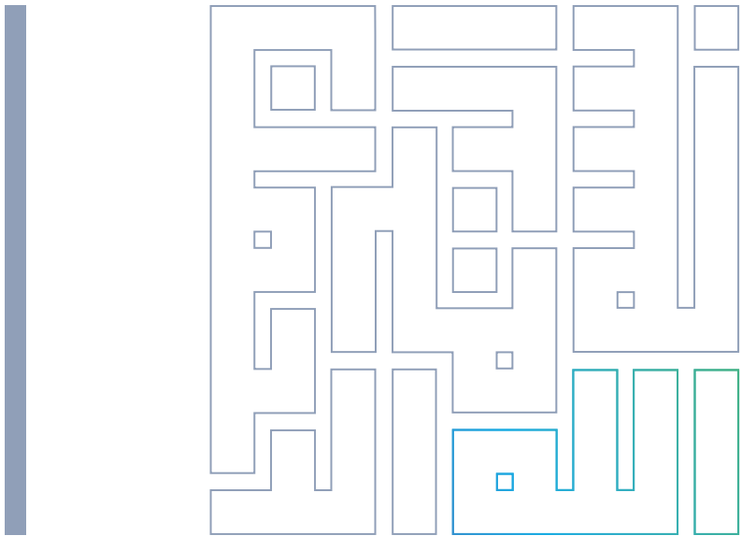
الهيئة العامة للإحصاء
General Authority for Statistics

Generic Statistical Business Process Model (GSBPM)

National Version - 2024
Kingdom of 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
5 Third: Background
5 Fourth: GSBPM Objectives and Usage
6 Fifth: GSBPM Development Methodology
7 Sixth: GSBPM Structure
9 Seventh: Applicability
10 Eighth: Phases of statistical business

11 **1. First phase: Specify needs**

11 1.1 First sub-process: Investigate needs
12 1.2 Second sub-process: Consult and confirm needs
12 1.3 Third sub-process: Establish output objectives
12 1.4 Fourth sub-process: Identify concepts
12 1.5 Fifth sub-process: Check data availability
13 1.6 Sixth sub-process: Prepare and submit business case

14 **2. Second Phase: Design**

14 2.1 First sub-process: Design outputs
15 2.2 Second sub-process: Design variable descriptions
16 2.3 Third sub-process: Design collection
17 2.4 Fourth sub-process: Design frame and sample
18 2.5 Fifth sub-process: Design processing and analysis
19 2.6 Sixth sub-process: Design production systems and workflow

20 **3. Third Phase: Build**

20 3.1 First sub-process: Reuse or build collection instruments
21 3.2 Second sub-process: Reuse or build processing and analysis components
21 3.3 Third sub-process: Reuse or build dissemination components
21 3.4 Fourth sub-process: Configure workflows
21 3.5 Fifth sub-process: Test production systems
22 3.6 Sixth sub-process: Test statistical business process
22 3.7 Seventh sub-process: Finalize production systems

23 **4. Fourth Phase: Collect**

23 4.1 First sub-process: Create frame and select sample
24 4.2 Second sub-process: Set up collection
25 4.3 Third Sub-process: Run Collection
26 4.4 Fourth sub-process: Finalize collection

Contents

28 **5. Fifth Phase: Process**

28 5.1 First sub-process: Integrate data
30 5.2 Second sub-process: Classify and code
30 5.3 Third sub-process: Review and validate
31 5.4 Fourth sub-process: Edit and impute
32 5.5 Fifth sub-process: Derive new variables and units
32 5.6 Sixth sub-process: Calculate weights
33 5.7 Seventh sub-process: Calculate aggregates
33 5.8 Eighth sub-process: Finalize data files

34 **6. Sixth phase: Analyze**

34 6.1 First sub-process: Prepare draft outputs
34 6.2 Second sub-process: Validate outputs
36 6.3 Third sub-process: Interpret and explain outputs
36 6.4 Fourth sub-process: Apply disclosure control
37 6.5 Fifth sub-process: Finalize outputs

38 **7. Seventh Phase: Disseminate**

38 7.1 First sub-process: Update output systems
39 7.2 Second sub-process: Produce dissemination products
40 7.3 Third sub-process: Manage release of dissemination products
41 7.4 Fourth sub-process: Promote dissemination products
41 7.5 Fifth sub-process: Manage user support

42 **8. Eighth Phase: Evaluate**

42 8.1 First sub-process: Gather evaluation inputs
43 8.2 Second sub-process: Conduct evaluation
45 8.3 Agree an action plan:

45 Ninth: Overarching Processes

45 9.1 Quality Management
47 9.2 Metadata Management
49 9.3 Data Management

50 Tenth: List of Acronyms

First: Introduction

Proceeding from GASTAT mission to provide updated, value-added, accurate, comprehensive and reliable statistical products and services based on the best international standards and practices; to take the lead in developing the statistical sector in support of decision-taking; and to realize its vision to be the most distinctive and innovative statistical reference in support of socio-economic growth of KSA, the GSBPM represents one of the key enablers to achieve this objective. GSBPM refers to the set of Processes and activities that statistical processes require in order for GASTAT statistical departments to provide high-quality, accurate, and complete statistics based on state-of-art worldwide best practices. The Model also serves as a means of organizing the statistical sector and shaping its operations as specified in Article 4 (2) of Council of Ministers Resolution No. 11 of 13 January 1437 AH on Approving the Organization of the General Authority for Statistics, which states: «GASTAT is the sole authority entrusted with carrying out, supervising and regulating statistical activities. To this end, it shall be vested with the following competencies: Carry out statistical activities in conformity with internationally recognized standards, including: Identify the statistical methodology».

GASTAT, being a member of the worldwide statistical community, incorporated international recommendations and practices into the National Version of GSBPM while keeping in mind its organizational environment.

Second: Terms and Definitions

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

1. **Authority:** General Authority for Statistics.
2. **Statistical Sector:** The General Authority for Statistics being the central body responsible for the entire statistical activities. It also comprises the integrated matrix of all departments, units and statistical divisions from all sectors and institutions within the state.
3. **The Model :** The (National Version) of the Generic Statistical Business Process Model.
4. **International Model :** Version (5.1) of the Generic Statistical Business Process Model (GSBPM) published by the United Nations Economic Commission for Europe (UNECE) in January 2019.
5. **Administrative Records:** Statistical-related paper or electronic records in which data or information is documented at all other entities, as well as those related to community conditions and activities at large.

6. **Statistical Product:** A structured statistical activity based on a series of scientific phases that aims to provide reliable statistical data on certain community features, whether by sampling or deep inventory.
7. **Statistical Assets:** Statistical data and information owned by GASTAT.
8. **Quality Portal:** Monitoring the accuracy of statistical data published to the website.
9. **Databases:** Storing statistical data using software and computer systems.
10. **Governance:** A set of rules and processes for managing and monitoring projects, as well as for organizing relationships between relevant stakeholders.
11. **Authorized Person:** The final decision-maker according to the governance models adopted at GASTAT.

Third: Glance

GASTAT establish this version to be a national version that is highly compatible with the GSBPM (5.1) issued in January 2019 by the United Nations Economic Commission for Europe (UNECE).

This national version is expected to witness further updates over the coming years to cope with the developments and models introduced by the High-Level Group for the Modernization of Official Statistics (HLG-MOS) such as the Generic Statistical Business Process Model (GSBPM).

Fourth: GSBPM Objectives and Usage

1. Improve and standardize the statistical processes used in the creation of statistics within both GASTAT and the statistical sector.
2. Facilitate collaboration and coordination across departments engaged in the statistics generating processes.
3. Provide a framework for assessing the quality of the statistical process, as well as a mechanism to facilitate the standardization of operational standards to the greatest extent possible, thereby promoting a culture of quality, accuracy, comprehensiveness, and the potential for highly productive statistics.
4. GSBPM enables the integration of international efforts on statistical metadata and data quality by providing a common framework and nomenclature to describe statistical processes.
5. Methodological criteria can be linked to the relevant phases or sub-processes and then classified and stored in a GSBPM-based structure.

6. Provide a tool to review and improve statistical classifications.
7. A starting point to develop processes aimed at deploying new products.
8. Enable governance of statistical activities and processes.
9. Create a professional and training resource for GASTAT›s and the statistical sector›s personnel.
10. Estimate the cost of different elements of the statistical process to determine the operational expenses, which can then be used to identify modernization efforts to enhance cost efficiency of the most expensive parts of the process.
11. Measure system performance by identifying components that are inefficient, redundant, or need to be replaced, and identifying gaps for which new components should be developed.
12. Provide ongoing assessment and development tools.
13. Provide a tool for harmonizing processes associated with non-statistical data providers (such as administrative and geospatial data), facilitate communication between statisticians and experts from other areas and coordinate related terminology.

Fifth: GSBPM Development Methodology

As previously indicated, the National Version was created in order to be compatible with GSBPM v. (5.1). Certain details, however, were added to provide more explanation. Among the most important factors considered in developing this GSBPM version of the model are the following:

- The international GSBPM v. (5.1) in its entirety serves as a basis and a starting point for developing and producing this National Version.
- Naming the phases in line with the international model.
- Structural design of phases in line with the international model.
- The «phases» in the National Version, like the international version, are made up of «sub-processes».
- The set of «activities» under each sub-process in the international version was considered a third level in the National Version, with the inclusion of certain activities not found in the international version. The work will be made consistent with the results of the UNECE «GSBPM Tasks» work-group.

Sixth: GSBPM Structure

The National Version of GSBPM consists of four levels, as follows:

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

Although each sub-process in the international version contains a set of activities, it has not been treated as a third level. It did not have the same detail and clarification adopted as the National Version.

