THE STATE STATISTICAL COMMITTEE
OF THE REPUBLIC OF AZERBAIJAN

GENERIC STATISTICAL BUSINESS PROCESS MODEL

(Azerbaijan version, fifth edition)

Baku - 2021

Confirmed by the decree number
04/07s of the State Statistical
Committee dated 20.08.2021.

GENERIC STATISTICAL BUSINESS PROCESS MODEL

(Azerbaijan version, fifth edition)

Baku - 2021

I. Introduction

One of the main goals of the official statistics is to reduce the cost of producing statistical products and to improve the quality of data. This requires the standardization of activities, the uniformity of the production mechanism, as well as the adjustment of the work to minimize management errors. In this reason, the General Statistical Business Process Model (GSBPM) can be taken as a flexible tool in order to required for the production of official statistics in the identification and explanation of the processes.

The main purpose of application of the GSBPM official statistical system is to provide a standardized terminology with a framework for the development of statistical metadata systems and processes. The 4th version of the GSBPM, one of the standard models for statistical process management in the international statistical system, was developed jointly by UNECE, Eurostat and OECD in April 2009. Taking into account it's broad application in 2012 the State Statistical Committee of the Republic of Azerbaijan prepared a national version based on this standard model. The "General Statistical Business Process Model (Azerbaijani version, first edition)" has been approved by order No 66/07 of the Committee on September 4, 2012. The first edition of GSBPM consisted of 9 main processes and 47 sub-processes. In 2013, the national version of the GSBPM was updated and 158 sub-process was added and the second edition of the Azerbaijani version was approved by the decree No. 30/07 of the Committee dated 12 April 2013.

The fourth version of the "General Statistical Business Process Model" has been revised and updated in 2013 and the fifth version of GSBPM developed by International organisations. Taking into account the application of the fifth version of GSBPM in all statistical activities at national and international levels, the national version of GSBPM was improved in 2015. The number of phases was reduced to 8, sub-processes to 44 and the sub sub-processes to 144. The third edition of the Azerbaijani version was approved by the Committee's Decree No 14/07 of 12 October 2015.

Applying this model in the State Statistical Committee the standardisation of all phases of statistical activity has been reflected in relevant decisions. Depending on the nature of statistical product, this model gives an opportunity to correlate activity with human resources and undertake system based on a defined technological route. For instance, all planned activities for the production of Consumer Price Index, including for specifying the needs for the indicator, designing, building, data collection, data processing, data analysis, data dissemination and evaluation. The main aim of conducting these activities on a regular basis serves to improve data quality and the production process.

The phases of GSBPM are defining the general framework of the model. Therefore, it is possible that the main processes and sub-processes are not the same for all products. Also, phases that are specific to any product may not be relevant for other products or activities. The main purpose here is to organise activities according to the adopted model.

In 2017 within the framework of the twinning project "Support to the State Statistical Committee in Harmonisation of the National Statistics System of the Republic of Azerbaijan in line with European standards" and as a result of the cooperation between experts from the statistical offices of Azerbaijan, Bulgaria, Germany and Slovenia on "Statistical Quality Management" topic further improvements, based on adopted international standards, of the Azerbaijani version of GSBPM has been done and was published the fourth edition.

A broad review of the 5th version of the GSBPM was conducted in 2018 to incorporate feedback based on practical implementation and to improve consistency with new models of High-Level Group for the Modernisation of Official Statistics[1], the GSIM and the GAMSO. Whilst the current version of the GSBPM (version 5.1) is considered as final at the time of release, it is also expected that future updates may be necessary in the coming years, either

[1] https://statswiki.unece.org/display/hlgbas

to reflect further experiences from implementing the model in practice or due to the evolution of the nature of statistical production. The fifth edition of the "General Statistical Business Process Model" has been updated based on version 5.1 of GSBPM.Thus, the current model consists of 8 phases, 44 subprocesses and 110 sub sub-processes covering standards for implementing the activities in the field of production of statistical products of State Statistical Committee.

Generally, the major differences between internationally accepted current GSBPM version 5.1 and the previous 5.0 version can be summarised as follows:

- A few sub-processes have been re-named to improve clarity;
- The duplication between the overarching processes in the GSBPM version 5.0 and the GAMSO has been resolved;
- Descriptions of the phases and the sub-processes have been updated to be less survey-centric and activities related to working with non-statistical data providers have been added where necessary;
- In recognition of the growing importance of integrating statistical data with geospatial data, descriptions have been expanded to include tasks needed to use geospatial data;
- Terminology has been modified where necessary to improve consistency with the GAMSO and the GSIM;
- Examples and descriptions have been updated and expanded to improve clarity.

II. Information about GSBPM

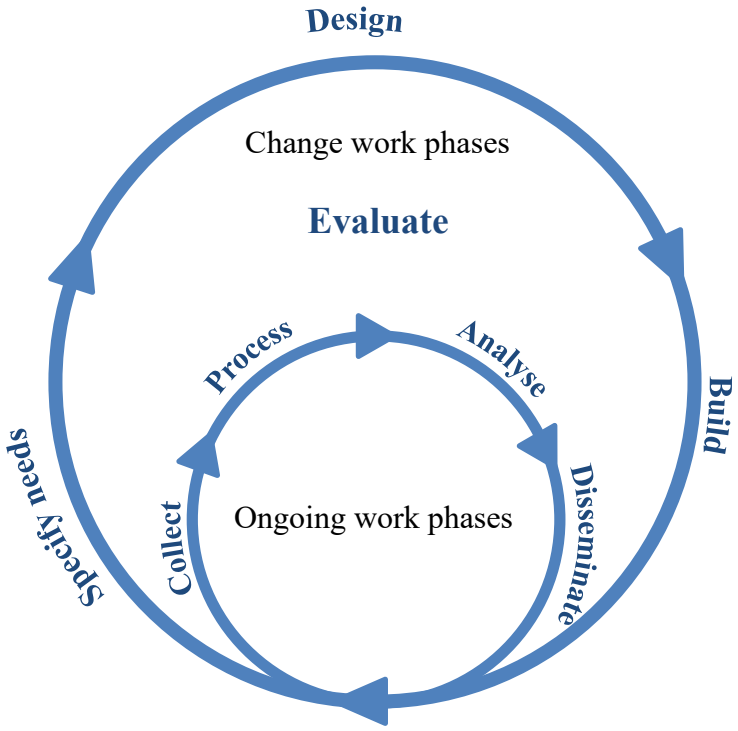
A statistical business process is a collection of related and structured activities and tasks to convert input data into statistical information. In the context of the GSBPM, organisations or groups of organisations perform statistical business processes to create official statistics to satisfy the needs of the users. The output of the process may be a mixed set of physical or digital products presenting data and metadata in different ways, such as publications, maps, electronic services, among others.

The GSBPM should be applied and interpreted flexibly. It is not a rigid framework in which all steps must be followed in a strict order, instead it identifies the possible steps in the statistical business process and the inter-dependencies between them. Although the presentation of the GSBPM follows the logical sequence of steps in most statistical business processes, the elements of the model may occur in different orders in different circumstances. Also, some sub-processes will be revisited, forming iterative loops, particularly within the Process and Analyse phases.

The GSBPM can be viewed as a checklist to make sure that all necessary steps have been considered or as a "cookbook" to identify all the "ingredients" of a statistical business process.

In many statistical organisations, the first few phases are only considered when a new output is created or when the process is revised as a result of an evaluation process. Once the output becomes part of “normal” ongoing activity, these phases are not undertaken (for example, it is not necessary to build new collection tools every time labour force survey data are collected). Following figure depicts this.

The GSBPM should therefore be seen more as a matrix, through which there are many possible paths.



Picture 1. Ongoing and change work phases of GSBPM

The structure of national version of the model

The national version of GSBPM comprises four levels:

- Level 0, the statistical business process (for example, census, entirely observation, selective observation, compilation, etc.);
- Level 1, the eight phases of the statistical business process;
- Level 2, the sub-processes within each phase;
- Level 3, the sub sub-processes of each sub-process.

Each sub-process on the phases should characterize the following identified features:

- entry of the process (access to the product);
- output of the process (result);
- goals (values);
- ownership (process management);
- methodological instructions (materials, documentation);
- opportunities;
- feedback mechanisms.

The GSBPM recognises several overarching processes with a strong statistical component that apply throughout the eight phases. These overarching processes included the list below. Quality management, metadata management and data management are elaborated further in Section of Overarching Processes. Activities that are carried out at the level of the organisation to support the statistical production are included in the GAMSO (see **“III. GSBPM and relationships with other models and frameworks (standards) ”**).

- **Quality management** – This process includes quality assessment and control mechanisms. It recognises the importance of evaluation and feedback throughout the statistical business process;

- **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;
- **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, systematising 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;
- **Provider management** – This includes cross-process burden management, as well as topics such as profiling and management of contact information (and thus has particularly close links with statistical business processes that maintain registers).

Using the GSBPM

The GSBPM is a reference model. It is intended that the GSBPM may be used by organisations to different degrees. An organisation may choose to either implement the GSBPM directly or use it as the basis for developing customised version of the model. It may be used in some cases only as a model to which organisations refer when communicating internally or with other organisations to clarify discussion. The various scenarios for the use of the GSBPM are all valid.
When organisations have developed organisation-specific adaptions of the GSBPM, they may make some specialisations to the model to fit their organisational context. The evidence so far suggests that these specialisations are not sufficiently generic to be included in the GSBPM itself.

III. GSBPM and relationships with other models and frameworks (standards)

Following the release of GSBPM, several models have been developed under the supervision of the High-Level Group for Official Statistics Modernization to support modernizing official statistics. Collectively, these are called modern statistical models. The following is a description of modern statistical models that are closely related to GSBPM.

GAMSO (The Generic Activity Model for Statistical Organizations)

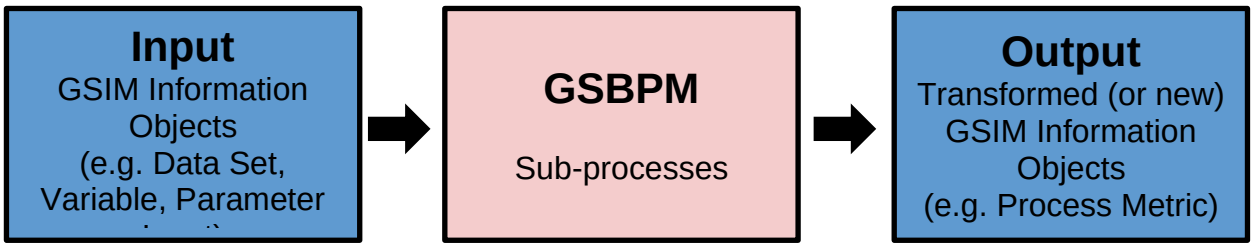
The GAMSO[2] describes and defines activities that take place within a typical statistical organisation. It extends and complements the GSBPM by adding activities needed to support statistical production (i.e. activities in the areas of strategy and leadership, capability development and corporate support). In the GSBPM 5.0, some of these activities were included as overarching processes. Activities that are not directly related to the production of statistics and/or are managed at a corporate or strategic level are now included in the GAMSO (e.g. human resource management, quality management activities that are carried out at the corporate level such as development of a quality framework). The GAMSO

[2] https://statswiki.unece.org/display/GAMSO

describes activities – that is, what statistical organisations do. It includes high level descriptions of these activities. On the other hand, the GSBPM focuses on the production process – it describes in more detail how statistical organisations undertake the activity of statistical production. Like the GSBPM, the GAMSO aims to provide a common vocabulary and framework to support international collaboration activities. Greater value will be obtained from the GAMSO if it is applied in conjunction with the GSBPM.

GSIM (The Generic Statistical Information Model)

The GSIM[3] is a reference framework for statistical information, designed to help modernise official statistics at both national and international levels. It enables generic descriptions of the definition, management and use of data and metadata throughout the statistical production process. It provides a set of standardised, consistently described information objects, which are the inputs and outputs for GSBPM sub-processes. The GSIM helps to explain significant relationships among the entities involved in statistical production and can be used to guide the development and use of consistent implementation standards or specifications. Like the GSBPM, the GSIM is one of the cornerstones for modernising official statistics and moving away from subject matter silos. It identifies around 130 information objects, examples include data sets, variables, statistical classifications, units, populations as well as the rules and parameters needed for production processes to run (e.g. data editing rules). The GSIM and the GSBPM are complementary models for the production and management of statistical information. As shown in following figure below, the GSIM helps to describe the GSBPM sub-processes by defining the information objects that flow between them, that are created in them and that are used by them to produce official statistics. Inputs and outputs can be defined in terms of information objects and are formalised in the GSIM.

