV.SA