- The statistical business phases of the Model are broken down into eight related phases as follows:

Statistical business processes (Overarching Processes)															
Specify needs		**Design**		**Build**		**Collect**		**Process**		**Analyze**		**Disseminate**		**Evaluate**	
1.1	Identify needs	2.1	Design outputs	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 descriptions	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	Manage release of dissemination products	8.3	Agree an action plan
1.4	Identify concepts	2.4	Design frame and sample	3.4	Configure workflows	4.4	Finalize collection	5.4	Edit and impute	6.4	Apply disclosure control	7.4	Promote dissemination products		
1.5	Check data availability	2.5	Design processing and analysis	3.5	Test production systems			5.5	Derive new variables and units	6.5	Finalize outputs	7.5	Manage user support		
1.6	Prepare and submit business case	2.6	Design production systems and workflow	3.6	Test statistical business process			5.6	Calculate weights						
				3.7	Finalize production systems			5.7	Calculate aggregates						
								5.8	Finalize data files						

Figure 1: The phases (level 1) and sub-processes (level 2) of the GSBPM

- 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)

- Phases and sub-processes are designed in a flexible manner. Some elements of the Model may fit a type of statistics more than others, while some elements may overlap in another product and be replicated until final approval is reached.
- The Model contains (8) phases and (44) sub-processes. The Model also describes a number of activities under each process, which describe the sub-processes in detail.

In addition to these levels, which are explained in Section Eight of this Version, Section Nine has been dedicated to the overarching processes that apply throughout the eight phases. These overarching processes include quality management, metadata management and data management:

1) Quality Management - This process includes quality assessment and control mechanisms. It recognizes the importance of evaluation and feedback throughout the statistical business process.

2) Metadata Management - Metadata are created/reused and processed within each phase, there is, therefore, a strong requirement for a metadata management system to ensure the appropriate metadata retain their links with data throughout the GSBPM. This includes process-independent considerations such as metadata custodianship and ownership, quality, archiving rules, preservation, retention and disposal.

3) Data Management - This includes process-independent considerations such as general data security, custodianship and ownership, data quality, archiving rules, preservation, retention and disposal.

- Process Data Management - This includes activities of registering, systematizing and using data about the implementation of the statistical business process. Process data can aid in detecting and understanding patterns in the data collected, as well as in evaluating the execution of the statistical business process as such.

  - Knowledge Management - This ensures that statistical business processes are repeatable, mainly through the maintenance of process documentation.
  - Resource Management - This includes cross-process burden management, as well as topics such as profiling and management of contact information; and thus has particularly close links with statistical business processes that maintain registers.

Seventh: Applicability

1. To produce a statistical product for the first time, all eight phases must be used:

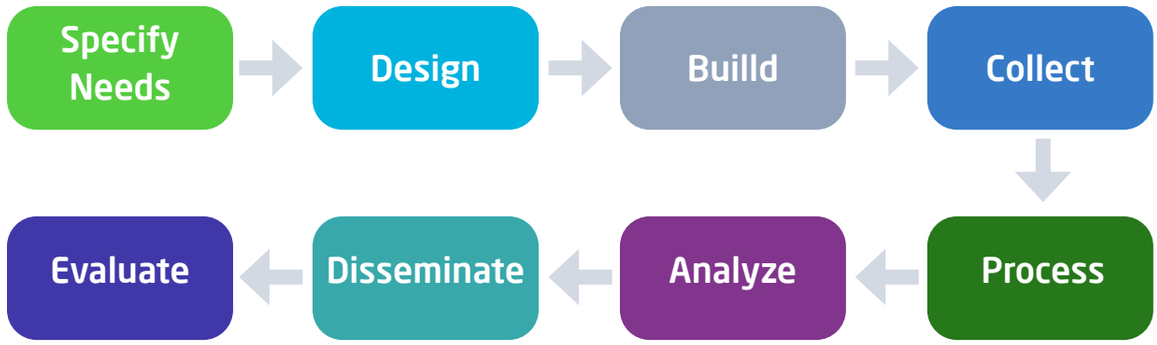


Figure (2): Phases of statistical business

2. Four phases are reviewed upon re-release of the statistical product:
   - Fourth Phase (Collection)
   - Fifth Phase (Processing)
   - Sixth Phase (Analysis)
   - Seventh Phase (Dissemination)
3. Each time a statistical product is produced, the number of reviews of the following phases should be determined:
   - First Phase (Specify Needs)
   - Second Phase (Design)
   - Third Phase (Build)
4. GSBPM applies several overarching statistical processes throughout the eight phases, as well as other details pertaining to quality management, metadata management, and data management. Also overarching processes will be updated following specifications released by the UNECE Supporting Standards group.
5. The Model is designed to be applicable regardless of the data source, so it can be used for the description and quality assessment of processes based on surveys, censuses, administrative registers, and other non-statistical or mixed sources.
6. The sectioning process used in the Model allows the phases to be applied sequentially or in parallel depending on the type of statistical product.

Eighth: Phases of statistical business

Phase		Description
First	Specify needs	Understand and confirm statistical needs, and identify possible solutions.
Second	Design	Design all subsequent sub-processes of the statistical process.
Third	Build	Build, organize, gather and test production solutions and systems so that they are ready for use in statistical implementation processes.
Fourth	Collect	Collect and validate data.
Fifth	Process	Process data to be converted from raw to usable statistical data.
Sixth	Analyze	Validate and interpret outputs.
Seventh	Disseminate	Prepare the products to be ready for dissemination.
Eighth	Evaluate	Conduct an evaluation of the statistical process.

01 | First phase: Specify needs

This phase aims to:

1. Understand the context surrounding the specific need or change.
2. Identify customer needs and requirements.
3. Set clear limits and targets for the statistical product.
4. Identify the range of solutions to meet needs.

This phase is triggered in two cases:

1. A need for new statistics is identified.
2. Amendments to (current) statistical product initiates a review.

This phase consists of (6) sub-processes:

- First sub-process : Specify needs
- Second sub-process : Consult and confirm needs
- Third sub-process : Establish output objectives
- Fourth sub-process : Identify concepts
- Fifth sub-process : Check data availability
- Sixth sub-process : Prepare and submit business case

Sub-processes and activities in the First Phase: Specify needs

1.1 First sub-process: Identify needs

This process is based on understanding customers› needs for the statistical product and reviewing observations and feedbacks on current statistics through certain procedures so that the actual need for the statistical product and the feasibility of its production or improvement in case the product exists, are identified.

- **Activities of the Process:**

1.1.1 Investigate and understand the specific statistical needs in order to consult with the customer.
1.1.2 Verify the background of the customers, the entity they represent and the reasons for the product›s demand or improvement if the product exists.
1.1.3 Identify the required statistics and objectives.
1.1.4 Identify priorities and challenges, and propose solutions to the customer.
1.1.5 Consider available statistical assets that might meet a customer›s need. Consider of practice amongst other (national and international) statistical organizations producing similar data and the methods used by those organizations.
1.1.6 Provide action plans from evaluations of previous iterations of the process or from other processes.

1.2 Second sub-process: Consult and confirm needs

- **Activities of the Process:**

1.2.1 Reach out to customers and hold workshops and meetings necessary to gain a deep understanding of their needs.
1.2.2 Consult with customers to make sure that the actual need is understood and to know their proposals on meeting such need.
1.2.3 Clarify concepts that need to be measured from the customers' point of view, and ensure that they are consistent with existing statistical standards.

1.3 Third sub-process: Establish output objectives

- **Activities of the Process:**

1.3.1 Document detailed statistical needs identified during the «Consult» phase.
1.3.2 Confirm statistical needs in writing by customers.
1.3.3 Agree with the customer on the preparation of the statistical need document, which includes the objectives of the statistical output needed to meet user needs identified during consultation with customers (Process 1.2), as well as details of the required products and services; and the concepts whose suitability should be measured, including time and quality, and expected outputs, including cross tabulation, diagrams, quality standards, priorities and expected challenges. Legal frameworks (e.g. relating to confidentiality), and available resources are likely to be constraints when establishing output objectives.
1.3.4 Agree the suitability of the proposed outputs and their quality measures with users.

1.4 Fourth sub-process: Identify concepts

- **Activities of the Process:**

1.4.1 Identify and define concepts that will be measured (these concepts are the same as those identified by the customer).
1.4.2 Align the required concepts with existing statistical standards whenever possible.
1.4.3 Choose concepts ideally from standard metadata.

1.5 Fifth sub-process: Check data availability

- **Activities of the Process:**

1.5.1 Research into availability of data and methodologies to help develop solutions that meet statistical needs.
1.5.2 Identify data sources that may fit in some way with statistical purposes (e.g. statistics, records, frameworks, processing and analysis models, data quality, timeliness, security and continuity of data supply; etc.).

1.5.3 Observe any restrictions or conditions regarding available data sources.

1.5.4 Coordinate with potential administrative data providers to discuss data availability and quality for the concepts being measured. Assess the division of responsibilities between data providers and GASTAT.

1.5.5 Check necessary ICT resources (e.g. data storage, technology required to handle incoming data and data processing) as well as any formal agreements with data providers for accessing and sharing the data (e.g. formats, delivery, accompanying metadata and quality check).

1.5.6 Collect information about legal frameworks that affect the extent to which that data is used.

1.5.7 Consider the common practices of international statistical agencies, and other agencies that produce similar data.

1.6 Sixth sub-process: Prepare and submit business case

The outputs of this sub-process are the future design and development decisions of the statistical product and their presentation to the customer approval to proceed to the detailed design phase.

- **Activities of the Process:**

1.6.1 Conduct business cases before reasonable solution options can be proposed (sometimes there may not be suitable solution options that can adequately meet identified needs).

1.6.2 Propose initial solution options, including:

- Propose one or more of the initial solution options to the concerned GASTAT committees, to meet previously confirmed statistical needs.
- Present a set of documented and estimated solution options, and review them after evaluating their feasibility and applicability.
- Suggest ways to fill any remaining gaps (compile new outputs from existing data sources, conduct analyses, develop new frameworks or classifications and conduct a new survey). This may include identifying possible partnerships with data holders. This sub-process also includes a more general assessment of the legal framework in which data would be collected and used, and may therefore identify proposals for changes to existing legislation or the introduction of a new legal framework.

1.6.3 Prepare accurate cost calculations of possible solution options to assess their feasibility for identifying and evaluating possible solution options.