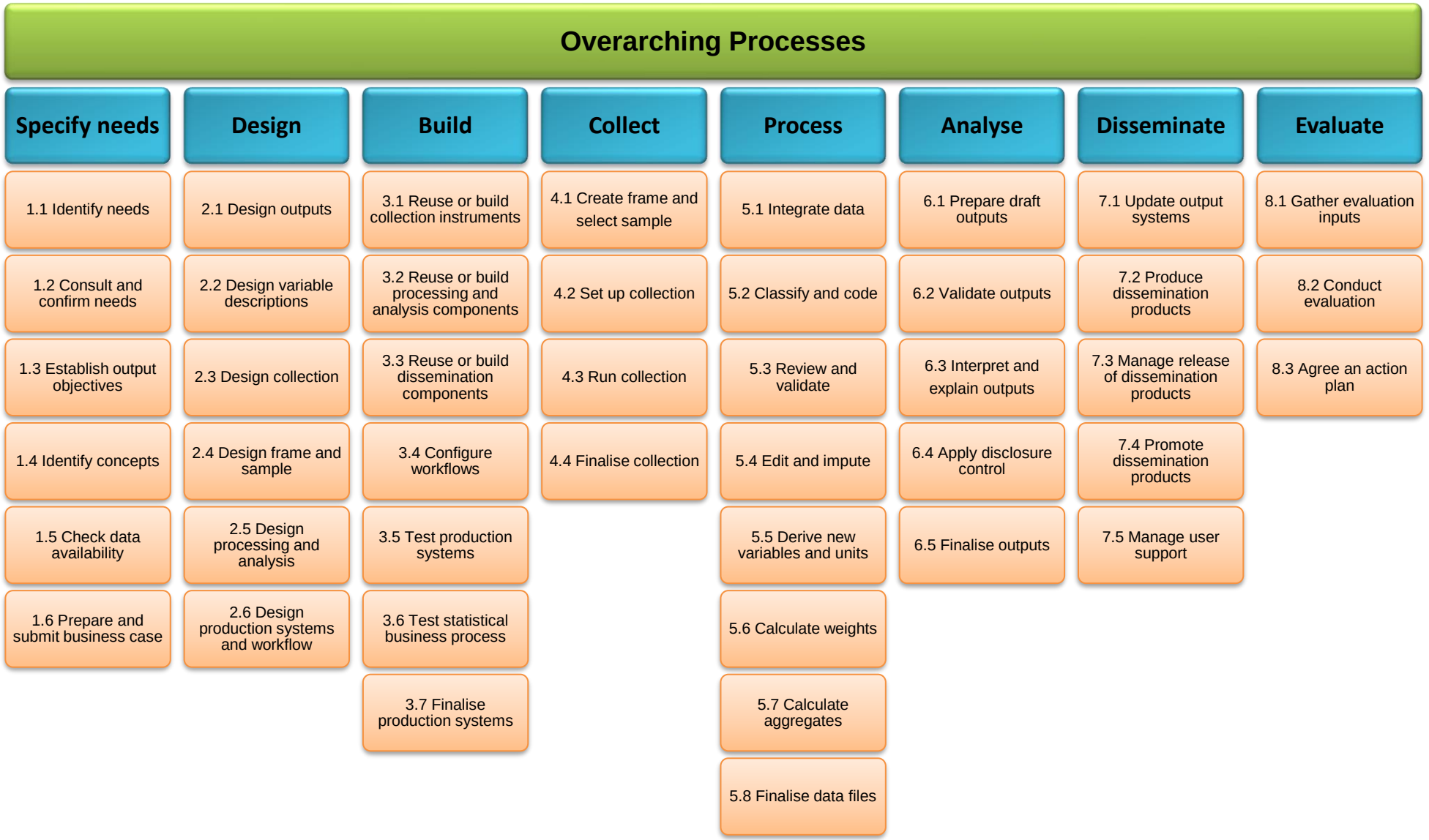


Picture 2. GSIM Information Objects as input and output instruments of the sub-processes of GSBPM

Greater value will be obtained from the GSIM if it is applied in conjunction with the GSBPM. Likewise, greater value will be obtained from the GSBPM if it is applied in conjunction with the GSIM. Nevertheless, it is possible (although not ideal) to apply one without the other. Applying the GSIM and the GSBPM together can facilitate the building of efficient metadata driven systems and help to harmonise statistical computing infrastructures.

[3] https://statswiki.unece.org/display/GSIM

IV. Description of model



V. Explanation of phrases, sub-processes and sub sub-processes

This phase is triggered when a need for new statistics is identified or feedback about current statistics initiates a review. It includes all activities associated with engaging stakeholders to identify their detailed statistical needs (current or future), proposing high level solution options and preparing a business case to meet these needs.
The "Specify needs" phase is broken down into six sub-processes, which are generally sequential, from left to right, but can also occur in parallel and can be iterative. These sub-processes are:

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

1.1.1 Identifying needs based on national priorities. The aim of this sub sub-process is to review the national strategical documents and policies for statistical information needs. It includes reviewing of strategies for development, road maps, decrees or other official documents adopted by the Azerbaijani government.

1.1.2 Study International methodologies, standards and practices and identifying needs. This sub sub-process is focused on research, review and study of relevant international methodologies, standards, practices and tendencies regarding the production of statistics. It includes also documenting of findings which can be used as input in "Design" phase, when the international methodology has to be reflected into the national one (e.g. harmonisation of definitions and concepts, improving statistical production tools, introducing new classifications and methods for collection, processing, analysing or dissemination of the data).

1.1.3 Study normative base and internal needs. This sub sub-process includes reviewing of the normative documents of the State Statistical Committee as decrees, decisions and other documents (e.g. Annual statistical program, strategies for development or improvement of the official statistics). During this process needs of other statistical domains are being identified.

1.1.4 Research, analyse users' needs and organise meetings with users. This sub sub-process includes assessment of user's satisfaction surveys, analytics of the use of statistics through the official web page of SSC as well as received questionnaires for data availability by national or international organisations. As the main input for analysing the user needs in sub-process 1.1 “Identify needs” the regularly organised meetings with key users plays very important role. The research and analysing the user needs include not only the scope of statistical products needed but also the tools and services for the dissemination of the products.

1.1.5 Prepare identified needs for discussions. This sub sub-process summarise all identified user needs in sub sub-processes from 1.1.1 to 1.1.4 in order to consult them and discuss it again with users and the Methodological Council of SSC.

1.2 Consult and confirm needs. This sub-process focuses on consulting with the internal and external stakeholders and confirming in detail the needs for the statistics. A good understanding of user needs is required so that the statistical organisation knows not only

what it is expected to deliver, but also when, ho and perhaps most importantly, why. For the second and subsequent iterations of this phase, the main focus will be on determining whether previously identified needs have changed. This detailed understanding of user needs is the critical part of this sub-process. This sub-process coverss three sub sub-processes:

1.2.1 Consultation of identified needs with internal and external stakeholders. This sub sub-process can be implemented together with activities performed in 1.1.4, where meetings with users are being organised. In these meetings, main users including data providers are invited. It aims to consult all stakeholders and receive feedback about identified user needs. In some cases, consultations are in written form (official letters or e-mails).

1.2.2 Specification of needs. This sub sub-process aims to check up the changes from previously identified needs and the level of their validity in the current iteration. For new business processes, it might include cross domain comparability of identified needs since the needs for new statistical products might be stressed in sub-process 1.1 of other related domains. For example needs for poverty indicators might be stressed by users of household budget survey or labour statistics and in the result, a new survey as the European “Statistics on income and living conditions (SILC)” might be proposed to cover the needs.

1.2.3 Discuss and confirm identified needs. This sub sub-process takes the results from 1.1.5 as well as 1.2.1 as an input for discussions of identified needs at the Methodological Council. The Council is reviewing and confirming the needs taking into consideration the relevance of the proposals, the institutional capacity, as well as any, constraints for answering thees needs on an institutional level.

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

1.3.1 Identify output objectives. In this sub sub-process, the outputs and their objectives are determined based on the results of 1.2.3 and taking into account the quality measures needed, confidentiality, possible periodicity and frequency for producing the outputs.

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

1.4.1 Compliance of the concepts according to the national legislation. This sub sub-process identified the level of compliance of the required concepts with the national legislation and their existence. In some cases, national concepts might be not existing or covering partly the concepts required. As result of which the existence of concepts on international level identified in sub sub-process 1.4.3 can be reviewed and applied.

1.4.2 Identify cross domain compliance of the concepts. This sub sub-process identify the level of compliance of the required concepts with existing concepts used by other statistical domains.

1.4.3 Ensure international comparability of the concepts. This sub sub-process identify the level of compliance between required and existing international concepts taking into account identified concepts in sub sub-process 1.4.1.

1.4.4 Consult identified concepts with stakeholders. This sub sub-process aims to inform the users (stakeholders) about findings and concepts planned to be used. It might be through organising a consultation meeting or in written form via sending official letters or e-mails.

1.5 Check data availability. This sub-process checks whether current sources of data could meet user requirements and the conditions under which they would be available including any restrictions on their use. An assessment of possible alternatives would normally include research into potential administrative or other non-statistical sources of data, to:

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

When existing sources have been assessed, a strategy for filling any remaining gaps in the data requirement is prepared. This may include identifying possible partnerships with data holders. This sub-process also includes a more general assessment of the legal framework in which data would be collected and used and may therefore identify proposals for changes to existing legislation or the introduction of a new legal framework. This sub-process covers four sub sub-processes:

1.5.1 Identify available data sources. This sub sub-process is focused on identifying available administrative or other data sources where any of the data needed is existing. It's an important step in reducing the burden on the individual respondents in case of defining of new or repetition of running production processes. Identification of the available data sources takes into account the results from sub sub-process 1.3.1 where the outputs and their objectives have been identified.

1.5.2 Analyse administrative and other data sources. This sub sub-process aims to study and analyse the identified available data sources in respect of data coverage and its metadata, methodology and concepts used as well as a review of any specific technical or legal documents which provide detailed information about the observed data source. An important result from this sub sub-process is also identification of the gaps between identified needs and available information in the administrative or other data sources.

1.5.3 Prepare plan (road map) for reducing the gaps. This sub sub-process aims to develop a plan using the results from 1.5.2 in order to determine possible methods for reducing the gaps between identified needs and available information in the administrative or other data sources (e.g. agreement for improvement of the administrative data source after consultations with data source owners, develop models for estimations or covering the gaps via standard data collection with respondents).

1.5.4 Determine proposals for changes to the existing normative legal basis. This sub sub-process aims to determine proposals for changes to the legal basis regulating the

identified administrative or other data sources, in case of need. The proposals can be developed together with the data source owners through written consultations or establishing specific cross institutional working groups. The proposals for changes may cover topics as data and metadata access (license), harmonisation of the methodology used in respect of definitions, classifications, coding applied and etc.

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

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

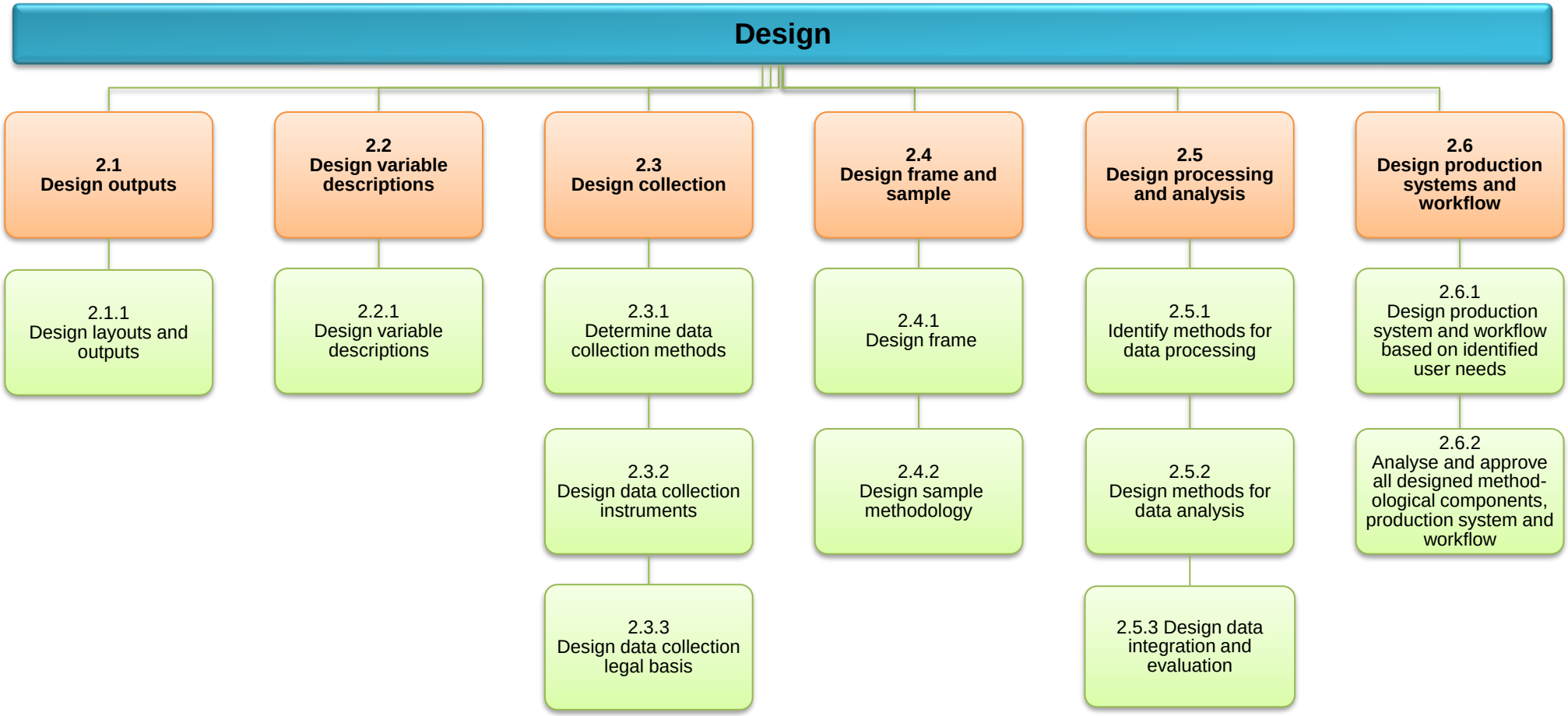
The business case describes options and makes recommendations. It may include the benefits, costs, deliverables, time frame, budget, required technical and human resources, risk assessment and impact on stakeholders for each option.
After the business case is prepared, it is submitted for approval to move to the next phase of the business process. At this sub-process, a “go”/“no go” decision is made. Typically, the business case is reviewed and formally approved or disapproved by the appropriate sponsors and governance committees. This sub-process covers four sub sub-processes:

1.6.1 Prepare working plan. In this sub sub-process, a working plan is developed based on results from previous sub sub-processes. The working plan can be new, revised or completely same as the working plan used in the previous period. In the case of revision of already existing working plan from previous periods, the level of changes is indicated. The working plan contains a description of planned activities by relevant phases of the business process, planned products and services to be produced as well as justification for the benefits.

