

The European Union's IPA 2013 Programme

“Support to the Improvement of Statistical Information System” - Albania

Europe Aid/136334/IH/SER/AL

Service contract no. (CRIS) AL/IPA2013/03

Output 1a1:

Blueprint of a corporate statistical business process architecture

Dragan Vukmirović



This project is funded by
The European Union

A project implemented by



August 2016

The Contents of this report are the sole responsibility of Ascenta and IN2 Consultants and can in no way be taken to reflect views of the European Union

Contents

PROJECT REFERENCES	4
REPORT OVERVIEW	5
BACKGROUND	6
DESCRIPTION OF REPORT OBJECTIVES	7
METHODOLOGICAL APPROACH	8
DESCRIPTION OF GENERIC STATISTICAL BUSINESS PROCESS MODEL - GSBPM	9
CONCLUSIONS AND RECOMMENDATIONS	10
ANNEX 1. GENERIC STATISTICAL BUSINESS PROCES MODEL - GSBPM	12
1. Specify Needs Phase	15
1.1. Identify needs for the statistics	15
1.2. Consult and confirm needs	15
1.3. Establish output objectives	15
1.4. Identify concepts.....	15
1.5. Check data availability	16
2. Design Phase	17
2.1. Design outputs	17
2.2. Definition and development of the methodology for collecting data and conducting survey 17	
2.3. Design frame and sample	18
2.4. Design processing and analysis.....	18
3. Build Phase	19
3.1. Build data collection instrument.....	19
3.2. Build or enhance dissemination components	19
3.3. Configure workflows.....	19
3.4. Testing instruments for data collection and data processing.....	19
3.5. Test statistical business process.....	20
3.6. Finalise production systems.....	20
4. Collect Phase	21
4.1. Select sample.....	21
4.2. Set up collection	21
4.3. Data collection.....	21
4.4. Finalise collection	22
5. Process Phase	23
5.1. Integrate data	23
5.2. Classify and code	23



5.3.	Review, validate and edit	24
5.4.	Impute.....	24
5.5.	Derive new variables and units	24
5.6.	Calculate weights	24
5.7.	Calculate aggregates	25
5.8.	Finalise data files.....	25
6.	Analyse Phase.....	26
6.1.	Prepare draft outputs	26
6.2.	Validate outputs	26
6.3.	Interpret and explain outputs	26
6.4.	Protection of confidential data	27
6.5.	Finalise outputs.....	27
7.	Disseminate Phase	28
7.1.	Update output systems.....	28
7.2.	Produce dissemination products.....	28
7.3.	Manage release of dissemination products.....	29
7.4.	Promote dissemination products.....	29
7.5.	Manage user support.....	29
8.	Evaluate Phase	30
8.1.	Gather evaluation inputs	30
8.2.	Conduct evaluation.....	30
8.3.	Agree an action plan	30
ANNEX 2.	QUALITY MANAGEMENT	32
ANNEX 3.	METADATA MANAGEMENT.....	34
ANNEX 4.	OTHER USES OF GSBPM.....	36
ANNEX 5.	RELATIONSHIPS WITH OTHER MODELS AND STANDARDS	38
ANNEX 6:	ABBREVIATIONS AND ACRONYMS.....	41
ANNEX 7:	REFERENCES.....	42



PROJECT REFERENCES

Project number	reference	Europe Aid/136334/IH/SER/AL
Project title	Support to the Improvement of Statistical Information System	
Contract number	AL/IPA2013/03	
Location	Albania	
Contractor	The project is implemented by the consortium who is composed by two companies Computas IT Romania and IN2	
Contracting Authority	Central Finance and Contracting Unit (CFCU) within the Ministry of Finance	
Beneficiary	The national statistical institution of Albania (INSTAT)	
Project start date	06 th of June 2016	
Project end date	05 th of June 2018	
Project duration	24 months	

Status of document:

Submitted on	03 th of August 2016	
Version	1.0	
Status	Draft	
	For approval	
	Approved	x

Change history:

Version	Date	Summary of Change
V1.0	03 th of August 2016	Initial version for the Beneficiary's and CA review