1.6.4 Support the business case by stating all sufficient details to provide the relevant reviewers and stakeholders with:

a. A clear understanding of what will be produced.
b. Justification and method of production (detailed).
c. Cost, risks and timing.

- Note: The level of detail in the business case needs to be commensurate with the expected cost, complexity, risk, business importance, as well as the quality of the required outputs.
- Note :Communicate with the customer to discuss the report and confirm whether GASTAT should pursue further development of any of the proposed solution options.
- Note: At this sub-process, a "go"/"no go" decision is made. Typically, the business case is reviewed and formally approved or disapproved by the appropriate sponsors and governance committees.

02 | Second Phase: Design

This phase is triggered in two cases:

1. Design a new product depending on the «Specify Needs» phase, understanding and confirming statistical needs, identifying possible solutions.
2. Redesign the product for improvement and development based on the outputs of the Eighth Phase (Evaluate).

This phase aims to:

1. Develop detailed practical solutions for statistical product design (including all identification work required prior to automation and data collection).
2. Apply a range of common reusable methodologies and instruments using international and national standards to reduce the length and cost of the design process and enhance the comparability and usability of outputs.

This phase consists of (6) sub-processes:

- First sub-process : Design outputs.
- Second sub-process : Design variable descriptions
- Third sub-process : Design collection
- Fourth sub-process : Design frame and sample
- Fifth sub-process : Design processing and analysis
- Sixth sub-process : Design production systems and workflow

Sub-processes and Activities in the Second Phase: Design

2.1 First sub-process: Design outputs

This sub-process contains the detailed design of the statistical outputs, products and services to be produced, including the related development work and preparation of the systems and tools used in the «Disseminate» phase.

- **Activities of the Process:**

2.1.1 Processes governing access to any confidential outputs are also designed here.

2.1.2 Outputs should be designed to follow existing standards wherever possible, so inputs to this process may include metadata from similar or previous collections (including extractions from statistical, administrative, geospatial and other non-statistical registers and databases), international standards, and information about practices in other statistical organizations from the (Identify Needs) sub-process.

2.1.3 Outputs may also be designed in partnership with other interested bodies, particularly if they are considered to be joint outputs, or they will be disseminated by another organization.

2.1.4 Identify mechanisms for providing access to detailed data through dissemination channels.

2.1.5 Identify the timing of the release.

2.1.6 Identify information security procedures.

2.1.7 Identify the quality metrics that are supposed to be disseminated.

2.1.8 Identify the statistical product promotion plan.

2.2 Second sub-process: Design variable descriptions

This sub-process includes:

- Define the variables collected through data collection instruments, as well as any other variables that will be derived from them in the Fifth Phase «Process».
- Define the units, classifications, standards and business frameworks that are used (existing national and international standards).
- Preparation of metadata for these concepts is necessary subsequent design, model, tabulation activities, and implementation.

- **Activities of the Process:**

2.2.1 Identify variables, including:
- Identify and define variables to be used to measure specific phenomena (variables to be collected by data collection instruments, as well as any other variables derived from them, ideally for processing, statistical analysis or dissemination).
- Select such variables directly from an existing set of metadata, or be derived from the metadata according to specific rules.

2.2.2 Define classifications, standards and statistical frameworks, including:
- Define of classifications, standards and statistical frameworks through which statistical data is classified
- Select classifications, criteria and statistical frameworks directly from a set of existing metadata that is available and homogeneous, or between different classifications and statistical frameworks.

2.3 Third sub-process: Design collection

This sub-process aims to identify and document data sources, methodologies, tools and supporting materials for the efficient and effective collection of data required from providers. In addition, data sources, methodologies, tools and supporting materials are tested, redeveloped and revised as necessary before adoption.
The activities in this sub-process will vary according to the type of collection instrument required, which can include computer assisted interviewing, paper questionnaires, administrative registers (e.g. by using existing service interfaces), data transfer methods, web-scraping technologies as well as technology for geospatial data. Direct or indirect use of administrative data may be introduced in the data collection mode for either controlling survey data or assisting it when capturing survey information.

- **Activities of the Process:**

2.3.1 Define the method of data collection, including:
- Identify data sources and methods of collection, including (source of data to be used for the collection of outputs and data channels)
- Determine the data collection method (online inquiries, file transfers, CAPI computer-assisted personal interviews, CATI computer-assisted telephone interviews or CAWI computer-assisted web interviews, etc.).
- Develop a plan to test and validate data collection method (content testing, sequence, usability, security, etc.)
- Develop release and dissemination plans, intensive follow-up and reconnection with data provider, training and study requirements.

2.3.2 Identify data collection tools and supporting materials, including:
- Identify the supporting tools and materials used to collect data such as (forms and questionnaires).
- Identify metadata on the content of questionnaires (questions, sequence, interviewer's instructions, modifications, online help, etc.), and reuse existing question units.
- Identify metadata for administrative data collection tools as a set of requirements using standard models (supporting materials, training in the collection of specific data on the subject, instructions for interviewers and enumerators, integrated tools if possible, dissemination materials associated with data collection).

2.3.3 Find data collection tools and supporting materials:
- Find data collection tools and relevant supporting materials for all types of data.
- Write questions of questionnaires and forms whenever the need arises.
- Prepare the necessary programs for data collection tools.

2.3.4 Approve the data collection method and tools and other metadata materials resulting from the processes, and supporting materials by the concerned bodies and GASTAT committees to be implemented in the Fourth Phase «Collect».

2.4 Fourth sub-process: Design frame and sample

This sub-process is concerned with identifying and improving the plan of the statistical framework and the survey sample and testing it for validation, and finding mechanisms to combine data sources (administrative and statistical records) for use within the statistical framework.

- **Activities of the Process:**

2.4.1 Identify target statistical units for the statistical product. This includes:

- Select the statistical units from a set of available data, or according to the specific data if the target statistical communities are not identified upon determining the scope of solutions.

2.4.2 Identify sampling units, including:

- Identify and define statistical community units from which data is collected.
- Identify and define the required units to ideally analyze, process and disseminate data (analysis units).
- Select the units required to analyze the data directly from an existing set of metadata, or be derived from the metadata according to specific rules.

2.4.3 Design the statistical frames plan, including:

- Design and document the optimal plan for creating statistical frameworks, to cover the target statistical community.
- Review lists, maps and analytical specifications of units from which data providers can be selected
- Use shared statistical records and frameworks as much as possible, and describe how to combine shared sources for the sample framework, if any, such as: Administrative and statistical records, censuses and information from other survey samples, which may include geospatial data and classifications.
- Identify the metadata required to establish the statistical and testing framework, validate and approve its use for the current survey cycle.
- Establish the production framework later in the Fourth Phase «Collect».

2.4.4 Design the sampling plan, including:

- Design and document the optimal plan to select a sample of the units from which the data are collected.
- Select sample from among the statistical frameworks designed to cover the target statistical community.
- Identify the most appropriate sampling methodologies, with a view to providing output with the required quality, with a minimum burden on data providers. The burden on the data provider is reduced by the use of rotation and overlap control methods.
- Define the metadata required for the application of the statistical framework and the allocation and selection of the sample.
- Test the sample, validate it and agree to use it in the current iteration of the project.
- Establish the production framework later in the Fourth Phase «Collect».

2.5 Fifth sub-process: Design processing and analysis

This sub-process designs of processing and statistical analysis plans which will be carried out during the Fifth Phase «Process» and Sixth Phase «Analyze». These plans include:
- Specifications and mechanisms of processing and analysis.
- Models of extracting the required outputs.
- Rules and standards for data linkage, coding and imputation for lost data.

Aspects of these plans may need to be revised during implementation, particularly when they are not performing as expected once the actual data for the statistical product are available.

- **Activities of the Process:**

2.5.1 Design data linkage plan, including:
- Design and document a multi-dataset linkage plan, which is used to support the scope of output aggregation.
- Validate inputs and outputs, including data linkage activities.
- Design data integration specifications from multiple sources.
- Identify partial or total data sets, required from shared data records (collected through this project, or made available from other projects).
- Match or link data sets using precise or probabilistic methods.
- Design methods to control statistical disclosure, and prioritize available data for the same variables to remove anything indicative of identity.

2.5.2 Design data auditing and estimation processes, Including:
- Design optimal models to audit and estimate data through:
- Estimate target statistical communities and variables for which there are no directly available data sources for the respective time period.
- Use statistical analysis methods and available test data (such data from records, and similar data from other sources).
- Estimates extracted through sample weighting processes.
- Seasonally adjusted expectations and estimates.

2.5.3 Design data validation plan, including:
- Design and document the optimal plan for data coding and validation used to infer the outputs (e.g. data inputs) as well as the outputs themselves.
- Validate data against expectations to ensure suitability for use. Validation activities include outliers (including missing data).
- Analyze, process, interpret and explain data, maintain confidentiality and revise outputs.
- Select and test a group of methods and processes, validate and finalize them prior to implementation in the Fifth Phase «Process» and Sixth Phase «Analyze», to ensure that the data validation process is both efficient and effective.
- Understand output priorities, operational constraints, and expectations of what inputs and outputs should be.

2.6 Sixth sub-process: Design production systems and workflow

This sub-process designs the procedures needed for the statistical product by setting an overview of all the required activities and the stakeholders involved in whole production process and ensuring that they fit together efficiently with no gaps or redundancies. This sub-process also defines the various systems and databases required during the process, identifies needs and assesses the appropriateness of the solutions of the existing systems, as well as how and when staff interact with the systems, and who will be responsible for what and when.

- **Activities of the Process:**

2.6.1 Design process procedures and needs for statistical business:
- Design processes and statistical business procedures to meet project needs, first at a high level and then move towards a more detailed level.
- Existing production solutions (e.g. services, systems and databases) should be examined first, to determine whether they are fit for purpose for this specific production process, then, if any gaps are identified, new solutions should be designed.
- Use shared business procedures and services when designing statistical business procedures as much as possible, from the Third Phase «Build» to the Eighth Phase «Evaluate».