1.6.2 Evaluate risks. This sub sub-process is used to evaluate the risks which might occur during implementation of proposed business processes and which might have a significant impact on the quality of results.

1.6.3 Estimate costs. This sub sub-process plays important role in providing estimations for the costs needed for each process planned to be implemented. It's based on the working plan prepared in sub sub-process 1.6.1.

1.6.4 Evaluation of the working plan and risks, compile of budget. In this sub sub-process, the working plan based on the results from previous sub-processes, the risk evaluation and proposed budget is provided to the "Methodological Council" and "Collegium" for review and approval. The approval of the whole business case gives "green light" for the “Design” process.


Design
2.1 Design outputs
2.1.1 Design layouts and outputs
2.2 Design variable descriptions
2.2.1 Design variable descriptions
2.3 Design collection
2.3.1 Determine data collection methods
2.3.2 Design data collection instruments
2.3.3 Design data collection legal basis
2.4 Design frame and sample
2.4.1 Design frame
2.4.2 Design sample methodology
2.5 Design processing and analysis
2.5.1 Identify methods for data processing
2.5.2 Design methods for data analysis
2.5.3 Design data integration and evaluation
2.6 Design production systems and workflow
2.6.1 Design production system and workflow based on identified user needs
2.6.2 Analyse and approve all designed method-ological components, production system and workflow

This phase describes the development and design activities and any associated practical research work needed to define the statistical outputs, concepts, methodologies, collection instruments and operational processes. It includes all the design elements needed to define or refine the statistical products or services identified in the business case. This phase specifies all relevant metadata, ready for use later in the business process, as well as quality assurance procedures. For statistical outputs produced on a regular basis, this phase usually occurs for the first iteration and whenever improvement actions are identified in the "Evaluate" phase of a previous iteration.
Design activities make substantial use of international and national standards in order to reduce the length and cost of the design process and enhance the comparability and usability of outputs. Organisations are encouraged to reuse or adapt design elements from existing processes and to consider geospatial aspects of data in the design to enhance the usability and value of the statistical information. Additionally, outputs of design processes may form the basis for future standards at the organisational, national or international levels.
The "Design" phase is broken down into six sub-processes, which are generally sequential, from left to right, but can also occur in parallel and can be iterative. These sub-processes are:

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

2.1.1 Design layouts and outputs. This sub sub-process is focused on designing the layouts and the content of outputs (statistical tables, indicators, special publications or news-releases) planned for dissemination. Design of the outputs follows internally established procedures and the dissemination policy of SSC as well as standards and objectives identified in sub sub-processes 1.1.2, 1.3.1 and 1.4.3.

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

2.2.1 Design variable descriptions. This sub sub-process includes defining of the descriptions of variables which have to be collected or derived. Descriptions give information about the name of the variables, types, definitions and codes lists to be used). In some cases, the descriptions of the variables take into account collection methods and instruments which are designed in sub sub-processes 2.3.1 and 2.3.2.

2.3 Design collection. This sub-process determines the most appropriate collection instruments and methods which may depend on the type of data collection (census, sample survey or other), the collection unit type (enterprise, person or other) and the available sources of data. The actual activities in this sub-process will vary according to the type of collection instrument required. The means: 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. This sub-process includes the design of the collection instruments, questions and response templates (in conjunction with the variables and statistical classifications designed in sub-process 2.2 “Design variable descriptions”). It also includes the confirmation of any formal agreements. Also, in the sub-process, related to the submission of additional data, as a memorandum of mutual agreements, it is substantiated the legal basis for the development of any formal agreements and for data collection. This sub-process is enabled by tools such as question libraries (to facilitate the reuse of questions and related attributes), questionnaire tools (to enable the quick and easy compilation of questions into formats suitable for cognitive testing) and agreement templates (to help standardise terms and conditions). This sub-process also includes the design of provider management systems that are specific to this business process. Where statistical organisations do not collect data directly (i.e. a third party controls the collection and processing of the data), this sub-process may include the design of mechanisms to monitor the data and the metadata to assess impacts of any change made by the third party. This sub-process covers three sub sub-processes:

2.3.1 Determine data collection methods. In this sub sub-process, the most appropriate data collection methods are determined (e.g. traditional collection via PAPI, CAPI, CATI, obtaining administrative data or using a mixture of collection methods). In many cases, the determination of the collection methods depends on the results from sub-process 1.5 and approved business case in 1.6.4.

2.3.2 Design data collection instruments. In this sub sub-process, data collection instruments are being designed. Collection instruments such as new questionnaires or modules of questions, data entry software, instructions for interviewers, respondents, IT department or data entry operators. In most cases, the data collection instruments have to be (re)designed based on identified and confirmed needs in 1.2.3 or the evaluation performed in 8.2.1 from previous iterations of the business process as well as using internal (institutional) templates with concrete rules how to prepare the collection instruments.

2.3.3 Design data collection legal basis. This sub sub-process is used to design the legal basis which supports the collection process. It may include the design of official letters with request for data access, agreements for data exchange, decrees of the head of SSC addressed to respondents informing about the purpose and needs etc. In many cases, the legal basis to be designed depends on data collection methods determined in 2.3.1.

2.4 Design frame and sample. This sub-process only applies to processes which involve data collection based on sampling, such as through statistical surveys. It identifies and specifies the population of interest, defines a sampling frame (and, where necessary, the register from which it is derived) and determines the most appropriate sampling criteria and methodology (which could include complete enumeration). Common sources for a sampling frame are administrative and statistical registers, censuses and information from other sample surveys. It may include geospatial data and classifications. This sub-process describes how these sources can be combined if needed. Analysis of whether the frame

covers the target population should be performed. A sampling plan should be made. The actual sample is created in sub-process 4.1 “Create frame and select sample”, using the methodology specified in this sub-process. This sub-process covers two sub sub-processes:

2.4.1 Design frame. This sub sub-process is used to design the frame for the sample. It includes identifying the data source or data bases (Census, administrative/official registers of any statistical units, etc.) which will be used for creating the sample frame in sub-process 4.1. It should also include analysis of target population coverage. As the specific practice of SSC here, specifications or instructions for the unit/department responsible for maintaining sampling procedures are developed.

2.4.2 Design sample methodology. This sub sub-process is used to design the sampling methodology, where it's also determined if a complete enumeration is going to be used or not. Instructions for creating the sample in sub-process 4.1.2 are also developed here. The sampling methodology includes the content, methods for defining the structure of the sample as well as procedures for replacement of sample units if needed.

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

2.5.1 Identify methods for data processing. In this sub sub-process identification of most suitable methods for data processing and analysis is done. It may include also the determination of whether the currently used method has to be revised or not, based on the review of international methodologies, standards and practices performed in 1.1.2.

2.5.2 Design methods for data analysis. In this sub sub-process, the methods and general methodology for processing and analyses are designed. It includes also the design of specifications for IT department regarding performing concrete methods for coding, validation rules, coverage of the imputation etc. In this sub sub-process also the specifications for IT unit about the content of intermediate working tables with aggregates are developed. These intermediate working tables with aggregates are obtained later in 5.7.1 through the special internally developed IT tool for obtaining aggregates from the primary database.

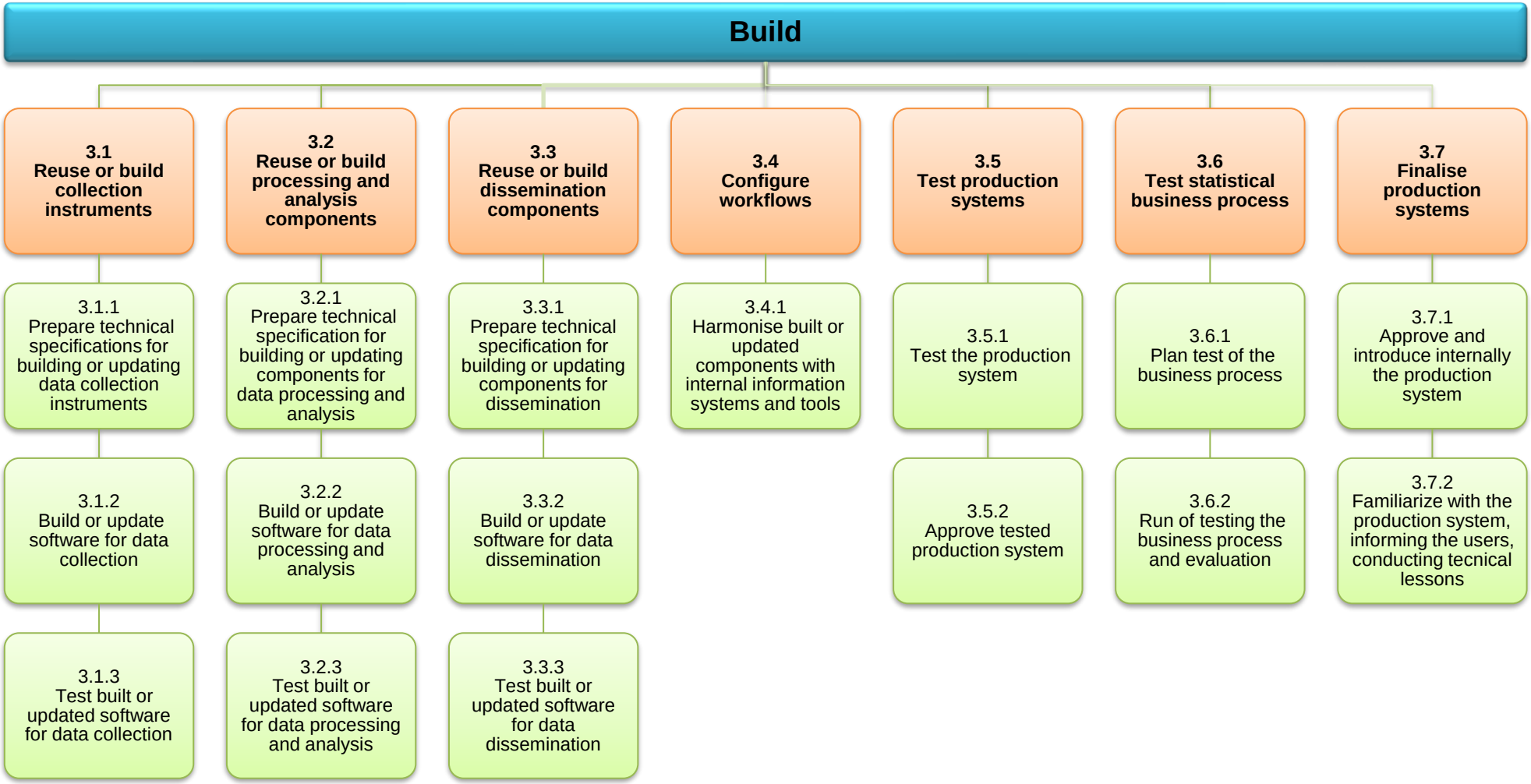
2.5.3 Design data integration and evaluation. In this sub sub-process, depending on the nature of work, it includes designing of methods of integration, verification and evaluation of data received from multiple sources.

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

specific production process, then, if any gaps are identified, new solutions should be designed. This sub-process also considers how staff will interact with systems and who will be responsible for what and when. This sub-process covers two sub sub-processes:

2.6.1 Design production system and workflow based on identified user needs. In this sub sub-process, the production system and the workflow are designed based on identified and approved user needs in sub sub-process 1.2.3. The design includes a description of the production solutions to be used. The workflow design describes the type of and sequence of activities to be performed from collection to dissemination of the statistical outputs. Identification of the responsible staff or units to perform or monitor the activities is determined also in this sub sub-process.

2.6.2 Analyse and approve all designed methodological components, production system and workflow. In this sub sub-process review of all designed methodological components, production system and workflow are performed. The reviewing body is the Methodological Council or Collegium of SSC. The review includes also consultations with relevant for the statistical domain ministries or state agencies of Azerbaijan regarding the methodology and some of the production processes. The results of the review can be an agreed methodological components, production system and workflow or recommendations for revision of any of the components. The review process can be repeated several times until final agreement of the Methodological Council and ministries is achieved. All designed and agreed methodological components, legal basis regarding collection designed in sub sub-process 2.3.3, production system and workflow are approved by the chairman of the SSC. In some cases, approvals from the relevant ministries are also required. Official approval determines that the work on the rest of production processes may start.