REPORT OVERVIEW

Project:	Support to the Improvement of Statistical Information System
Component:	Component 1 - Redesign and improve corporate statistical business processes and enhance institutional capacities in the production and usage of Agricultural and Labour Force statistics
Sub-component:	1a. Redesign and improve corporate statistical business processes
Activity:	ACTIVITY 1a.1 Assist INSTAT in designing a blueprint of a corporate statistical business process architecture
Sub-activities:	1a.1.1: Assessment of national level and European regulations and their synchronisation 1a.1.2: Assessment of the inter institutional reporting protocols 1a.1.3: Creation of an inventory of internal and inter institutional processes 1a.1.4: Creation of an inventory of existing and envisioned tools 1a.1.5: Review of known and envisioned needs of INSTAT and relevant stakeholders 1a.1.6: Report preparation
Output:	1a.1. Blueprint of a corporate statistical business process architecture
Expert:	Dragan Vukmirovic
Date of report:	03 th of August 2016
Reporting period:	01 th July – 03 th of August 2016
Venue:	Tirana, Albania



BACKGROUND

The production processes for the various statistical domains within INSTAT over recent years have been subject of changes due to increment of the work program and the introduction of the new methodologies, technologies and techniques. This has contributed to the progress of the Institute towards its alignment with the European best practices and the EU acquis in statistics. However, this expansion may lead to the devoting insufficient attention to the standardisation and integration of processes and the usage of common tools and applications.

The report is related to the work within the IPA 2013 Project: Support to the Improvement of Statistical Information System. The **main objectives** of the Project are:

- Improving the quality, sustainability and efficiency of statistical data production as produced by INSTAT and other official statistics producers in Albania.
- Strengthening the institutional capacities of INSTAT and further aligning the statistics produced in Albania with the EU acquis in statistics.

The **specific purpose** of this project is to strengthen the capacity of INSTAT in the three main selected domains:

- Corporate statistical business processes, especially in the fields of Agricultural and Labour Force statistics;
- Corporate ICT infrastructure; with inclusion centralised Data Warehouse(s);
- Communication function and statistical coordination role of INSTAT with focus on dissemination, and quality management.

The project is conceptualized into three main components with large number of activities, sub-activities and tasks having a different degree of complexity and largely depending on a number of various factors, of internal and external nature.

Component 1 - Redesign and improve corporate statistical business processes and enhance institutional capacities in the production and usage of Agricultural and Labour Force statistics

Component 2 - Develop and test system for upgraded corporate ICT infrastructure with inclusion of establishment and pilot implementation of centralised Data Warehouse(s)

Component 3 - Strengthen communication function and statistical coordination role of INSTAT with focus on dissemination and quality management.

Functional interrelations between project/components results and activities may be presented graphically in the simplified way:

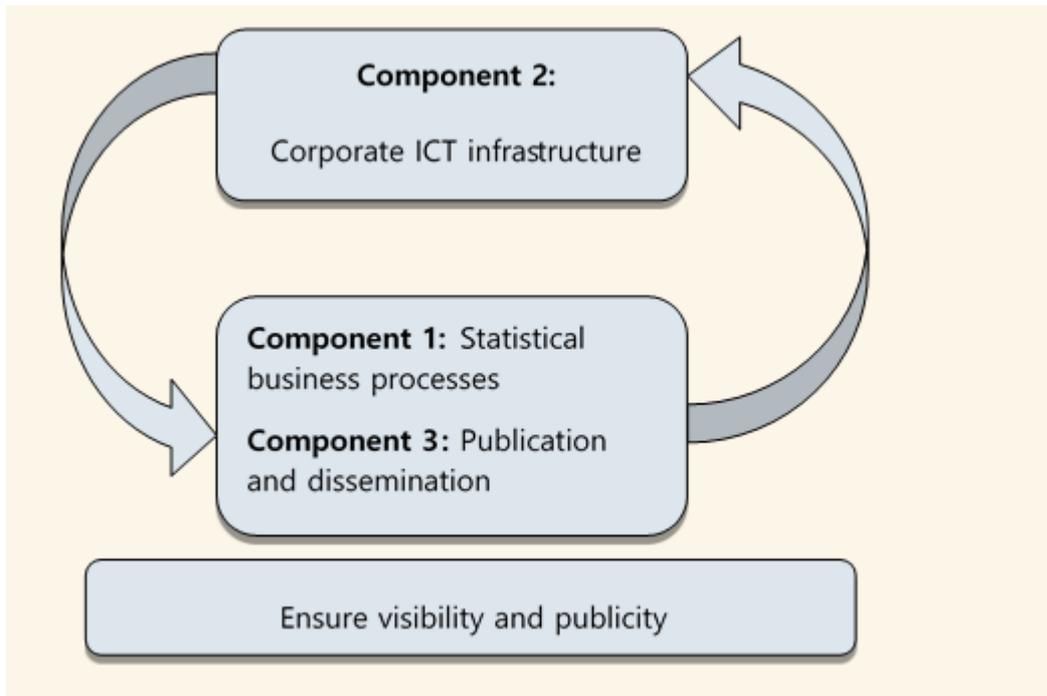


Figure 1. Functional interrelations between project/components results and activities