2.6.2 Design a plan to test, document and release statistical procedures. This includes:
- Test the operation of combined business tracks through default data (for example, from the previous project cycle) in the Third Phase «Build».
- Test and diagnose processing outputs to see whether they have been done as expected or not.

2.6.3 Assess the suitability of existing system solutions, including:
- Assess shared system solutions to protect the project for the purpose for which it was developed.
- Identify the requirements for new solutions when needed.
- Address new requirements through GASTAT governance committees' meetings; and ideally promoted as shared services.

03 | Third Phase: Build

«Build» as a phase in statistical business production: Build, organize, gather and test production solutions and systems so that they are ready for use in statistical implementation processes. This includes the development and improvement of data collection, processing, analyzing and dissemination systems by taking advantage of previous cycles of the statistical process. This phase is reviewed every time a statistical product is produced through:

- Compile statistical business procedures needed for the project in terms of shared processes and services.
- Design, create, test, manage and organize shared services as institutional shared resources across all departments entrusted with statistical business. Organizers must use simulation or test data to ensure that their processes and flows are in the right direction.

This phase consists of (7) sub-processes:

- First sub-process : Reuse or build collection instruments
- Second sub-process : Reuse or build processing and analysis components
- Third sub-process : Reuse or build dissemination components
- Fourth sub-process : Configure workflows
- Fifth sub-process : Test production systems
- Sixth sub-process : Test statistical business process
- Seventh sub-process : Finalize production systems

Sub-processes and Activities in the Third Phase: Build

3.1 First sub-process: Reuse or build collection instruments

This sub-process describes the activities to build and reuse the collection instruments to be used during the «Collect» phase.

- **Activities of the Process:**

3.1.1 The collection instruments are built based on the design specifications created during the third sub-process of the «Design» phase.

3.1.2 Prepare and test the contents and functioning of the collection instrument (e.g. cognitive testing of the questions in a questionnaire).

3.1.3 It is recommended to consider the direct connection of collection instruments to a metadata system, so that metadata can be more easily captured in the collection phase. This procedure can save work in later phases. Capturing the metrics of data collection (paradata) is also an important consideration in this sub-process for calculating and analyzing process quality indicators.

3.2 Second sub-process: Reuse or build processing and analysis components

This sub-process describes the activities to reuse existing components or build new components needed for the «Process» and «Analyze» phases.

- **Activities of the Process:**

3.2.1 Reuse existing components or build new components needed for the «Process» and «Analyze» phases, as designed in the fifth sub-process of the «Design» phase. They may include: Dashboard functions and features, transformation functions, geospatial data services, provider and metadata management services.

3.3 Third sub-process: Reuse or build dissemination components

This sub-process describes the activities to build new components or reuse existing components needed to disseminate statistical outputs.

- **Activities of the Process:**

3.3.1 Reuse existing components needed to disseminate statistical outputs or build new components as designed in the «Design» phase, to provide dissemination via web services, open data outputs, geospatial statistics, maps, or individual data access.

3.4 Fourth sub-process: Configure workflows

This sub-process simulates the preparation of statistical workflow procedures to be tested in the next sub-process.

- **Activities of the Process:**

3.4.1 Configure and test workflow mechanism and functions, as well as systems used within the business processes, from data collection through to dissemination based on the design created in sixth sub-process of the design phase (Design production systems and workflows).

3.4.2 Modify workflow for a specific purpose, assembling the workflows for the different phases together and configuring systems accordingly.

3.5 Fifth sub-process: Test production systems

In this sub-process, statistical workflow procedures are tested to launch the project in the production environment.

- **Activities of the Process:**

3.5.1 Test the assembled services in the design phase and newly configured services as a result of gaps in the current service list. This includes technical testing and sign-off of new programs and routines.

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

3.5.3 Test of interactions between assembled and configured services, and ensure that the whole production solution works in a coherent way.

3.6 Sixth sub-process: Test statistical business process

- **Activities of the Process:**

3.6.1 Manage a field test or pilot of the statistical business process. Typically, it includes a small-scale data collection, to test the collection instruments, followed by processing and analysis of the collected data, to ensure the statistical business process performs as expected. It may be necessary to go back to a previous sub-process and make adjustments to collection instruments, systems or components of the major statistical business process.

If testing of some of the data collected is required on a very large scale (e.g. a comprehensive census pilot) this test may be repeated until the process is completed satisfactorily.

3.7 Seventh sub-process: Finalize production systems

- **Activities of the Process:**

3.7.1 Put the assembled and configured processes and services, including modified and newly-created services, into production ready for use, including:

- Producing documentation about the process components, including technical documentation and user manuals.
- Training the employees on how to operate the process.
- Moving the process components into the production environment and ensuring they work as expected in that environment.

- Generally, in this «Build» phase, testing, metadata revision, or other features may be repeated to configure statistical workflow procedures before they are ready to be approved for launch.

04 | Fourth Phase: Collect

This phase collects all necessary information (e.g. data, metadata and paradata), using different collection modes (e.g. extraction from sampling surveys, administrative sources, databases) and loads them into databases to be ready for extraction by internal units in the Fifth Phase «Process» (under GASTAT's security policies).

This phase does not include any transformations of the assembled data (unless data loading is needed), as these are all done in the Fifth Phase «Process».

The «Collect» Phase is broken down into three sub-processes covering a number of activities that can occur in parallel. In particular, collection, loading and integration verification activities can be iterative.

This Phase consists of (4) sub-processes:

- First sub-process: Create frame and select sample
- Second sub-process: Set up collection
- Third sub-process: Run collection
- Fourth sub-process: Finalize collection

Sub-processes and Activities in the Fourth Phase: Collect

4.1 First sub-process: Create frame and select sample

This sub-process is the beginning of the operational phase of the statistical product, and is concerned with the creation of the statistical frame and sample selection.
Where the process is repeated regularly, some or all of these activities may not be explicitly required for each iteration.

- **Activities of the Process:**

4.1.1 Create frame:
- Establish the collection process frame as defined in procedure «2.4.3 design plan of statistical frames».
- Validate and approve the use of the statistical frame once it is established.

4.1.2 Select sample:
- Select the collection process sample as defined in Activity «2.4.4 Design sampling plan».
- Validate and approve use of the selected sample. This procedure is not usually relevant for processes based entirely on the use of pre-existing sources (e.g. administrative registers) as such processes generally create frames from the available data. However, variables from administrative and other non-statistical sources can be used as auxiliary variables in the construction of sampling design.

- Verify the readiness of the selected sample for subsequent data collection activities.
- Review and update sample data of data providers (e.g. contact or housing data).
- Prepare and process sampling reports (e.g. response states, percentage, etc.).

4.2 Second sub-process: Set up collection

This sub-process ensures that the people, processes and technology (e.g. web-based applications, GPS system) are ready to collect data and metadata, in all modes as designed during the Design Phase, including the strategy, planning and training activities.

- **Activities of the Process: (For survey data):**

4.2.1 Prepare a collection strategy.

4.2.2 Identify workforce needs to meet requirements during the various phases of the project.

4.2.3 Conduct nomination, training and employment processes for the workforce to ensure a sufficient number of well-trained personnel is available at the appropriate time to carry out the required work.

4.2.4 Training collection staff.

4.2.5 Training system using supervised machine learning techniques.

4.2.6 Ensuring collection resources are available (e.g. laptops, collection apps, APIs).

4.2.7 Agreeing on terms with any intermediate collection bodies, (e.g. for computer assisted telephone interviewing, web services).

4.2.8 Configuring collection systems to request and receive the data.

4.2.9 Ensuring the security of data to be collected.

4.2.10 Preparing collection instruments (e.g. printing questionnaires, pre-filling them with existing data, loading questionnaires and data onto interviewers› computers, APIs, web scraping tools).

4.2.11 Providing information for respondents (e.g. drafting letters or brochures explaining the purpose of the survey, notifying respondents when online reporting instruments will be made available).

4.2.12 Translating of materials (e.g. into the different languages spoken or used in the country).

- **Activities of the Process: (For non-survey data):**

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

4.2.13 Examining and evaluating requests for data.

4.2.14 Initiating contacts with organizations providing the data, and sending an introductory package with details on the process of acquiring the data.

4.2.15 Checking detailed information about files and metadata with the data provider and receiving a test file to assess if data are fit for use.

4.2.16 Arranging secure channels for data transmission.

4.3 Third Sub-process: Run Collection

This sub-process is concerned with collecting and loading data repeatedly. This sub-process includes:

- The initial request for data from its providers and subsequent reception up to loading of relevant data and metadata to databases.
- Elements of managing the relationship between GASTAT and data providers to ensure that it remains positive.

This sub-process aims to get data from its multiple and secure channels using the relevant collection instruments.

- **Activities of the Process:**

4.3.1 Data request, including:

- Sending notifications to data providers for the purpose of requesting data in accordance with the agreements contained in the sub-process 2.3 (Design data collection). Requests for such data may be made electronically (for example, via secure e-mail).
- Send reminders to follow up in case the expected data is not received.

4.3.2 Respond to queries received from data providers and collectors:

- Respond to queries received from data providers (e.g. Refusal to provide required data, difficulties, and changes in communication or business structure details).
- Recording the nature of queries and GASTAT›s response in each case. (e.g.: Sample and statistical frame retention procedures.

4.3.3 Receiving data, including:

- Receive the required data using the methods agreed upon in the data collection plan (e.g.: CAPI computer-assisted personal interviews, CATI computer-assisted telephone interviews, CAWI computer-assisted web interviews or other data collection methods)..
- Geo-coding may need to be done at the same time as collection of the data by using inputs from GPS systems; in addition to the other coding from the used statistical classifications.
- Receiving electronic files, knowing that the actual loading of relevant temporary or permanent databases is dealt with in Activity 4.4.4 below.
- Monitoring and controlling data collection and making any changes necessary to improve data quality.