Build
3.1 Reuse or build collection instruments
3.2 Reuse or build processing and analysis components
3.3 Reuse or build dissemination components
3.4 Configure workflows
3.5 Test production systems
3.6 Test statistical business process
3.7 Finalise production systems
3.1.1 Prepare technical specifications for building or updating data collection instruments
3.1.2 Build or update software for data collection
3.1.3 Test built or updated software for data collection
3.2.1 Prepare technical specification for building or updating components for data processing and analysis
3.2.2 Build or update software for data processing and analysis
3.2.3 Test built or updated software for data processing and analysis
3.3.1 Prepare technical specification for building or updating components for dissemination
3.3.2 Build or update software for data dissemination
3.3.3 Test built or updated software for data dissemination
3.4.1 Harmonise built or updated components with internal information systems and tools
3.5.1 Test the production system
3.5.2 Approve tested production system
3.6.1 Plan test of the business process
3.6.2 Run of testing the business process and evaluation
3.7.1 Approve and introduce internally the production system
3.7.2 Familiarize with the production system, informing the users, conducting tecnical lessons

This phase builds and tests the production solution to the point where it is ready for use in the "live" environment. The outputs of the "Design" phase are assembled and configured in this phase to create the complete operational environment to run the process. New services are built by exception, created in response to gaps in the existing catalogue of services sourced from within the organisation and externally. These new services are constructed to be broadly reusable in alignment with the business architecture of the organisation where possible.
For statistical outputs produced on a regular basis, this phase usually occurs for the first iteration, following a review or a change in methodology or technology, rather than for every iteration.
The “Build” phase is broken down into seven sub-processes, which are generally sequential, from left to right, but can also occur in parallel and can be iterative. The first three sub-processes are concerned with the development and improvement of systems used in collection, processing, analysis and dissemination of data. The last four sub-processes focus on the end-to-end process. These sub-processes are:

3.1 Reuse or build collection instruments. This sub-process describes the activities to build and reuse the collection instruments to be used during the "Collect" phase. The collection instruments are built based on the design specifications created during the "Design" phase. A collection may use one or more modes to receive the data (e.g. personal or telephone interviews; paper, electronic or web questionnaires; SDMX web services). Collection instruments may also be data extraction routines used to gather data from existing statistical or administrative registers (e.g. by using existing service interfaces). This sub-process also includes preparing and testing the contents and functioning of that collection instrument (e.g. cognitive testing of the questions in a questionnaire). It is recommended to consider the direct connection of collection instruments to a metadata system, so that metadata can be more easily captured in the collection phase. Connecting metadata and data at the point of capture can save work in later phases. Capturing the metrics of data collection (paradata) is also an important consideration in this sub-process for calculating and analysing process quality indicators. This sub-process covers three sub sub-processes:

3.1.1 Prepare technical specifications for building or updating data collection instruments. In this sub sub-process, technical specifications for building or updating data collection instruments are prepared. The specifications are based on the results from subprocesses of the “Design” phase, where the most appropriate and needed instruments are identified. The specifications include the type of the instruments to be used, instructions for defining specific quality control mechanisms such as logical checks, skips between questions etc. This sub sub-process is performed based on consultations with IT department and software developers in cases when the data collection instruments have to be developed or updated by an external software developers. The IT department provides also an important input to the specifications regarding further harmonisation to be performed in sub sub-process 3.4.1. In the cases where administrative or other data sources are going to be collected specifications might be consulted with the data source owners in regard to services and procedures.

3.1.2 Build or update software for data collection. In this sub sub-process, collection instruments are built or updated in cooperation between the subject matter department' methodologists and IT department or external software developers) based on the specifications developed in 3.1.1.

3.1.3 Test built or updated software for data collection. In this sub sub-process initial tests of the collection components in regard to coverage of specifications, quality control

measurements, the performance of the data collection instruments are being performed. Usually done by ITs and subject matter department' methodologists. In some cases, data entry staff or chosen respondents can be involved in the tests of the data collection software.

3.2 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 “Analyse” phases, as designed in the "Design" phase. Services may include dashboard functions and features, information services, transformation functions, geospatial data services, workflow frameworks, provider and metadata management services. This sub-process covers three sub sub-processes:

3.2.1 Prepare technical specification for building or updating components for data processing and analysis. In this sub sub-process, technical specifications for building or updating components for processing and analysing are prepared. The specifications are based on the results from subprocess 2.5 of the “Design” phase. This sub sub-process is performed based on consultations between IT unit, subject matter department' methodologists and experts from the "Centre for Scientific Research and Statistical Innovation" who are responsible for sampling (if it's relevant). IT unit also provides important input to the specifications regarding further harmonisation to be performed in sub sub-process 3.4.1.

3.2.2 Build or update software for data processing and analysis. In this sub sub-process components for processing and analyse are built or updated in cooperation between the subject matter department' methodologists, IT unit (or external software developer) and experts from the "Centre for Scientific Research and Statistical Innovation" (if it's relevant). The built or updated components are based on the specifications developed in 3.2.1.

3.2.3 Test built or updated software for data processing and analysis. In this sub sub-process, initials tests of built or updated components for process and analyse are performed. In regard to coverage of the specifications developed in 3.4.1. Usually done by ITs and methodologists who are responsible to process and analyse collected data.

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

3.3.1 Prepare technical specification for building or updating components for dissemination. In this sub sub-process, technical specifications for building or updating components for dissemination are prepared. The specifications are based on the results from subprocesses of “Design” phase as well as determined dissemination methods, tools and services planned in the business case (sub-process 1.6). The specifications include the list of the products to be disseminated, types and functionalities of the tools for dissemination, rules and standards for visualization, relations with quality reports and metadata and others. This sub sub-process is performed based on consultations between subject matter departments’ methodologists, experts from departments responsible for dissemination and IT and software developers in cases when building or updating of the components has to be outsourced. The IT department provides also an important input to the specifications regarding further harmonization to be performed in sub sub-process 3.4.1

3.3.2 Build or update software for data dissemination. In this sub sub-process component for dissemination are built or updated in cooperation between the subject matter departments' methodologists, experts from dissemination and IT departments (or external software developer). The built or updated components are based on the specifications developed in 3.3.1.

3.3.3 Test built or updated software for data dissemination. In this sub sub-process, initial tests of built or updated components for dissemination are performed. Objectives of the tests are cheking of coverage of the specifications developed in 3.4.1 as well as the operational performance of the components. Usually done by IT experts and department responsible for dissemination of the statistical products.

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

3.4.1 Harmonise built or updated components with internal information systems and tools. This sub-process includes harmonization of built or updated components (tools) for collection, processing, analysing or dissemination with already set (developed) internal tools and information systems. Harmonization activities such as updates of the internal metadata system, internal registers (e.g. "Register of Indicators"), data bases (e.g. "Passports of Rayons") or classification systems. It includes also the configuration of IT tools and services for transformation processes of services for communication between different tools or databases. This process involves subject matter departments who prepared the technical specifications in sub-processes 3.1, 3.2 and 3.3.

3.5 Test production systems. This sub-process is concerned with the testing of assembled and configured services and related workflows. It includes technical testing and sign-off of new programmes and routines, as well as confirmation that existing routines from other statistical business processes are suitable for use in this case. Whilst part of this activity concerning the testing of individual components and services could logically be linked with sub-process 3.1, 3.2 and 3.3, this sub-process also includes testing of interactions between assembled and configured services and ensuring that the whole production solution works in a coherent way. This sub-process covers two sub sub-processes:

3.5.1 Test the production system. In this sub sub-process, the production system and performance of the components and related services which have been built or updated (sub-processes 3.1, 3.2 and 3.3) and harmonized with internal information systems and databases (sub sub-process 3.4.1) are being tested.

3.5.2 Approve tested production system. In this sub sub-process, the tests of the production system components which have been built or updated are being evaluated and approved by authorized unit or responsible person who should accept (receive) the components from IT department or software developer. The process includes checking the coverage and answering the requirements of the specifications developed in sub sub-processes 3.1.1, 3.2.1 and 3.3.1.

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

3.6.1 Plan test of the business process. This sub sub-process is focused on the preparation of concrete plan with actions for testing the run of the business process from the collection to dissemination at once. It's based on determined main aim and most suitable environment for testing (in-house testing or conducting a real pilot survey with respondents) of the business process.

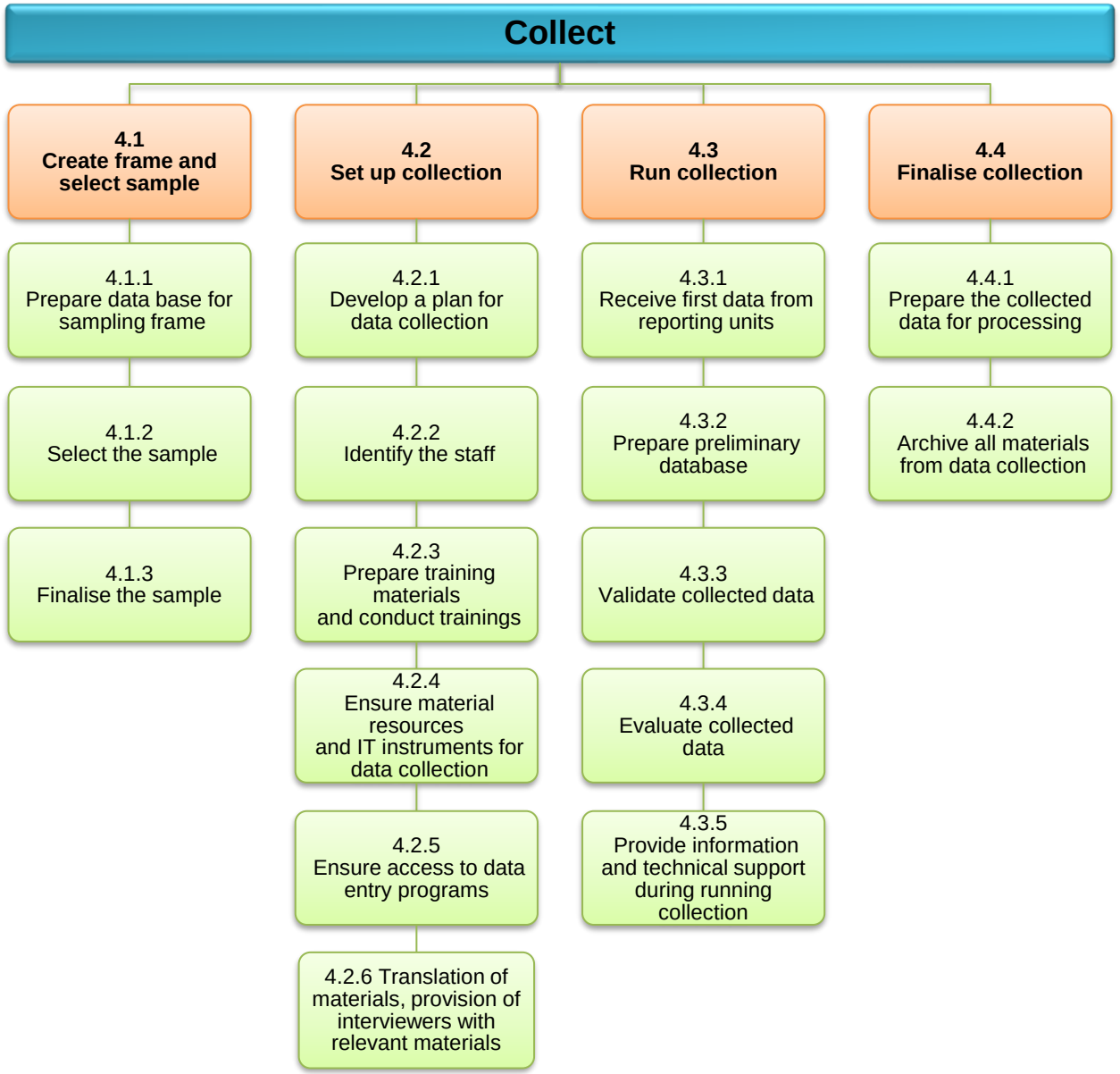
3.6.2 Run of testing the business process and evaluation. This sub sub-process is for running the test of the business process. The most important results from this sub sub-process are developing an evaluation report on how the business process is running and/or how to be improved. It might include recommendations for repetition of 3.6. or redesign of any methodological or production process components and approve them again through sub sub-process 2.6.2. The recommendations might concern the current or next iterations of the business process depending on the importance and the possibilities to reflect the changes before the start of the data collection phase.

3.7 Finalise production systems. Including two sub sub-processes, this sub-process implements activity of transferring to "ready for use" production of the assembled and configured processes and services, including modified and newly-created services, which are ready for use in practice. This activity includes the following:

- producing documentation about the process components, including technical documentation and user manuals;
- training the users on how to operate the process;
- moving the process components into the production environment and ensuring they work as expected in that environment (this activity may also be part of sub-process 3.5 "Test production systems").

3.7.1 Approve and introduce internally the production system. The aim of this sub sub-process is to approve the evaluation report from 3.6.2 after all tests and adjusted methodological or production process components (if it was recommended by the evaluation report). The approval obtained in this sub sub-process confirms that the production system is ready for the collection phase. After approval of the production system, a general presentation to the management of SSC is organized.