4.3.4 Loading data and metadata to databases:

- Loading all collected data and metadata into the relevant databases (information may be loaded before or after confirmation of integrity in sub-process 4.4 below, and may include reloading new versions of information received after information is pre-accounted for and stored. Data will be obtained as long as it is acceptable).

- Storing data in a temporary location before transferring and loading it to a permanent database (e.g. load questionnaire tools to databases, upload data collected using administrative tools to a temporary database for validation).
- Load information associated with queries and contacts of data providers.

4.3.5 Archiving physical data collection tools:

- Archiving physical data collection tools (e.g.: paper questionnaires) and disposal after being electronically stored.
- Applying automatic archiving of information derived from other types of collection tools in accordance with the generally accepted archival rules at the end of the relevant procedure.

4.4 Fourth sub-process: Finalize collection

This sub-process is concerned with the basic integrity of any data collected externally that is deemed «appropriate for the purpose for which it was collected» and prepares it for use in the activities of the Fifth Phase «Process».

- **Activities of the Process:**

4.4.1 Reviewing integrity of collected data:

- Reviewing integrity of data collected from all data channels, through the ideal use of a set of applicable automated rules, whether before or after the data is loaded to a temporary or permanent shared database.
- Reviewing the integrity of the administrative data by comparing data volume and content, and verify the format of the files received to achieve the expected targets.
- Reviewing the integrity of standard metadata to ensure the resulting loading (e.g. avoiding negative values while expecting positive ones, avoiding data format errors).
- Reviewing the integrity of data sets collected through questionnaires by verifying acceptable response targets (e.g. acceptability of response rates, submission of responses to major variables).
- Adjusting any deviations detected through integrity checks and addressing them in activities 4.4.2, 4.4.3 and 4.4.4 below, depending on the type of follow-up required.

4.4.2 Monitoring data collectors:

- Adjusting and addressing deviations detected through desk checks.
- Obtaining missing data not yet received from the data providers.
- Responding to queries associated with samples.
- Managing the workforce performance by regularly providing feedback on performance, rewarding and praising high-quality performance, dealing effectively with poor performance, etc.
- Preparing regular initial reports on the project workforce.

4.4.3 Monitoring data providers:

- Direct contact with data providers to help resolve and address deviations detected during desk checks and obtaining missing data that has not been yet received.

- **Note**: The quality of the data provided is not verified in this process, as this is done in sub-process 5.3 (Review and validate).

4.4.4 Adjusting and addressing data integrity deviations, including:

- Adjusting and addressing deviations detected during desk reviews which do not warrant follow-up (follow-up is made in activities 4.2.2 and 4.2.3 above). An example is the review of information publicly available on the Internet.
- Reviewing data entered from paper models.
- Encoding variables received from other bodies.
- Converting data to appropriate formats to be loaded to shared databases.

4.4.5 Approving collected data, including:

- Approving data collected externally to move forward to the next processing procedures.
- Preparing and using reports containing relevant standards for data collection to review the approval process.

- **Note**: Data can remain acceptable and loaded to databases after this closure phase as long as it is appropriately identifiable.

05 | Fifth Phase: Process

The collected data is processed at this phase by converting it into a set of statistical outputs for two purposes:

1. Statistical analysis that reveals broader understanding of data.
2. Disseminate to customers using various rules or forms approved for dissemination.

- This phase occurs in each iteration of statistical production.
- The processes and activities 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.
- The «Process» and «Analyze» phases may also commence before the «Collect» phase is completed. This enables the compilation of provisional results and extend the time available for validation.
- This phase can be done in parallel and can be iterative with the Sixth Phase «Analyze» Analysis can reveal a broader understanding of the data, which might make it apparent that additional processing is needed.
- The difference between data validation and editing procedures at both the «Process» and «Analyze» phases:
  - At the Process Phase: Validation of data entered into the models and used in the production of the outputs.
  - At the Analyze Phase: Validation is made for post-processing outputs.

This phase consists of (8) sub-processes:

- First sub-process : Integrate data
- Second sub-process : Classify and code
- Third sub-process : Review and validate
- Fourth sub-process : Edit and impute
- Fifth sub-process : Derive new variables and units
- Sixth sub-process : Calculate weights
- Seventh sub-process : Calculate aggregates
- Eighth sub-process : Finalize data files

Sub-processes and Activities in the Fifth Phase: Process

5.1 First sub-process: Integrate data

This sub-process integrates data from one or more sources. The input data can be from a mixture of external or internal sources, and a variety of the collection instruments, including extracts of

administrative and other non-statistical data sources. Administrative data or other non-statistical sources of data can substitute for all or some of the variables directly collected from survey.

This sub-process also includes harmonizing or creating new figures that agree between sources of data. The result is a set of linked data.

Data sets may need to be linked several times during this «Process» phase in order to increase the quality of the entered data.

- **Activities of the Process:**

5.1.1 Extraction and compilation of data, including:
- Extracting or compiling of data records from relevant databases, allowing greater effectiveness of monitoring certain phenomena. Obtaining integrated statistics such as national accounts.
- Combining spatial demographic data with statistical or other non-statistical data.
- Preparing the linking process or other organizing activities.

5.1.2 Data matching, including:
- Matching or integrating multiple datasets (micro/macro) that are required for organization or other audit activities (e.g. Extracting statistical accounts and analyses and validating inputs or outputs), whether under complete matching or statistical matching:
  - Complete matching: Linking records from two different data sets believed to belong to the same agency or statistical community.
  - Statistical matching: Linking records that do not necessarily belong to the same individual, using information concerning similar individuals. It is used when there is little or no overlap between the datasets being integrated.

5.1.3 Prioritizing data, including:
- Prioritizing data when two or more data sources contain data on the same variable (with potentially different values). When linking multiple datasets, data from the same agency or statistical community is usually available on both data sets.
- Making decisions (in the form of predetermined rules), according to data of higher priority, and on the basis of which data will be kept in the linked dataset (e.g. based on the corresponding specified working time, data quality, etc.).

5.1.4 Anonymizing data, including:
- Clearing all personal identifiers from linked or unlinked datasets to ensure compliance with legal requirements that protect individuals› privacy.
- Stripping individual›s name, address and other identifiers from the collected survey data and from administrative files as soon as possible after the process. (After all linking processes are completed).

- **Note:** Removing identifiers from datasets (usually input datasets) is different from producing confidential outputs in sub-process 6.4 (Apply disclosure control), where different methods are used and the focus is on the produced output.

5.2 Second sub-process: Classify and code

This sub-process is concerned with classifying and coding input data from micro and macro data. For example, coding routines may assign numeric codes to coded responses according to a pre-determined statistical classification system. This sub-process also classifies input data at the macro level into different classifications or coding systems using indexes.

- **Activities of the Process:**

5.2.1 Automatic classification and coding of values, including:
- Assigning all categories.
- Coding variable values using automated methods (in bulk or immediately during the review of individuals› responses).
- Re-assigning or re-encoding if necessary (e.g. coding based on geographic data, economic and professional activities.

5.2.2 Classifying and coding values not automatically classified, including:
- Identifying instances in which codes cannot be automatically assigned to variables.
- Assigning or re-assigning these instances clerically (e.g. if available information increases).
- Use of clerical coding is kept to a minimum (e.g. Assigning industrial and professional codes where descriptions cannot be automatically assigned using search tables).

5.3 Third sub-process: Review and validate

This sub-process examines data to identify potential deviations, problems, errors and discrepancies found in the inputs of the «Process» phase, and adjusts and addresses them to be entered in the forms designed to calculate weights and aggregates in Sub-process 5.5. Deviations can be defined as certain errors in data, or data of a questionable nature which may arise from wrong classifications in coding, data loss, and copying errors, etc.

This sub-process aims to:

1. Ensuring that these inputs are «valid for the purpose for which they were collected» by improving them up to acceptable quality priorities as effectively as possible.
2. Identifying and modifying automated data, including imputation of missing or unreliable data.

It may be run iteratively, validating data against pre-defined edit rules, usually in a set order. It may flag data for automatic or manual inspection or editing. Reviewing and validating can apply to data from any type of source, before and after integration, as well as imputed data from the next sub-process (Edit and impute). Whilst validation is treated as part of the «Process» phase, in practice, some elements of validation may occur alongside collection activities, particularly for modes such as computer assisted collection. Whilst this sub-process is concerned with detection and localization of actual or potential errors, any correction activities that actually change the data is done in the sub-process (Edit and impute).

The validation process may be applied at different times during the project (e.g. before or after the data linking, or after changing the entered data upon validating the outputs in sub-process «6.2 validate outputs»).

- **Activities of the Process:**

5.3.1 Detecting and addressing missing inputs and outliers that can be automatically corrected, including:
- Detecting and addressing deviations in entry data that are capable of automatic correction, using or mathematical rules or solutions.
- Applying methods of imputing missing or unreliable data, as well as a set of other rules designed to identify deviations automatically.
- Modifying or changing data (prior to editing activities or at other times during the review process).
- Applying the treatment methods required for the current cycle data and comparing them with the previous cycles.

5.3.2 Detecting outliers that must be manually modified, including:
- Automating outlier detection in data inputs using specific rules (these rules may be in the form of traditional amendments or selected editorial procedures).
- Detecting deviations that greatly affect outputs using the same rules above.

- **Note:** Detection is limited to these types of outliers that are significant or large enough to warrant a clerical investigation. Significant deviations may include distributed workload that requires adjustment.

5.3.3 Analyzing and adjusting outliers, including:
- Using investigative data analysis methods and utilizing statistical expertise to detect and address outliers.
- Identifying the causes of deviations detected in input data. This approach helps to detect and understand extreme deviations. Large deviations here pertain to data distributions (such as outliers) that are not suitable for known predetermined rules).
- Reviewing data in table, graph, or map format.

5.4 Fourth sub-process: Edit and impute

- **Activities of the Process:**

5.4.1 Identifying appropriate methods to address outliers.
5.4.2 Applying these methods in instances where quick addressing is not possible and if the detected deviations involve changing or modifying some data, the traditional form of addressing is to leave the data unchanged.
5.4.3 Documenting the causes of deviation according to the phases of Design, Collect or Run in a previous activity.
5.4.4 Addressing outliers, and flagging them as changed.