3.7.2 Familiarize with the production system, informing the users, conducting tecnical lessons. This sub sub-process aims to familiarize users of the production systems with its structure, procedures and instructions for operating with its components. Familiarization is done through certain technical and other trainings; as well as developing and providing user manuals and other technical documentations. The activities of this sub sub-process are performed often with the support of the IT department.


Collect
4.1 Create frame and select sample
4.1.1 Prepare data base for sampling frame
4.1.2 Select the sample
4.1.3 Finalise the sample
4.2 Set up collection
4.2.1 Develop a plan for data collection
4.2.2 Identify the staff
4.2.3 Prepare training materials and conduct trainings
4.2.4 Ensure material resources and IT instruments for data collection
4.2.5 Ensure access to data entry programs
4.2.6 Translation of materials, provision of interviewers with relevant materials
4.3 Run collection
4.3.1 Receive first data from reporting units
4.3.2 Prepare preliminary database
4.3.3 Validate collected data
4.3.4 Evaluate collected data
4.3.5 Provide information and technical support during running collection
4.4 Finalise collection
4.4.1 Prepare the collected data for processing
4.4.2 Archive all materials from data collection

This phase collects or gathers all necessary information (e.g. data, metadata and paradata), using different collection modes (e.g. acquisition, collection, extraction, transfer) and loads them into the appropriate environment for further processing. Whilst it can include validation of data set formats, it does not include any transformations of the data themselves, as these are all done in the "Process" phase. For statistical outputs produced regularly, this phase occurs in each iteration.
The "Collect" phase is broken down into four sub-processes, which are generally sequential, from left to right, but can also occur in parallel and can be iterative. These sub-processes are:

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

4.1.1 Prepare data base for sampling frame. In this sub sub-process, the data base identified in 2.4.1 for sampling frame is prepared for selection of the sample for the current iteration of data collection. Here also the quality measures are taken into account as well as maintenance of the data bases when it's used by several statistical domains for different statistical surveys or in the case of planning rotations.

4.1.2 Select the sample. In this sub sub-process, the sample has been selected from the sample frame prepared in sub sub-process 4.1.1 and according to sample methodology designed in 2.4.2.

4.1.3 Finalise the sample. In this sub sub-process selected sample is finalized with adding information important for collection and processing phases. For example - ID's of interviewers, statistical units or information which will be needed later for calculation of the quality indicators.

4.2 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. It takes place over a period of time, as it includes the strategy, planning and training activities in preparation for the specific instance of the statistical business process. Where the process is repeated regularly, some (or all) of these activities may not be explicitly required for each iteration. For one-off and new processes, these activities can be lengthy. For survey data, this sub-process includes:

- preparing a collection strategy;
- training collection staff;
- training system using supervised machine learning techniques;
- ensuring collection resources are available (e.g. laptops, collection apps, APIs);

- agreeing on terms with any intermediate collection bodies, (e.g. sub-contractors for computer assisted telephone interviewing, web services);
- configuring collection systems to request and receive the data;
- ensuring the security of data to be collected;
- preparing collection instruments (e.g. printing questionnaires, pre-filling them with existing data, loading questionnaires and data onto interviewers' computers, APIs, web scraping tools);
- providing information for respondents (e.g. drafting letters or brochures explaining the purpose of the survey, notifying respondents when online reporting instruments will be made available);
- translating of materials (e.g. into the different languages spoken or used in the country).

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

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

This sub-process covers six sub sub-processes:

4.2.1 Develop a plan for data collection. In this sub sub-process, the survey responsible unit prepares a plan for the data collection process. The plan is setting the responsible structures (RSOs) for all activities and the unit or concrete employees (methodologists) who will provide methodological support to interviewers or respondents. The plan also determines all deadlines for field work and sending the reports via data entry software to the central database. The plan includes also explanations of the institutional procedures for approaching respondends for any clarifications or validation of reported data if needed.

4.2.2 Identify the staff. In this sub sub-process, the staffs for data collection (interviewers) are being identified. The process includes preparation of the individual contracts and communication with the staff. In the case of business statistics or other similar domains where respondents/reporting units directly submit the data to the central database, a responsible contact person from SSC is being identified. This contact person has to check and update the list of contacts of the respondents.

4.2.3 Prepare training materials and conduct trainings. In this sub sub-process, training materials as instructions and methodological notes for interviewers and respondents are being prepared. The training sessions can include also presenting the changes from the previous periods regarding the collection instruments such as questionnaire forms, supporting legal or other documents, etc. Prepared materials are presented in training sessions for interviewers or sent directly to the respondents/reporting units in case of data collection without interviewers. In some cases, training sessions may include also presentation and test of new or updated data collection/entry software CATI method is applied.

4.2.4 Ensure material resources and IT instruments for data collection. In this sub sub-process collection materials and instruments (questionnaires, instructions, methodological notes, official letters and decrees) which were designed in sub-process 2.3 are printed (prepared) and distributed to relevant structures or actors in the collection process.

4.2.5 Ensure access to data entry programs. In this sub sub-process access to data entry program is ensured by providing usernames and passwords, as well as related links and instructions. It includes installing the data collection/entry software to mobile devices (tablets) or personal computers at the RSOs when it's needed. Ensuring access to data collection/entry programs it might mean that the running of the data collection can start (if it's indicated in the plan prepared in sub sub-process 4.2.1).

4.2.6 Translation of materials, provision of interviewers with relevant materials. In this sub sub-process, depending on the nature of the statistical survey (observation) to be carried out, if necessary, questionnaires will be translated into various foreign languages and materials will be presented to the relevant interviewers.

4.3 Run collection. This sub-process is where the collection is implemented. The different collection instruments are used to collect or gather the information which may include raw microdata or aggregates produced at the source, as well as any associated metadata. It can include the initial contact with providers and any subsequent follow-up or reminder actions. It may include manual data entry at the point of contact or fieldwork management, depending on the source and collection mode. It records when and how providers were contacted and whether they have responded. Depending on the geographical frame and the technology used, geo-coding may need to be done at the same time as collection of the data by using inputs from GPS systems, putting a mark on a map, etc. This sub-process also includes the management of the providers involved in the current collection. This management ensures the relationship between the statistical organisation and data providers remains positive and recording and responding to comments, queries and complaints. Proper communication with reporting units and minimisation of the number of non-respondents contribute significantly to a higher quality of the collected data. For administrative, geographical or other non-statistical data, the provider is either contacted to send the information or sends it as scheduled. This process may be time consuming and might require follow-ups to ensure that data are provided according to the agreements. In the case where the data are published as “open data” and exist in machine-readable form, they may be freely accessed and used. This sub-process may also include the monitoring of data collection and making any necessary changes to improve data quality. This includes generating reports, visualising and adjusting the acquisition process to ensure the data are fit for use. When the collection meets its targets, it is closed and a report on the collection is produced. Some basic checks of the structure and integrity of the information received may take place within this sub-process, (e.g. checking that files are in the right format and contain the expected fields). This sub-process covers five sub sub-processes:

4.3.1 Receive first data from reporting units. In this sub sub-process, first interim reports from RSOs or respondents/reporting units are received. The process includes monitoring of the timeliness and temps of receiving the reports according to planned time schedule. The monitoring is performed by the survey responsible department in the Central office.

4.3.2 Prepare preliminary database. In this sub sub-process, survey responsible department and statisticians are provided with access to internal IT tools in order to review the preliminary collected data.

4.3.3 Validate collected data. In this sub sub-process, basic automated validation procedures are applied, such as technical validation for problems in the logical control measures performance or consistency of the data formats according to the requirements. In most cases these automated validation procedures are done by the IT unit and results are provided to the subject matter departments for evaluation and further procedures.

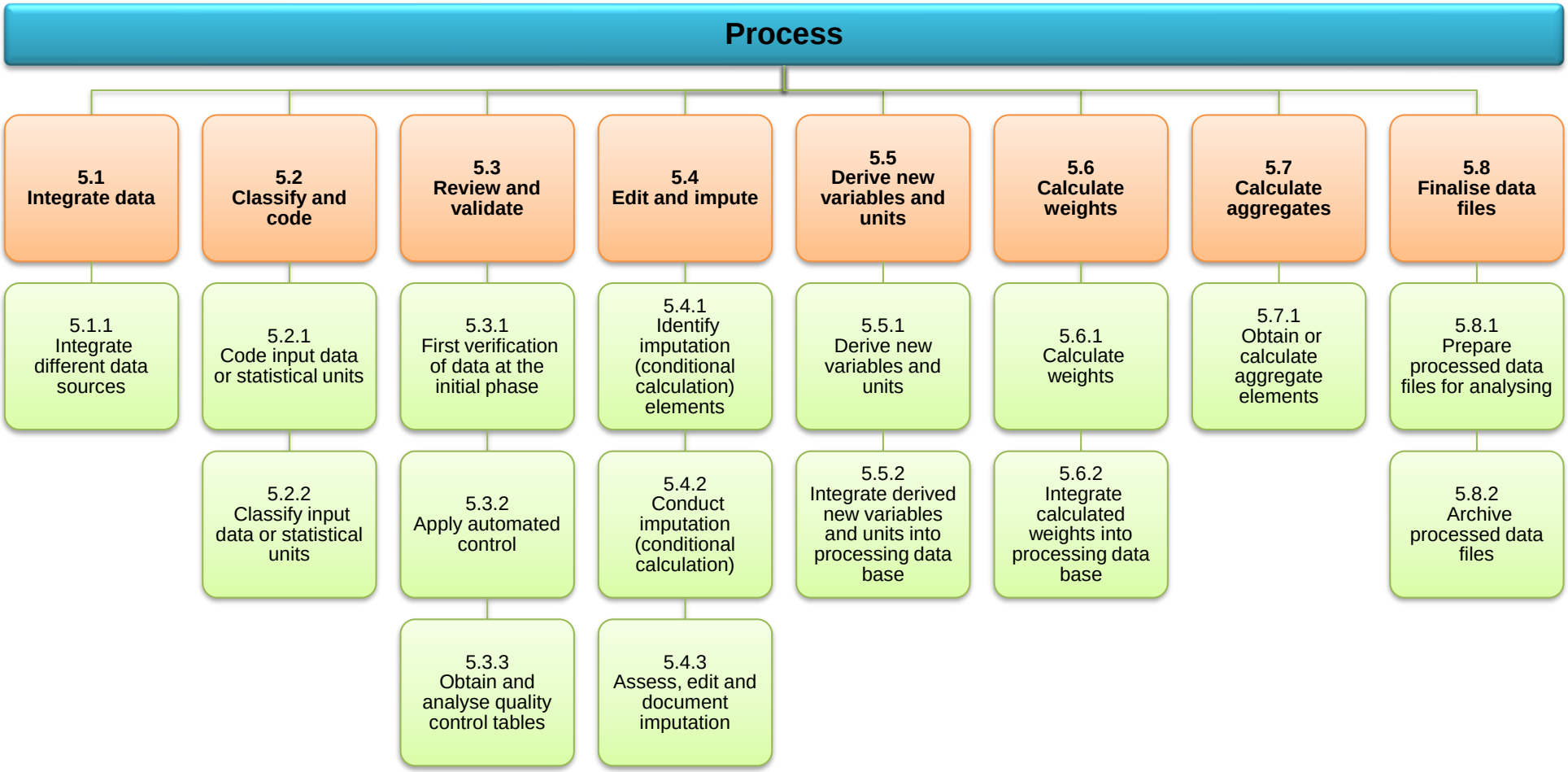
4.3.4 Evaluate collected data. This sub sub-process is for evaluation of the results from sub sub-process 4.3.3 and take measures to solve the technical problems. In case of identified technical problems and after additional measures are taken respondents are being approached to submit again their reports. The procedures on approaching the respondents for clarifications or re-submitting the reports are part of the plan developed in sub sub-process 4.2.1.

4.3.5 Provide information and technical support during running collection. In this sub sub-process activity for providing methodological or logistical support to the interviewers, respondents, RSOs etc. are foreseen. Usually the support is provided by the statistical unit/department in Central Office responsible for the overall methodology of the survey.

4.4 Finalise collection. This sub-process includes loading the collected data and metadata into a suitable electronic environment for further processing. It may include manual or automatic data capture, for example, using clerical staff or optical character recognition tools to extract information from paper questionnaires or converting the formats of files or encoding the variables received from other organisations. It may also include analysis of the metadata and paradata associated with collection to ensure the collection activities have met requirements. In cases where there is a physical collection instrument, such as a paper questionnaire, which is not needed for further processing, this sub-process manages the archiving of that material. When the collection instrument uses software such as an API or an app, this sub-process also includes the versioning and archiving of these. This sub-process covers two sub sub-processes:

4.4.1 Prepare the collected data for processing. In this sub sub-process, collected and validated data is prepared for processing and analysise. The data is uploaded in relevant databases if needed. Final checks are performed and documented.