5.5 Fifth sub-process: Derive new variables and units

This sub-process derives data for variables and units that are not explicitly provided in the collection, but are needed to deliver the required outputs. It derives new variables by applying arithmetic formulae to one or more of the variables that are already present in the dataset, or applying different model assumptions.
This activity may need to be iterative during the Fifth Phase «Process» prior to certain activities such as data linking, coding, validation and obtaining outputs, as some derived variables may themselves be based on other derived variables. It is therefore important to ensure that variables are derived in the correct order. New units may be derived by aggregating or splitting data for collection units, or by various other estimation methods.

- **Activities of the Process:**

5.5.1 Applying a set of estimation models designed in sub-process 2.6 (Design production system and workflows) to current data to create the required outputs.

5.5.2 Extracting new classes and units, including:
- Deriving new statistical communities from aggregated data (e.g. individuals under the age of 15 with disabilities).
- Including survey and target communities as well as statistical frame and analysis into the statistical communities.
- Deriving the values of new statistical units by aggregating or splitting data for collection or reporting units, or by various other estimation methods (e.g.: Deriving households where the collection units are persons using household derivation processes).

5.5.3 Deriving new variables, including:
- Deriving values for new variables necessary to aggregate the required outputs, or to validate them by applying arithmetic formulae to one or more of the variables that are already present in the dataset.
- Applying derivations to datasets before or after data linking. This activity may need to be iterative, as some derived variables may themselves be based on other derived variables.
- Ensuring that variables are derived in correct order (e.g. deriving total values, change ratios, price relations, two combined estimates, change hours to days and apply exchange rates).

5.6 Sixth sub-process: Calculate weights

This sub-process creates weights for unit data records according to the methodology developed in sub-process 2.5 (Design processing and analysis). It also converts certain statistical inputs into different outputs using specific and tested modeling methods when designing the methodology under activity «2.6.3 design data validation plan».

- **Activities of the Process:**

5.6.1 Creating weights for unit data records according to the methodology developed in sub-process 2.5 (Design processing and analysis). For example, weights can be used to «gross-up» data to make them representative of the target population in the survey, this is based on projections and the results of population estimates for the survey period.

5.6.2 Adjusting for non-response in total enumerations.

5.6.3 Weighting variables for adjustment purposes. It may also include correcting weights for benchmarking indicators (such known population totals).

5.7 Seventh sub-process: Calculate aggregates

Examples of outputs from this sub-process include the creation of aggregated data as population totals from individual data or lower-level aggregates, and the calculation of corresponding sampling errors for related aggregates, administrative data aggregates, accounts, serial scales, indicators, time series, forecasts or predictions, error measurements, averages, dispersion and quality standards.

- **Activities of the Process:**

5.7.1 Produce gross-up data at different levels of details (e.g. data at the micro, intermediate or macro level).
5.7.2 Applying weights specified in activity 5.6.1" Calculate weights".
5.7.3 Deriving appropriate totals.

5.8 Eighth sub-process: Finalize data files

- **Activities of the Process:**

5.8.1 Loading statistical outputs (data and metadata) to relevant output databases and should be stored in formats appropriate to the scope of the output analysis.
5.8.2 Sometimes this may not be final outputs, particularly where there are strong time pressures, and requirements to produce both preliminary and final estimates.
5.8.3 Validating statistical outputs after loading.

06 | Sixth phase: Analyze

This phase is concerned with the detailed examination and validation of statistical outputs against expectations and comments. Relevant documents are prepared and content is developed for dissemination. This sub-process includes activities that enable statistical analysts to understand and interpret the resulting outputs.
This phase is done in each cycle for statistical outputs produced regularly، Its outputs include: Statistics, relevant metadata and comments for both internal and external release.
The outputs of this phase can also be used as input for other sub-processes (e.g., using new source analysis as input for the «Design» phase).
The «Analyze» phase includes all statistical outputs, regardless of the source of the data.

This Phase consists of (5) sub-processes:

- First sub-process : Prepare draft outputs
- Second sub-process : Validate outputs
- Third sub-process : Interpret and explain outputs
- Fourth sub-process : Apply disclosure control
- Fifth sub-process : Finalize outputs

Sub-processes and Activities in the Sixth Phase: Analyze

6.1 First sub-process: Prepare draft outputs

- **Activities of the Process:**

6.1.1 Collecting factual information about the statistical scope (statistical filed) by making the most of the common database available at GASTAT.
6.1.2 Prepare draft outputs from the «Calculate weights and totals» and «load outputs to databases» processes in the «seventh and eighth sub-process» of the «Process Phase».

6.2 Second sub-process: Validate outputs

This sub-process validates the quality of the produced outputs in accordance with a general quality framework and customers' expectations. The team entrusted with the procedures of this sub-process must have an in-depth understanding, review, analysis and interpretation of the statistical outputs from many perspectives using different tools and methods for validation. These aspects include comparing statistics with previous cycles (as far as possible), comparing statistics with other relevant data (internal and external), examining statistics or analyzing them into the component parts, deriving patterns or key messages from the data and reviewing quality standards related to the data.

- **Activities of the Process:**

6.2.1 Detecting and addressing outlier outputs that can be automatically corrected, including:
- Predetermining the correction rules for detecting outlier outputs.
- Detecting deviations in produced data that are capable of automatic correction by applying correction rules.
- Checking that the coverage and response rates are as required.
- Comparing the statistics with previous cycles to validate data.
- Checking that the associated metadata, paradata and quality indicators are present and in line with expectations.
- Checking geospatial consistency of the data.
- Comparing the statistics against other relevant data (both internal and external).
- Investigating inconsistencies in the statistics.
- Performing macro editing.
- Validating the statistics against expectations and domain intelligence.
- This procedure is usually reiterative every time a new version of the outputs changes.

6.2.2 Analyzing outputs and adjusting deviations, including:

- Assessing the outputs and get an accurate understanding after they›ve been validated for release.
- Adjusting significant outlier outputs detected using mechanical methods in activity «6.2.1» or through statistical visual and clerical analyses.
- Statistical analysis of outputs using statistical analysis methods, presentation methods (tables, graphs, maps) and tools ideally using a group of internal and external data sources.
- Assessing the success of outputs to meet expectations (using the scope information collected in activity «6.1.1») to be compared with other data sources.
- Testing assumptions, understanding key components, quality indicators, etc.

6.2.3 Addressing outlier outputs, including:

- Addressing outputs that are different from expectations to ensure that they are correct and valid for release.
- Applying a group of methods to address any deviations (despite the tendency to use clerical methods when there are very significant variations from expectations). If differences from expectations can be justified and explained, the outputs will remain unchanged.
- Documenting all addressing decisions to be a reference in the future (as a rule and quality).
- Addressing data deviation from the source (e.g. price quality adjustments, derivation processing, etc.).
- This procedure is done in each reiteration of the activities of the current sub-process.

6.3 Third sub-process: Interpret and explain outputs

This sub-process aims to interpret data outputs, where comments and relevant supporting documents explaining these outputs are compiled. Documents and reports related to statistical analyses can be used even if they have been prepared for the internal approval processes of GASTAT.

- **Activities of the Process:**

6.3.1 Confirm outputs for release:

- Selecting outputs valid for release and dissemination with the determination of the level of dissemination according to the rules and policies of dissemination and based on validation and preservation of confidentiality achieved in sub-process 6.4.

- **Note:** Not all statistical product outputs loaded into databases are necessarily considered «highly relevant» for release or dissemination. This activity selects and confirms outputs that are considered valid for release.

6.3.2 Prepare clarifications for release:

- Preparing comments to support the data being released. Such comments may be in the form of text, presentations, table, graph, or map.
- Storing comments in a ready-made template for product assembly in the sub-process «7.2 produce dissemination products».

6.3.3 Prepare credentials for release:

- Collecting information supporting the release of data, including: Interpretation, summary reports, confidentiality assurance, and any other necessary metadata.
- Preparing official documents using GASTAT approved forms.
- Filling in documents as automatically as possible (trough the materials used to validate outputs in sub-process 6.2).
- Including a comprehensive report or explanation that reveals the full picture along with individual reports collected regarding specific outputs.

- **Note:** For internal purposes, information supporting the release of micro data and macro data is also collected, including: interpretation, summary reports, confidentiality assurance, and any other necessary metadata; as well as internal reports for analysis purposes only.

6.4 Fourth sub-process: Apply disclosure control

This sub-process ensures that the data and metadata to be disseminated do not breach the appropriate rules on confidentiality according to GASTAT dissemination policies, or to the process-specific methodology created in sub-process 2.5 (Design processing and analysis).
This may include checks for primary and secondary disclosure and dissemination, as well as the application of data suppression or perturbation techniques and output checking.

- **Activities of the Process:**

6.4.1 Monitoring and disclosure of confidential data:
- Using statistical analysis methods and means of presentation for disclosure of confidential data (similar to those used in activity «6.2.2 analyze outputs and adjust deviations»).

6.4.2 Handling confidential data:
- Applying automated or clerical control methods to suppress or perturb data upon disclosure and dissemination.
- Examining the outputs after applying this process.
  - Methods applied to control, slight modification or deletion of data may limit the amount of detail available, requiring permission from the data provider to disseminate it.
  - The degree and method of statistical disclosure control may vary for different types of outputs. For example, the approach used for microdata sets for research purposes will be different to that for published tables, finalized outputs of geospatial statistics or visualizations on maps.

6.5 Fifth sub-process: Finalize outputs

This sub-process aims to approve data, metadata, and comments that interpret the data and load them to dissemination databases after they reach the required quality level.

- **Activities of the Process:**

6.5.1 Convening internal meetings to discuss approval:
- Meetings between statisticians responsible for the production of outputs and relevant specialized experts at GASTAT will be held to discuss the quality of data outputs and verification of consistency.

6.5.2 Approving content for release, including:
- Translating statistical content prepared for release into English including metadata and comments (text and presentations) and other supporting technical documents.
- Approving statistical content prepared for release (in both Arabic and English) including metadata and comments (text and presentations) and other supporting technical documents.

6.5.3 Checking data outputs with customers:
- Reviewing data outputs on customers› and beneficiaries› requests by taking their views on the data outputs prior to dissemination.

- **Note:** This activity is specific to the surveys carried out by GASTAT at the request of beneficiary parties.

6.5.4 Loading final content to protected dissemination databases, including:
- Verifying the written approval of the authorized person to the statistical product outputs.
- Loading statistical content to relevant protected dissemination databases.
- Making statistical product data available for aggregating products and services in sub-process «7.2 produce dissemination products».

07 | Seventh Phase: Disseminate

This phase assembles, organizes and releases statistical products and statistical content by means of dissemination to customers. Products are assembled and services are provided based on the approved statistical content ready for release which has been loaded to secure dissemination databases under activity «6.5.4 load final content to secure dissemination databases".
For statistical products produced regularly, this phase occurs in each iteration, including: An integrated statistical product disseminated through dissemination channels approved by GASTAT.

This Phase consists of (5) sub-processes:
- First sub-process: Update output systems
- Second sub-process: Produce dissemination products
- Third sub-process: Manage release of dissemination products
- Fourth sub-process: Promote dissemination products
- Fifth sub-process: Manage user support

Sub-processes and Activities in the Seventh Phase: Disseminate

7.1 First sub-process: Update output systems

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

- **Activities of the Process:**

7.1.1 Preparing and loading of outputs to public dissemination databases, including:
- Formatting and preparing data and metadata ready to be put into public dissemination databases.
- Loading data and metadata into public dissemination databases.
- Ensuring data are linked to the relevant metadata.
  - Formatting, loading and linking of metadata should preferably mostly take place in earlier phases, but this sub-process includes a final check that all of the necessary metadata are in place ready for dissemination.

7.2 Second sub-process: Produce dissemination products

This sub-process produces the dissemination products, as previously designed in sub-process 2.1 (Design outputs), to meet customers' needs. Promotional martials are also prepared, and may include printed publications, press releases and websites. The products can take many forms including interactive graphics, tables, maps and public-use microdata sets. Typical sub-processes include:

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

- **Activities of the Process:**

7.2.1 Assembling regular products:
- Assembling products made available to beneficiaries and regularly released (e.g. data tables, time series, publications, special analyses, graphs, data presentations, etc.) of approved data, associated text, presentations, and technical documents derived from secure dissemination databases.

7.2.2 Assembling customized products:
- Assembling products specifically designed to meet the needs of specific customers (e.g. special tables and analyses, confidential files for unit records, etc.).

- **Note:** These products are paid except for international requests.

7.2.3 Preparing promotional materials:
- Preparing and carrying out promotional plans for products with a view to advertising.
- Considering the appropriate methods for each product according to its importance.

- **Note:** Promotional plans include study of target audience and access strategy.

7.2.4 Approving regular products for release:
- Approving statistical products, including promotional materials, for release.
- Reviewing statistical products to ensure that the specific needs of the customer are met, in accordance with GASTAT policies and procedures.
- If the products are not approved for a first time, the following sub-activities will be applied:

- Returning the product to the relevant phase to address the core problem.
- Reassembling products from secure dissemination databases.
- Applying the approval process again.

- **Note:** Content should not be fixed by custom modifications at the dissemination phase itself, but should be modified on databases.

7.3 Third sub-process: Manage release of dissemination products

This sub-process ensures that all elements for the release are in place including managing the timing of the release. It includes briefings for specific groups such as the press or decision makers from relevant officials, facilitating access to statistical content in public dissemination databases through multiple dissemination channels for different beneficiaries on various devices, as well as the arrangements for any pre-release embargoes. It also includes managing access to confidential data by authorized user groups according to GASTAT access policies. The disseminated content is also archived in accordance with the archiving policy designed in over-arching activity «9.3 data management».

- **Activities of the Process:**

7.3.1 Exporting data to public dissemination databases, including:
- Copying data from a protected internal storage server to a platform available to beneficiaries.
- Making data available for transfer, and accessing it through a set of tabulated or visualizing tools for interactive data on the web site.

7.3.2 Briefing relevant stakeholders prior to release, including:
- Briefing relevant stakeholders about statistical content prior to release to lift data embargoes. Embargoed content may be accessed through secure databases (often key results only) or restricted access prior to release to important external authorized customers (for example, ministers, key government departments).
- Facilitating partners' review of forthcoming release, to allow for producing a shared release with another party, and assist in public launching or similar event.

7.3.3 Releasing products and services to customers, including:
- Approving statistical content for public dissemination.
- Releasing statistical content through specified dissemination channels according to the date of dissemination announced in GASTAT Statistical Calendar.
- Briefing relevant customers and providing them with data to be disseminated.
- Observing the statistical product dissemination date as previously posted on GASTAT website.

7.3.4 Providing customers with the post-release of results, including:
- Providing (key) results to specific customer groups by product in different templates.
- Continuing the submission process and responding to clarification requests.
- Marketing the product after its release by establishing relationships with the related customers.

- **Note:** The submission templates include (reports and publications, presentations, participation in statistical product-related events).

7.3.5 Dealing with media interviews and queries, including:
- Responding to media queries on the disseminated statistical product.
- Monitoring interaction and measuring feedback in the media and bringing them up to the relevant department for future improvements.

7.3.6 Maintaining disseminated content, including:
- Applying the archiving plan from over-arching activity «9.3 data management» for externally sourced statistical content (data, metadata, comments),
- and ensuring that the content has been successfully archived.

7.4 Fourth sub-process: Promote dissemination products

This sub-process concerns the active promotion of the statistical products produced in a specific statistical business process.

- **Activities of the Process:**

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

7.4.2 Using tools including websites, wikis and blogs to facilitate the process of communicating statistical information to users.

7.5 Fifth sub-process: Manage user support

This sub-process ensures collection and review of customers' queries, service requests and requirements to be answered and met.

- **Activities of the Process:**

7.5.1 Receiving customers' queries and questions, including:
- Receiving queries and questions from customers and beneficiaries of data and information related to the products and services of GASTAT.
- Receiving requests for availability and access to accurate data.
- These queries and requests should be regularly reviewed to provide an input to the overarching quality management process, as they can indicate new or changing user needs.
- Monitoring interaction and measure customers' feedback and submit them to the relevant department to be used for improvements.

7.5.2 Responding to customers' queries, including:
- Following up on the release day to ensure that the dissemination date is adhered to.
- Responding to customers' queries shortly after the release (e.g., questions regarding released statistics, availability of additional data, etc.).
- Recording queries to support the «Evaluate» phase.
- Handling subsequent queries, including requests for customized outputs, products or services, through Customer Relationship & Support Department.

08 | Eighth Phase: Evaluate

This phase manages the evaluation of a specific instance of a statistical business process, as opposed to the more general overarching process of statistical quality management described in (Overarching Processes). It relies on inputs gathered throughout the production phases to be evaluated based on specific procedures leading to the analysis and identification of improvements required for product quality assurance.

It can take place at the end of the instance of the process, but can also be done on an ongoing basis during the statistical production process.

For statistical products produced regularly, this phase occurs in each iteration, including: Evaluation and development of improvements plan.

For statistical outputs produced regularly, evaluation should, at least in theory be considered as the decision point for post-evaluation procedures:

- Determining whether future iterations should take place.
- If so, whether any improvements should be implemented.
- Providing the decision as to whether the next iteration should start from the «Specify Needs»phase, or from some later phase (often the «Collect» phase).

This Phase consists of (3) sub-processes:

- First sub-process : Gather evaluation inputs
- Second sub-process : Conduct evaluation
- Third sub-process : Agree an action plan

Sub-processes and Activities in the Eighth Phase: Evaluate

8.1 First sub-process: Gather evaluation inputs

This sub-process aims to compile all forms of evaluation elements, including: (feedback from users, process metadata, system metrics, staff suggestions, and reports of progress against an action plan agreed during a previous iteration) and made them available to the team that will carry out the evaluation. For the evaluation of certain processes it can be necessary to perform specific activities such as small surveys, (e.g. post-enumeration surveys, re-interview studies, survey on effectiveness of dissemination).

- **Activities of the Process:**

8.1.1 Design evaluation plan

- Design and document the project success evaluation plan, including achieved outputs and used resources (e.g. data sources, registries, etc.), as well as the capabilities used to provide the sources of production activities (e.g. methods, processing, systems, standards, skills etc.).
- Identify the entities interested in the product for consultation and determination of the timing and methods for collecting inputs and criteria to be used for evaluation, as well as the effectiveness of the evaluation of quality portals.
- Identify, test and validate monitoring, performance and output criteria in this procedure.

8.1.2 Gather quantitative evaluation inputs:

- Gathering a set of quantitative information or measurable inputs to enhance the evaluation process.
- Automating the evaluation inputs, which includes the metrics on individual processes, metrics or measures relating to output quality, such as: Effects of imputation of missing values, revisions to relative standard errors, etc.).

8.1.3 Gather qualitative evaluation inputs:

- Collection of qualitative or textual information using a variety of methods (e.g.: User meetings, internal meetings to discuss outstanding issues, meetings, workshops, etc.).
- This information is collected from:
  A. Production staff.
  B. Customers.
  C. Data providers.
  D. Service providers.

- **Note:** The information is compiled (clerically), and some information collected during the project is added to it such as: Comments on output quality during review.

8.2 Second sub-process: Conduct evaluation

In this sub-process, evaluation inputs are analyzed and collected in the prepared report or dashboard. The evaluation report includes (action plan, any quality issues, recommendations for changes where appropriate). These recommendations can cover changes to any phase or sub-process for future iterations of the process, or can suggest that the process is not repeated. The action plan is then presented to the relevant body to set priorities and move forward.
The evaluation can take place at the end of the whole process (ex-post evaluation) for selected activities, during its execution in a continuous way, or throughout the process, thus allowing for quick fixes or continuous improvement. The resulting report should note any quality issues specific to this iteration of the statistical business process.