4.4.2 Archive all materials from data collection. This sub sub-process is for archiving all developed and used materials and instruments from 4.2.3., as well documentation of performance and activities taken to support the respondents or interviewers in the collection. It includes also documentation about problems occurred during the evaluation of the collected data. These documents usually are developed in sub-process 4.3.4 and 4.3.5 by IT experts and survey responsible methodologists


Process
5.1 Integrate data
5.2 Classify and code
5.3 Review and validate
5.4 Edit and impute
5.5 Derive new variables and units
5.6 Calculate weights
5.7 Calculate aggregates
5.8 Finalise data files
5.1.1 Integrate different data sources
5.2.1 Code input data or statistical units
5.2.2 Classify input data or statistical units
5.3.1 First verification of data at the initial phase
5.3.2 Apply automated control
5.3.3 Obtain and analyse quality control tables
5.4.1 Identify imputation (conditional calculation) elements
5.4.2 Conduct imputation (conditional calculation)
5.4.3 Assess, edit and document imputation
5.5.1 Derive new variables and units
5.5.2 Integrate derived new variables and units into processing data base
5.6.1 Calculate weights
5.6.2 Integrate calculated weights into processing data base
5.7.1 Obtain or calculate aggregate elements
5.8.1 Prepare processed data files for analysing
5.8.2 Archive processed data files

This phase describes the processing of input data and their preparation for analysis. It is made up of sub-processes that integrate, classify, check, clean and transform input data, so that they can be analysed and disseminated as statistical outputs. For statistical outputs produced regularly, this phase occurs in each iteration. The sub-processes in this phase can apply to data from both statistical and non-statistical sources (with the possible exception of sub-process 5.6 “Calculate weights”, which is usually specific to survey data).
The "Process" and "Analyse" phases can be iterative and parallel. Analysis can reveal a broader understanding of the data, which might make it apparent that additional processing is needed. Sometimes the estimates being processed might be already published aggregates (undertaken according to a Revision Policy). Activities within the "Process" and "Analyse" phases may also commence before the "Collect" phase is completed. This enables the compilation of provisional results where timeliness is an important concern for users and increases the time available for analysis.
The "Process" phase is broken down into eight sub-processes, which may be sequential, from left to right, but can also occur in parallel and can be iterative. These sub-processes are:

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

- combining data from multiple sources, as part of the creation of integrated statistics such as national accounts;
- combining geospatial data and statistical data or other non-statistical data;
- data pooling, with the aim of increasing the effective number of observations of some phenomena;
- matching or record linkage routines, with the aim of linking micro or macro data from different sources;
- data fusion - integration followed by reduction or replacement;
- prioritising, when two or more sources contain data for the same variable, with potentially different values.

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

5.1.1 Integrate different data sources. This sub sub-process is focused on the integration of data sources (files) collected internally (from different surveys) or externally (obtained by administrative or other sources). This integration is relevant in cases when the outputs to be produced are a combination of already collected data fromdifferent domains. In most cases, this sub sub-process is applicable for statistical departments like National Accounts.

5.2 Classify and code. This sub-process classifies and codes the input data. For example, automatic (or clerical) coding routines may assign numeric codes to text responses according to a pre-determined statistical classification to facilitate data capture and processing. Some questions have coded response categories on the questionnaires or

administrative source of data, others are coded after collection using an automated process (which may apply machine learning techniques) or an interactive, manual process. This sub-process covers two sub sub-processes:

5.2.1 Code input data or statistical units. In this sub sub-process coding of input data according to predefined classifications, national or internal code lists etc. is performed. This coding procedures are needed for further calculations or obtaining derived variables and aggregates during the next processing stages. In some cases, coding might be performed after imputation of units or other elements done in sub sub-process 5.4.2.

5.2.2 Classify input data or statistical units. This sub sub-process is focused to classify input data to concrete classifications or to new classes or groups for later use in analyses or calculation of aggregates. Classifying of input data is based on designed outputs in sub sub-process 2.1.1. In some cases classifying might be performed after imputation of units or other elements done in sub sub-process 5.4.2.

5.3 Review and validate. This sub-process examines data to identify potential problems, errors and discrepancies such as outliers, item non-response and miscoding. It can also be referred to as input data validation. It may be run iteratively, validating data against pre-defined edit rules, usually in a set order. It may flag data for automatic or manual inspection or editing. Reviewing and validating can apply to data from any type of source, before and after integration, as well as imputed data from sub-process 5.4 “Edit and impute”. Whilst validation is treated as part of the “Process” phase, in practice, some elements of validation may occur alongside collection activities (particularly for modes such as computer assisted collection). Whilst this sub-process is concerned with detection and localisation of actual or potential errors, any correction activities that actually change the data is done in sub-process 5.4 “Edit and impute”. This sub-process covers three sub sub-processes:

5.3.1 First verification of data at the initial phase. In this sub sub-process, not automated, but manual verification of the input data is performed. This process might start even before the data collection is finalised, as the subject matter departments has access to review preliminary collected data in sub sub-process 4.3.2. The process includes detailed review and analysing of item-non responce and it’s resons, checks for major data inconsistencies, wrong classification and coding of the data. In the case of need after identified inconsistencies or errors in this process and if the reporting period is not ended, repetition of the collection process might be applied. The procedures on approaching the respondents for clarifications or re-submitting the reports are part of the plan developed in sub sub-process 4.2.1.

5.3.2 Apply automated control. In this sub sub-process, automated control procedures are performed according to the plan and defined deadlines. In the case of need after identified inconsistencies in this process and if the reporting period is not ended, repetition of the collection process might be applied.

5.3.3 Obtain and analyse quality control tables. In this sub sub-process, the quality control tables provided by the internally developed system are analysed. This system for quality control tables is developed by SSC for detecting errors and inconsistency of initially collected data. It's used by the methodologists to analyse the quality of the received input data and to verify if the initially generated aggregates (from collected data) meets the expectations. In the case of need quality control tables are sent to RSOs for verification and further processing.

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

- determining whether to add or change data;
- selecting the method to be used;
- adding / changing data values;
- writing the new data values back to the data set and flagging them as changed;
- producing metadata on the editing and imputation process.

5.4.1 Identify imputation (conditional calculation) elements. In this sub sub-process identification of the imputation elements is done. It includes identification of coverage of the imputation needed as well as the types of elements such as statistical units, items or variables to be imputed. The reason might be non-response or use of a combination of data sources planned in sub sub-process 1.5.3.

5.4.2 Conduct imputation (conditional calculation). In this sub sub-process imputation procedures are applied based on results from 5.4.1. Relevant elements are imputed to the initial input data, before deriving new variables or units. In some cases it might need to repeat sub-process 5.2 where procedures for classification and coding of the input data are applied.

5.4.3 Assess, edit and document imputation. In this sub sub-process assessment of the applied imputation is conducted. Assessment includes calculation of imputation rate and other processing quality indicators.

5.5 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, as some derived variables may themselves be based on other derived variables. It is therefore important to ensure that variables are derived in the correct order. New units may be derived by aggregating or splitting data for collection units or by various other estimation methods. Examples include deriving households where the collection units are persons or enterprises where the collection units are legal units. This sub-process covers two sub sub-processes:

5.5.1 Derive new variables and units. In this sub sub-process, new variables or units are derived based on results from sub-process 2.2. Derived new variables or units are needed for obtaining aggregates or to calculate indicators, indices or other statistical outputs.

5.5.2 Integrate derived new variables and units into processing data base. In this sub sub-process derived new variables or units in sub sub-process 5.5.1 are integrated into processing data base. In some cases, new variables might be derived through creating and calculating weights in sub sub-process 5.6.1.

5.6 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”. For example, weights can be used to "gross-up" data to make them representative of the target population (e.g. for sample surveys or extracts of scanner data) or to adjust for non-response in total enumerations. In other situations, variables may need weighting for normalisation

purposes. It may also include weight correction for benchmarking indicators (e.g. known population totals). This sub-process covers two sub sub-processes:

5.6.1 Calculate weights. In this sub sub-process, weights are created or calculated. The reasons can be non-response, grossing up factors, deriving new variables or units (sub sub-process 5.5.2) etc. In many cases, the sampling experts are involved in this sub-process.

5.6.2 Integrate calculated weights into processing data base. In this sub sub-process calculated or created weights are integrated into processing data base for further calculations or obtaining aggregates. In many cases the sampling experts are involved in this sub-process.

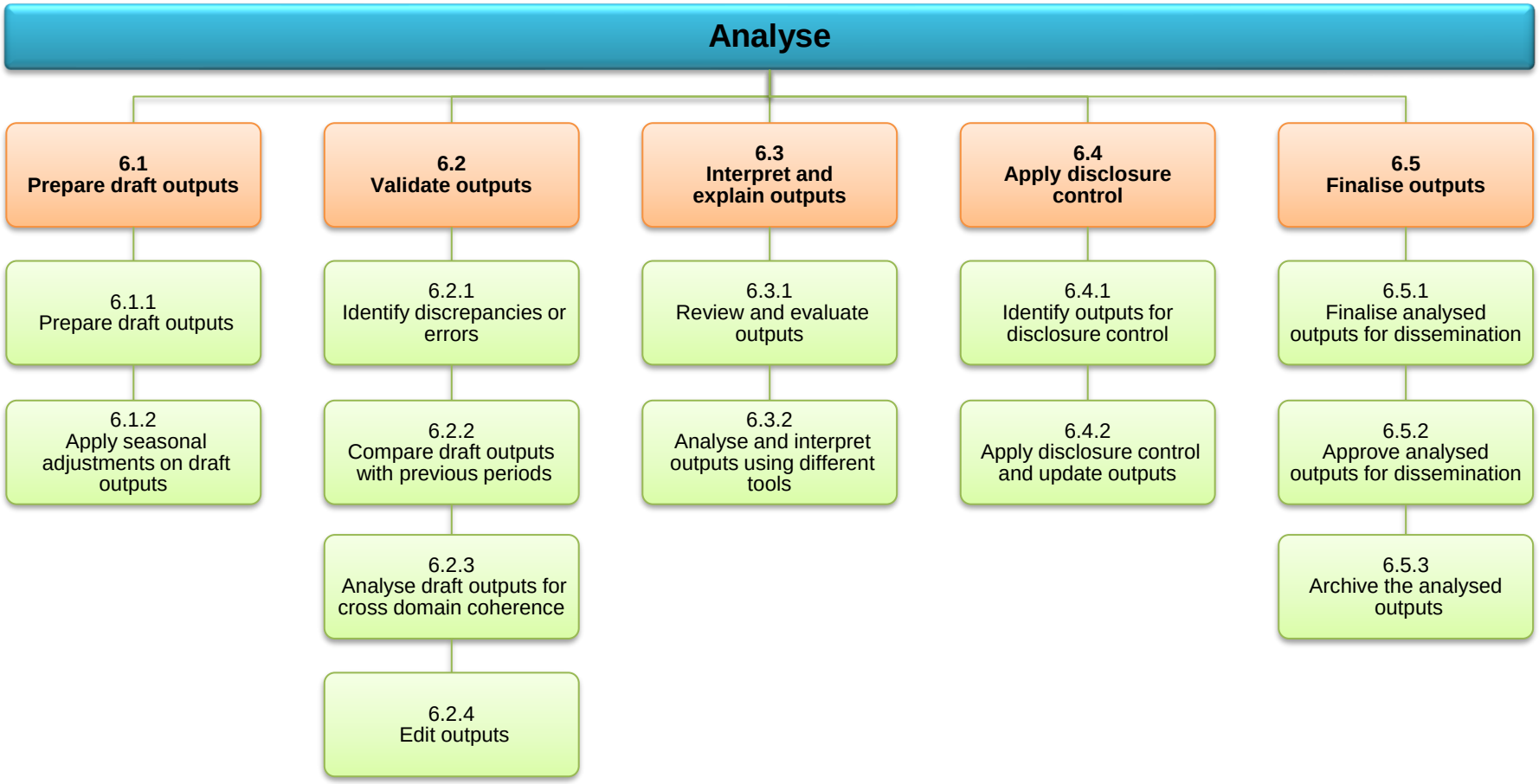
5.7 Calculate aggregates. This sub-process creates aggregate data and population totals from microdata or lower-level aggregates. It includes summing data for records sharing certain characteristics (e.g. aggregation of data by demographic or geographic classifications), determining measures of average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

5.7.1 Obtain or calculate aggregate elements. In this sub sub-process, aggregates are obtained or calculated. Obtaining may be done through a request to the primary database by a special tool which provides aggregates or calculates new aggregates derived from initial ones. It also might include applying weights from 5.6 to derive appropriate totals.

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

5.8.1 Prepare processed data files for analysing. In this sub sub-process, the processed data files (data bases) are prepared for analysing. The data files should contain the microdata, weights and aggregates processed in sub-process 5.5.

5.8.2 Archive processed data files. In this sub sub-process archiving of the processed data files and documentation about procedures performed during the phase "Process". In some cases, recalculation of quarterly data or to summarising it for the reference period is needed. For that reason, processed aggregates and data files, including all supporting files and documents needed for this processing (e.g. SPSS syntax files, data files for National Accounts, tables with weights) are being archived.