- **Activities of the Process:**

8.2.1 Analyze evaluation inputs, including:
- Accurate analysis of evaluation inputs collected in sub-process «8.1 gather evaluation inputs».
- Comparing the results of the analysis with the expected results of the project, to determine whether or not they have been achieved.

8.2.2 Identify improvements, including:
- Identifying a set of improvements or potential solutions to the problems identified by the analyses.
- Discussing identified improvements with those concerned to bring them before the authorized person.

- **Note:** Improvements can be made to processes, methodologies, systems, personnel skills, standards and statistical frameworks.

8.2.3 Provide data quality instructions to data providers, including:
- Providing data providers with information and resources with a view to improving data quality.
- Shedding light on advantages obtained by data providers, as well as GASTAT and its customers.

- **Note:** This can also deal with internal source units and external administrative data providers.

8.2.4 Measure customer satisfaction, including:
- Developing specific mechanisms to measure customer satisfaction and identify elements of satisfaction.
- Coordinating with customers to express their views about the product through direct contact or workshops.
- Measuring customer satisfaction with the statistical product.
- Dealing with dissatisfied customers and explaining the situation to them.
- Analyzing the results and submitting them to the relevant department for improvement.
- Measuring the satisfaction of the statistical community.

- **Note:** Customer satisfaction questionnaire can be carried out on several products or a specific product.

8.2.5 Measuring customer use of indicator:
- Measuring customer use of exported statistical indicators and the extent to which these indicators meet customers› objectives and requirements.

8.2.6 Prepare evaluation report, including:
- Documentintg the results of the evaluation analyses.
- Preparing the evaluation report based on the results of documented analyses.
- Making recommendations for improvements.

- **Note:** The report includes: Recommended changes to various strategies and the data set collected during the survey, impact assessments, proposed changes to the issuance of lists and risk mitigation strategies, and an action plan indicating the appropriate timing, priority and responsibility to undertake all identified improvements.

8.3 Agree an action plan:

- **Activities of the Process:**

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

- It should also include consideration of a mechanism for monitoring the impact of those actions and improvements to the statistical process and its results, which may, in turn, provide an input to evaluations of future iterations of the statistical process.

09 | Quality Management, Metadata Management & Data Management

In addition to the overarching processes applied during the eight production phases discussed above, this section will address other processes related to quality management, metadata management and data management.

In general, these processes aim to increase the possibility of completing the work on time and in accordance with the defined budget and quality levels required to mitigate risks that the product may face at any of the previous eight phases of production, and also strengthen the aspects of project management as one major group.

Overarching processes include:

9.1 Quality Management

- The main goal of quality management within the statistical business process is to understand and manage the quality of the statistical sources, processes and products. There is general agreement among statistical organizations that quality should be defined according to the ISO 9000-2015 standard: «The degree to which a set of inherent characteristics of an object fulfils requirements». Thus, quality is a complex and multi-faceted concept, usually defined in terms of several quality dimensions. The dimensions of quality that are considered most important depend on user perspectives, needs and priorities, which vary between processes and across groups of users.

- While quality management is linked to the «evaluate» phase, quality management should be present at all phases of other statistical work processes. It is necessary to evaluate procedures at each separate phase taking into account that evaluation is unlikely to be feasible for every iteration of every part of every statistical business process. Thus, there should be systematic way for evaluation according to a pre-determined schedule that allows for the review of all main parts of the process within a specified time period.

- By conducting these evaluations, a number of proposals (feedback) are formed, which

certainly include the metadata generated by the various sub-processes that are used as an input to process quality management. These evaluations can be applied within a specific process, or across several processes that use common components.

- Quality control procedures also play a key role in quality management, at the level that should be implemented within sub-processes to prevent and monitor errors and sources of risk. These procedures must be documented, so that they can be used for quality reporting.
- Quality management within GASTAT should be based on the specific quality framework adopted by senior management. Such a framework need not be fully compatible with the quality frameworks within other statistical organizations, as this reinforces the importance of cross-organizational benchmarking.
- «Quality loop» reinforces the approach to continuous improvements and organizational learning as the quality processes go through the elements of such loop as elaborated in the below figure:

Figure 3: Quality loop

- **Examples of quality management activities include:**
  - Setting quality criteria at each phase of the project to be used in the process.
  - Setting process quality targets and monitoring compliance.
  - Examining process metadata and quality indicators.
  - identifying and assessing risks, Including: financial risk, employee benefit risk, customer relationship risk, etc.,
  - Implementing risk treatments to ensure fit-for-purpose quality.
  - Monitoring project action plan and budget.
  - Seeking and analyzing user feedback.
  - Reviewing processes and documenting lessons learned.
  - Internal or external auditing on the process.
  - Preparing workflow reports.
  - Adopting corrective actions.

9.2 Metadata Management

- Metadata has an important role and must be managed at an operational level within the statistical production process. When aspects of metadata management are considered at corporate or strategic level (e.g. there are metadata systems that impact large parts of the production system), it should be considered in the framework of the Generic Activity Model for Statistical Organizations (GAMSO).
- Because metadata exists at every phase of the statistical process, the overall management process of metadata focuses on the creation/revision, updating, use and archiving of statistical metadata.
- Ensuring that these metadata are captured as early as possible, and stored and transferred from phase to phase alongside the data they refer to. Therefore, metadata management strategy and systems are vital to the operation of this model, and these can be facilitated by the Generic Statistical Information Model (GSIM).
- GSIM is a reference framework of information objects, which enables generic descriptions of the definition, management and use of data and metadata throughout the statistical production process. GSIM supports a consistent approach to metadata, facilitating the primary role for metadata, that is, that metadata should uniquely and formally define the content and links between information objects and processes in the statistical information system.
- The METIS Common Metadata Framework identifies the following sixteen core principles for metadata management, all of which are intended to be covered in the overarching metadata management process, and taken into the consideration when designing and implementing a statistical metadata system.

- **The principles are presented in four groups:**

Metadata handling	1. Statistical Business Process Model: Manage metadata with a focus on the overall statistical business process model. 2. Active not passive: Make metadata active to the greatest extent possible. Active metadata are metadata that drive other processes and actions. Treating metadata this way will ensure they are accurate and up-to-date. 3. Reuse: Reuse metadata where possible for statistical integration as well as efficiency reasons. 4. Versions: Preserve history (old versions) of metadata
Metadata Authority	1. Registration: Ensure the registration process (workflow) associated with each metadata element is well documented so there is clear identification of ownership, approval status, date of operation, etc. 2. Single source: Ensure that a single, authoritative source («registration authority») for each metadata element exists. 3. One entry/update: Minimize errors by entering once and updating in one place. 4. Standards variations: Ensure that variations from standards are tightly managed/approved, documented and visible
Relationship to Statistical Cycle / Processes	1. Integrity: Make metadata-related work an integral part of business processes across the organization. 2. Matching metadata: Ensure that metadata presented to the end-users match the metadata that drove the business process or were created during the process. 3. Describe flow: Describe metadata flow with the statistical and business processes (alongside the data flow and business logic). 4. Capture at source: Capture metadata at their source, preferably automatically as a by-product of other processes. 5. Exchange and use: Exchange metadata and use them for informing both computer based processes and human interpretation. The infrastructure for exchange of data and associated metadata should be based on loosely coupled components, with a choice of standard exchange languages, such as XML.
Users	1. Identify users: Ensure that users are clearly identified for all metadata processes, and that all metadata capturing will create value for them. 2. Different formats: The diversity of metadata is recognized and there are different views corresponding to the different uses of the data. Different users require different levels of detail. Metadata appear in different formats depending on the processes and goals for which they are produced and used. 3. Availability: Ensure that metadata are readily available and useable in the context of the users' information needs (whether an internal or external user).

9.3 Data Management

- Data management is essential as data are produced within many of the activities in the statistical business process and are the key outputs. The main goal of data management is to ensure that data are appropriately used and usable throughout their lifecycle. Managing data throughout their lifecycle covers activities such as planning and evaluation of data management processes as well as establishing and implementing processes related to collection, organization, use, protection, preservation and disposal of the data.
- How data are managed will be closely linked to the use of the data, which in turn is linked to the statistical business process where the data are created. Both data and the processes in which they are created must be well defined in order to ensure proper data management.
- Examples of data management activities include
  - Establishing a governance structure and assigning data stewardship responsibilities.
  - Designing data structures and associated data sets, and the flow of data through the statistical business process.
  - Identifying database (repositories) to store the data and administration of the database.
  - Documenting the data (e.g. registering and inventorying data, classifying data according to content, retention or other required classification).
  - Determining retention periods of data. and the risks that may encounter data archiving, and find solutions to overcome them, such as:
    - Limited institutional data and metadata archiving and disposal bases (for example, deciding on which releases need to be retained).
    - Difficulties that may face the archiving of administrative data collected through the project.
  - Securing data against unauthorized access and use.
  - Safeguarding data against technological change, physical media degradation, data corruption.
  - Performing data integrity checks (e.g. periodic checks providing assurance about the accuracy and consistency of data over its entire lifecycle).
  - Performing disposition activities once the retention period of the data is expired.

10 | Tenth: List of Acronyms

GSBPM	Generic Statistical Business Process Model: A flexible tool to describe and define the set of business processes needed to produce official statistics.
GAMSO	Generic Activity Model for Statistical Organizations: A reference framework describing and defining the activities that take place within a typical statistical organization.
GSIM	Generic Statistical Information Model: A reference framework of information objects, which enables generic descriptions of the definition, management and use of data and metadata throughout the statistical production process.
GPS	Global Positioning System
HL-G-MOS	High-Level Group for the Modernization of Official Statistics.
ICT	Information and Communication Technology (ICT)
UNECE	United Nations Economic Commission for Europe.
CAPI	Computer-assisted personal interviews
CATI	Computer-assisted telephone interviews.
CAWI	Computer-assisted Web Interviews.

STATS.GOV.SA