Analyse
6.1
Prepare draft outputs
6.1.1
Prepare draft outputs
6.1.2
Apply seasonal adjustments on draft outputs
6.2
Validate outputs
6.2.1
Identify discrepancies or errors
6.2.2
Compare draft outputs with previous periods
6.2.3
Analyse draft outputs for cross domain coherence
6.2.4
Edit outputs
6.3
Interpret and explain outputs
6.3.1
Review and evaluate outputs
6.3.2
Analyse and interpret outputs using different tools
6.4
Apply disclosure control
6.4.1
Identify outputs for disclosure control
6.4.2
Apply disclosure control and update outputs
6.5
Finalise outputs
6.5.1
Finalise analysed outputs for dissemination
6.5.2
Approve analysed outputs for dissemination
6.5.3
Archive the analysed outputs

In this phase, statistical outputs are produced and examined in detail. It includes preparing statistical content (including commentary, technical notes, etc.) and ensuring outputs are “fit for purpose” prior to dissemination to users. This phase also includes the sub-processes and activities that enable statistical analysts to understand the data and the statistics produced. The outputs of this phase could also be used as an input to other sub-processes (e.g. analysis of new sources as input to the “Design” phase). For statistical outputs produced regularly, this phase occurs in every iteration. The "Analyse" phase and sub-processes are generic for all statistical outputs, regardless of how the data were sourced.
The "Analyse" phase is broken down into five sub-processes, which are generally sequential, from left to right, but can also occur in parallel and can be iterative. These sub-processes are:

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

6.1.1 Prepare draft outputs. This sub sub-process is focused on preparing of initial (draft) outputs derived from the calculated aggregates in 5.7. or updated outputs from previous iterations. It also includes additional calculation of indices, new aggregations when it’s needed. Quality indicators as non-response rates, sampling errors or others required indicators are also calculated in parallel with the preparation of the draft outputs. In many cases, these calculations are performed by the "Centre for Scientific Research and Statistical Innovation". The quality indicators might be used for additional validation of outputs in sub-process 6.2. or as additional information in sub sub-process 8.2.1.

6.1.2 Apply seasonal adjustments on draft outputs. In this sub sub-process calculation of trends, seasonally adjusted time series are performed.

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

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

6.2.1 Identify discrepancies or errors. In this sub sub-process, the draft outputs are checked again for errors and discrepancy, including for such which might occurred during the process phase.

6.2.2 Compare draft outputs with previous periods. In this sub sub-process produced draft outputs are compared with the previous reference periods.

6.2.3 Analyse draft outputs for cross domain coherence. In this sub sub-process produced draft outputs are analysed for cross domain coherence when similar outputs are produced by other statistical domains.

6.2.4 Edit outputs. In this sub sub-process, macro editing of the outputs is performed in case of need. Mainly after performed cross domain coherence analyses.

6.3 Interpret and explain outputs*.* This sub-process is where the in-depth understanding of the outputs is gained by statisticians. They use that understanding to interpret and explain the statistics by assessing how well the statistics reflect their initial expectations, viewing the statistics from all perspectives using different tools and media and carrying out in-depth statistical analyses such as time-series analysis, consistency and comparability analysis, revision analysis (analysis of the differences between preliminary and revised estimates), analysis of asymmetries (discrepancies in mirror statistics), etc. This sub-process covers two sub sub-processes:

6.3.1 Review and evaluate outputs. In this sub sub-process, a review and evaluation of the outputs are done. The process is focused on reviewing and comparing the expectations about the structures of the results, trends over time, indices.

***6.3.2 Analyse and interpret outputs using different tools*.** In this sub sub-process, analyse and interpretation of the outputs are done. It includes preparing of additional interpretations, analysing the results by using different presentation or visualization tools as tables, graphics, maps etc. In some cases, additional (textual or graphical explanations) descriptions are prepared.

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

6.4.1 Identify outputs for disclosure control. This sub sub-process is focused on identifying of outputs which are an object of disclosure control. The Identification is based on determined objectives in sub sub-process 1.3.1 and internal rules for applying disclosure control. The process includes not only the identificiation of confidential outputs but also units and statistical figures.

***6.4.2 Apply disclosure control and update outputs*.** This sub sub-process is focused on applying disclosure control techniques for different type of statistical products (e.g. tables, graphs and maps) determined in sub-process 2.1. The activities performed in this sub sub-process include updating the outputs with the results of applied disclosure control. In some cases, the rate of disclosed figures or units is documented and later included as explanatory or additional notes to the final outputs. (sub sub-process 6.5.1) Disclosure control is applied

according to results of sub sub-process 6.4.1 as well as legal documents and guidelines describing procedures how to prepare outputs with confidentiality treatment.

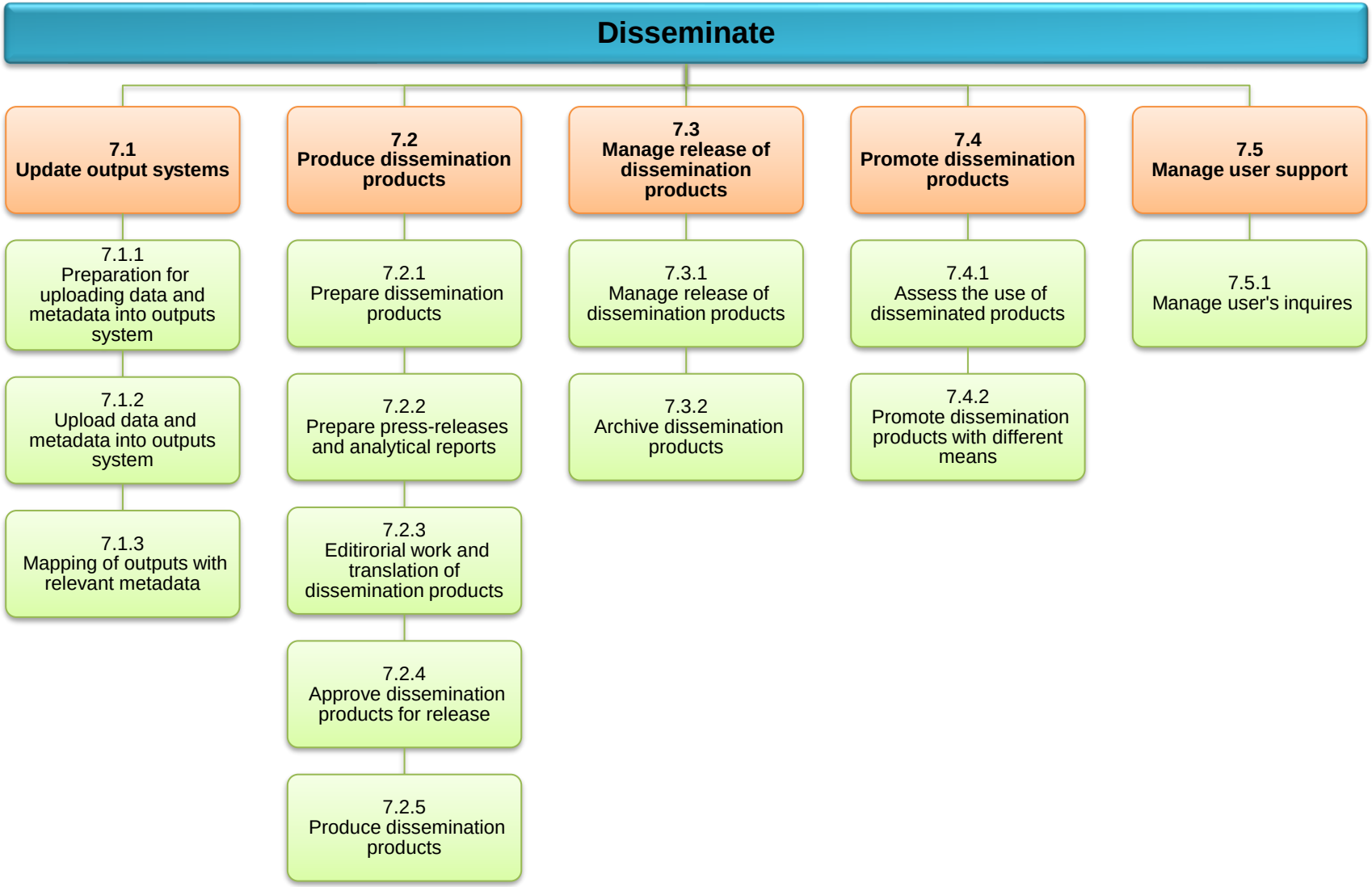
6.5 Finalise outputs. Including three sub sub-processes, this sub-process ensures the statistics and associated information are fit for purpose and reach the required quality level and are thus ready for use. It includes:

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

6.5.1 Finalise analysed outputs for dissemination. In this sub sub-process final consistency checks, preparing of supporting information and discussions with experts from related departments (as dissemination department) is performed.

6.5.2 Approve analysed outputs for dissemination. In this sub sub-process finalised (analysed) outputs are prepared for review and approval before dissemination process. The finalised outputs are approved by the Chairman.

6.5.3 Archive the analysed outputs. In this sub sub-process, all analysed outputs and related documentation are archived.


Disseminate
7.1
Update output systems
7.1.1
Preparation for uploading data and metadata into outputs system
7.1.2
Upload data and metadata into outputs system
7.1.3
Mapping of outputs with relevant metadata
7.2
Produce dissemination products
7.2.1
Prepare dissemination products
7.2.2
Prepare press-releases and analytical reports
7.2.3
Editirorial work and translation of dissemination products
7.2.4
Approve dissemination products for release
7.2.5
Produce dissemination products
7.3
Manage release of dissemination products
7.3.1
Manage release of dissemination products
7.3.2
Archive dissemination products
7.4
Promote dissemination products
7.4.1
Assess the use of disseminated products
7.4.2
Promote dissemination products with different means
7.5
Manage user support
7.5.1
Manage user's inquires

This phase manages the release of the statistical products to users. It includes all activities associated with assembling and releasing a range of static and dynamic products via a range of channels. These activities support users to access and use the products released by the statistical organisation. For statistical products produced regularly, this phase occurs in each iteration.
The “Disseminate” phase is broken down into five sub-processes, which are generally sequential, from left to right, but can also occur in parallel and can be iterative. These sub-processes are:

7.1 Update output systems. Including three sub sub-processes, this sub-process manages the update of systems (e.g. databases) where data and metadata are stored ready for dissemination purposes, including:

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

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

7.1.1 Preparation for uploading data and metadata into outputs system. In this sub sub-process finalised and approved outputs are prepared for upload into outputs data base. It can be an intermediate environment in the servers where IT department keeps the data before the official release time. It can be also structured data base of tables and files ready to be entered (uploaded) by subject matter departments directly to the dissemination data base according to the release time.

7.1.2 Upload data and metadata into outputs system. In this sub sub-process, the finalised statistical outputs (tables with indicators etc.) and related metadata files are uploaded to outputs data base or any environment provided by IT department for release the data online.

7.1.3 Mapping of outputs with relevant metadata. In this sub sub-process, the mapping between the outputs and metadata files is ensured. It can be performed by providing the instruction to the IT department about the mapping between the files.

7.2 Produce dissemination products. This sub-process which covers 5 sub subprocesses produces the dissemination products, as previously designed in sub-process 2.1 “Design outputs”, to meet user needs. They could include printed publications, press releases and websites. The products can take many forms including interactive graphics, tables, maps, public-use microdata sets, linked open data and downloadable files. Typical steps include:

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

7.2.1 Prepare dissemination products. This sub sub-process is focused on the preparation of product components and integrate them in final dissemination products.

7.2.2 Prepare press-releases and analytical reports. In this sub sub-process, press-releases and analytical reports to the governmental institutions are being prepared. It's a

standard procedure to prepare such products and being released and distributed through established dissemination and communication channels with users of these products; in general media and governmental institutions.

7.2.3 Editirorial work and translation of dissemination products. This sub sub-process is focused on the editorial work of textual content to be published according to internal dissemination policy of State Statistical Committee. It includes also a translation of the dissemination products in English language. In some cases translation in other languages might be required.

7.2.4 Approve dissemination products for release. In this sub sub-process, dissemination products which are prepared for release are being reviewed and approved by the relevant body of SSC as the Collegium and/or the Chairman of SSC.

7.2.5 Produce dissemination products. In this sub sub-process production of physical copies of statistical products is foreseen; Printing of publications, brochures and other materials.

7.3 Manage release of dissemination products. This sub-process ensures that all elements for the release are in place including managing the timing of the release. It includes briefings for specific groups such as the press or ministers, as well as the arrangements for any pre-release embargoes. It also includes the provision of products to subscribers and managing access to confidential data by authorised user groups, such as researchers. Sometimes an organisation may need to retract a product, for example, if an error is discovered. This is also included in this sub-process. This sub-process covers two sub sub-processes:

7.3.1 Manage release of dissemination products. In this sub sub-process, the real release is managed. It's based on the internal rules and data release calendar. It includes a procedure for distribution of printed copies of the products, sending analyses for the governmental bodies. Publishing of the data online and informing the users about the availability and access (internet, library, etc.) to the data.

7.3.2 Archive dissemination products. In this sub sub-process archiving of all produced and disseminated products is done. It includes archiving of the press releases and reports and other information supporting the release of the statistical products.

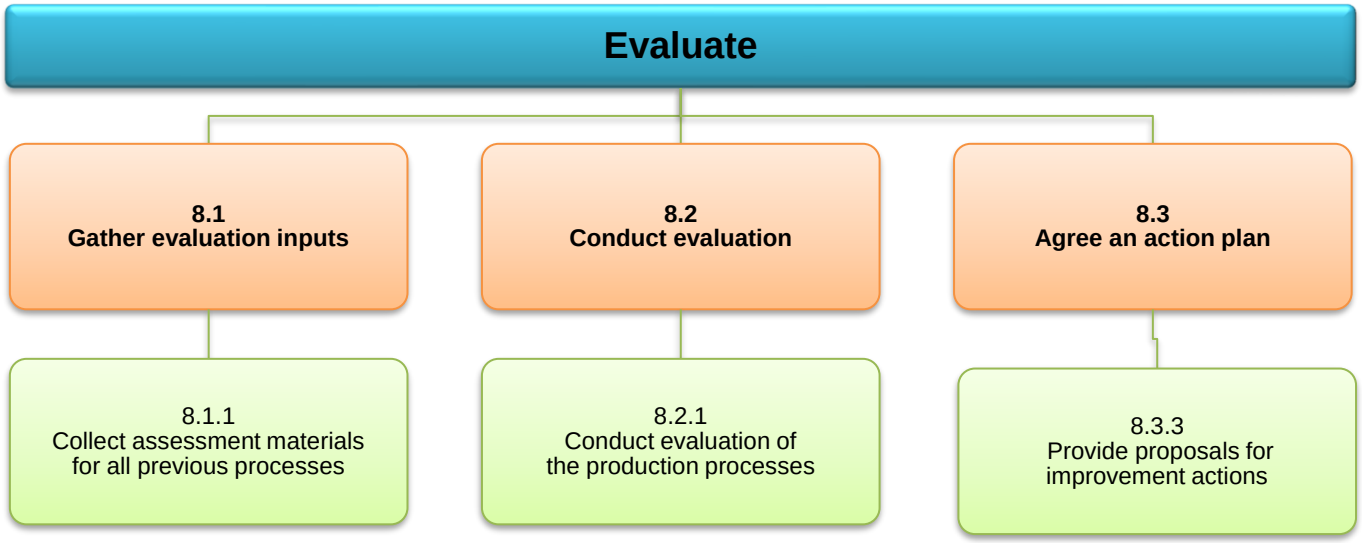
7.4 Promote dissemination products. Whilst marketing in general can be considered to be an overarching process, this sub-process concerns the active promotion of the statistical products produced in a specific statistical business process, to help them reach the widest possible audience. It includes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

7.4.1 Assess the use of disseminated products. In this sub sub-process assessment of use and user's satisfaction is done. It includes also collecting feedback in a systematic way from users and analysinges it. The analyses of the feedback and user satisfaction are an important input in the overall evaluation of the production process performed in sub sub-process 8.2.1 as well as for the sub sub-process 1.1.4 during the next iteration of the production.

7.4.2 Promote dissemination products with different means. This sub sub-process includes promotion activities based on identified needs in 1.1.4 or analyses of user satisfaction and use of the data through the institutional web page. The process includes also organisation of specific events (such as press conferences, specific to the topic workshops or sending promotional materials to different user groups, social media promotion through Facebook page). In some cases promotion of the statistical products are done by participating in national or international events (seminars, workshops) related to the use of statistical data or other related topics

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

7.5.1 Manage user's inquires. In this sub sub-process, user inquiries are reviewed and answered. It includes also consultations of users how to obtain the data needed.


Evaluate
8.1
Gather evaluation inputs
8.2
Conduct evaluation
8.3
Agree an action plan
8.1.1
Collect assessment materials
for all previous processes
8.2.1
Conduct evaluation of
the production processes
8.3.3
Provide proposals for
improvement actions

This phase manages the evaluation of a specific instance of a statistical business process, as opposed to the more general overarching process of statistical quality management described in Section of Overarching Processes. It can take place at the end of the instance of the process, but can also be done on an ongoing basis during the statistical production process. It relies on inputs gathered throughout the different phases. It includes evaluating the success of a specific instance of the statistical business process, drawing on a range of quantitative and qualitative inputs and identifying and prioritising potential improvements.For statistical outputs produced regularly, evaluation should, at least in theory occur for each iteration, determining whether future iterations should take place and if so, whether any improvements should be implemented. However, in some cases, particularly for regular and well established statistical business processes, evaluation might not be formally carried out for each iteration. In such cases, this phase can be seen as providing the decision as to whether the next iteration should start from the “Specify needs” phase or from some later phase (often the “Collect” phase).
The “Evaluate” phase is broken down into three sub-processes, which are generally sequential, from left to right, but can also occur in parallel and can be iterative. These sub-processes are:

8.1. Gather evaluation inputs. Evaluation material can be produced in any other phase or sub-process. It may take many forms, including feedback from users, process metadata (paradata), system metrics and staff suggestions. Reports of progress against an action plan agreed during a previous iteration may also form an input to evaluations of subsequent iterations. This sub-process gathers all of these inputs, compiles quality indicators and makes them available for the person or team producing the evaluation. The collection of some of these evaluation materials can be automated and take place in a continuous way throughout the whole process, as defined by the quality framework (see Quality Management Section). On the other hand, for the evaluation of certain processes it can be necessary to perform specific activities such as small surveys, (e.g. post-enumeration surveys, re-interview studies, survey on effectiveness of dissemination). This sub-process covers one sub sub-process:

8.1.1 Collect assessment materials for all previous processes. In this sub-process, all assessment materials from previous processes are being collected. This includes also evaluation reports with recommendations for improvement of the business process during the next iterations, developed in sub sub-process 3.6.2. The process activities include finalizing the quality related documents (quality reports, DESAP questionnaires, a mapping tables for GSBPM and other internal quality documents).

8.2. Conduct evaluation. This sub-process analyses the evaluation inputs, compares them to the expected/target benchmarking results (when available) and synthesises them into an evaluation report or control dashboard. The evaluation can take place at the end of the whole process (ex-post evaluation) for selected activities, during its execution in a continuous way or throughout the process, thus allowing for quick fixes or continuous improvement. The resulting report should note any quality issues specific to this iteration of the statistical business process as well as highlight any deviation of performance metrics from expected values and should make recommendations for changes if appropriate. These recommendations can cover changes to any phase or sub-process for future iterations of the process or can suggest that the process is not repeated. This sub-process covers one sub sub-process:

8.2.1 Conduct evaluation of the production processes. In this sub sub-process, all collected materials from 8.1.1 are being provided to the evaluation body (e.g. an internal working group, the Collegium or the Methodological Council) for review and conduction of final evaluation of the business case and overall quality of the business process. As result, the evaluation report is being prepared with a statement(s) for the need of concrete short or long term changes. In the case of determined need of changes in the business process, a list of recommendations with proper justifications is being developed.

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

8.3.1 Provide proposals for improvement actions. In this sub sub-process developing an action plan for improvements is foreseen. The action plan takes into account the recommendations given in the evaluation process 8.2.1. It's an important input for the "Design" phase during the next iteration of the production process. It also includes steps or actions for monitoring the implementation of the recommendations in future, through monitoring mechanism proposed (and established) on an institutional level.

VI. Overarching Processes

The GSBPM recognises several overarching processes that apply throughout the production phases and across statistical business processes. Some of these overarching processes are listed in Section II (II. Information about GSBPM, page 4). The processes of quality management, metadata management and data management are further elaborated in this Section.

Quality Management

Quality concerns organisations, products, sources and processes. In the present framework, quality management overarching process refers to product and process quality. Quality at an institutional level (e.g. adoption of a Quality Policy or Quality Assurance Framework) is considered in the GAMSO. 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 organisations 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.

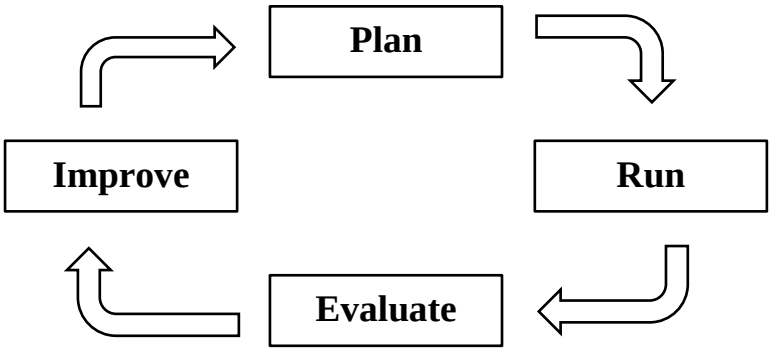
In order to improve quality, quality management should be present throughout the business process model. It is closely linked to the “Evaluate” phase, however, quality management has both a deeper and broader scope. As well as evaluating iterations of a process, it is also necessary to evaluate separate phases and sub-processes, ideally each time they are applied, but at least according to an agreed schedule. Metadata generated by the different sub-processes themselves are also of interest as an input for process quality management. These evaluations can apply within a specific process or across several processes that use common components.

In addition, a fundamental role in quality management is played by the set of quality control actions that should be implemented within the sub-processes to prevent and monitor errors and sources of risks. These should be documented and can be used for quality reporting.

Within an organisation, quality management will usually refer to a specific quality framework and may therefore take different forms and deliver different results within different organisations. The current multiplicity of quality frameworks enhances the importance of the benchmarking and peer review approaches to evaluation and whilst these approaches are unlikely to be feasible for every iteration of every part of every statistical business process, they should be used in a systematic way according to a pre-determined schedule that allows for the review of all main parts of the process within a specified time period.

Broadening the field of application of the quality management overarching process, evaluation of groups of statistical business processes can also be considered, in order to identify potential duplication or gaps.

All evaluations result in feedback, which should be used to improve the relevant process, phase or sub-process, creating a quality loop that reinforces the approach to continuous improvements and organisational learning.



Picture 3. Quality cycle

Examples of quality management activities include:

- assessing risks and implementing risk treatments to ensure fit-for-purpose quality;
- setting quality criteria to be used in the process;
- setting process quality targets and monitoring compliance;
- seeking and analysing user feedback;
- reviewing operations and documenting lessons learned;
- examining process metadata and quality indicators;
- internal or external auditing on the process.

Quality indicators support a process-oriented quality management. A suggested list of quality indicators for phases and sub-processes of the GSBPM as well as for the overarching quality and metadata management processes can be found at the “Quality Indicators for the GSBPM – for Statistics derived from Surveys and Administrative Data Sources”. Among others, they can be used as a checklist to identify gaps and/or duplication of work in the organisation.

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 GAMSO.
Good metadata management is essential for the efficient operation of statistical business processes. Metadata are present in every phase, either created, updated or carried forward from a previous phase or reused from another business process. In the context of this model, the emphasis of the overarching process of metadata management is on the creation/revision, updating, use and archiving of statistical metadata, though metadata on the different sub-processes themselves are also of interest, including as an input for quality management. The key challenge is to ensure that these metadata are captured as early as possible and stored and transferred from phase to phase alongside the data they refer to. Metadata management strategy and systems are therefore vital to the operation of this model and these can be facilitated by the GSIM.
The 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. The 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	i. Statistical Business Process Model: Manage metadata with a focus on the overall statistical business process model; ii. Active not passive: Make metadata active to the greatest extent possible. Active metadata are metadata that drive other processes and actions. Treating metadata this way will ensure they are accurate and up-to-date; iii. Reuse: Reuse metadata where possible for statistical integration as well as efficiency reasons; iv. Versions: Preserve history (old versions) of metadata.
Metadata Authority	i. Registration: Ensure the registration process (workflow) associated with each metadata element is well documented so there is clear identification of ownership, approval status, date of operation, etc.; ii. Single source: Ensure that a single, authoritative source (“registration authority”) for each metadata element exists; iii. One entry/update: Minimise errors by entering once and updating in one place; iv. Standards variations: Ensure that variations from standards are tightly managed/approved, documented and visible.
Relationship to Statistical Cycle / Processes	i. Integrity: Make metadata-related work an integral part of business processes across the organisation; ii. Matching metadata: Ensure that metadata presented to the end-users match the metadata that drove the business process or were created during the process; iii. Describe flow: Describe metadata flow with the statistical and business processes (alongside the data flow and business logic); iv. Capture at source: Capture metadata at their source, preferably automatically as a by-product of other processes; v. Exchange and use: Exchange metadata and use them for informing both computer based processes and human interpretation. The infrastructure for exchange of data and associated metadata should be based on loosely coupled components, with a choice of standard exchange languages, such as XML.
Users	i. Identify users: Ensure that users are clearly identified for all metadata processes and that all metadata capturing will create value for them; ii. Different formats: The diversity of metadata is recognised and

	there are different views corresponding to the different uses of the data. Different users require different levels of detail. Metadata appear in different formats depending on the processes and goals for which they are produced and used; iii. Availability: Ensure that metadata are readily available and useable in the context of the users' information needs (whether an internal or external user).

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, organisation, 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;
- securing data against unauthorised 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.

Acronyms

UNECE	-	The United Nations Economic Commission for Europe
OECD	-	The Organisation for Economic Co-operation and Development
EUROSTAT	-	Statistical office of the European Union
GAMSO	-	Generic Activity Model for Statistical Organisations
GSBPM	-	Generic Statistical Business Process Model
GSIM	-	Generic Statistical Information Model
METIS	-	The "brand name" for work on Statistical Metadata under the Conference of European Statisticians
GIS	-	Geographic Information System
SSC	-	State Statistical Committee
RSO	-	Regional Statistical Organisations
IT	-	Information technologies
DESAP	-	Development of Self Assessment Programme
PAPI	-	Personality and Preference Inventory
CAPI	-	Computer assisted personal interviews
CATI	-	Computer assisted telephone interviews
XML	-	eXtensible Mark-up Language: A language that defines a set of rules for encoding documents in a format that is both human-readable and machine-readable
SPSS	-	Statistical Package for the Social Sciences

Contents