The activities to be performed during the Project have been singled out according to the objectives of the Project and were further defined according to discussions made during the meetings with the beneficiary's - Institute of Statistics of Albania (INSTAT) nominated experts. The institutional framework provides solid background, so that the project is managed in an efficient and transparent way at high professional level and respecting existing legislation.

This report, *Blueprint of a corporate statistical business process architecture*, is an output 1a.1. for the:

Component 1 – Redesign and improve corporate statistical business processes and enhance institutional capacities in the production and usage of Agricultural and Labour Force statistics;

Sub-component 1a. – Redesign and improve corporate statistical business processes;

Activity 1a.1 – Assist INSTAT in designing a blueprint of a corporate statistical business process architecture.

DESCRIPTION OF REPORT OBJECTIVES

Clearly defined business process is the essential for the successful production of the official statistics. The development of the business process is the base for the implementation of this Project but also for the further modernization of the statistical production processes in INSTAT. For the full implementation of the business process model several activities should be undertaken:

- the review of the existing processes *with focus on data flows*,
- identification of necessary improvements, and
- assistance in the implementation of the business process model.



Further on, the project will help INSTAT to update and upgrade in a structured way the processes' documentations in the framework of enhanced Quality management. The targeted result is the improvement of the efficiency and effectiveness of the INSTAT's operation and at the same time preparation the Institute for the establishment of centralised data warehouse(s) and for the introduction of the new information technologies and methodologies. Also the statistical business process environments shall be reviewed and upgraded.

Expected results related with the Activity 1a.1 is formalized as **Output 1a.1**. Blueprint of a corporate statistical business process architecture for INSTAT, designed based on an inventory of the actual processes, taking into account the expectations with respect to future changes. The architecture should follow best practices as implemented in the ESS, and preferably be aligned with the Generic Statistical Business Process Model (GSBPM)¹, as published by UNECE. The blueprint lays down the reference model for the other activities to be undertaken under the Component as well as form the basis for the IT systems developments as identified under Component 2.

METHODOLOGICAL APPROACH

The methodological approach for creating this Report was based on several sub-activities.

- 1a.1.1: Assessment of national level and European regulations and their synchronisation
- 1a.1.2: Assessment of the inter institutional reporting protocols
- 1a.1.3: Creation of an inventory of internal and inter institutional processes
- 1a.1.4: Creation of an inventory of existing and envisioned tools
- 1a.1.5: Review of known and envisioned needs of INSTAT and relevant stakeholders
- 1a.1.6: Report preparation

The first step was to review the Input documents and other necessary information:

- List of regulations at national level
- List of regulations at European and global level
- List with description of interinstitutional reporting protocols
- List of internal and interinstitutional processes
- List of tools (existing and envisioned)
- The list of all administrative data sources available and supporting documentation (such as definitions, data collection and checking methodologies, thresholds, quality reports and summary statistics
- Quality reports and definitions.
- Adapted Global Assessment of the National Statistical System of Albania, 2010
- Light Peer Review of the Implementation of the European Statistics Code of Practice in the Republic of Albania - Final Report, 2014
- Report on Limited Peer Review of the Institute of Statistics of Albania, 2015
- Official Statistics Programme for the period 2012-2016

¹ <http://www1.unece.org/stat/platform/display/GSBPM/GSBPM+v5.0>



- Sector review of Implementation of the Labour Force Survey in Albania, Draft report ver. March, 2016
- IPA 2014 Multi-beneficiary statistical cooperation programme, MISSION REPORT - DRAFT, Pilot Project 18, Developing a Metadata System, May, 2016
- IT/software specifications.

Furthermore, consultation with INSTAT staff regarding the business process models in use in INSTAT, as well as on the production of metadata, implementation of Single Integrated Metadata Structure (SIMS), the Euro-SDMX Metadata Structure (ESMS), the ESS Standard for Quality Reports Structure (ESQRS) were conducted. After consultation it was agreed to adopt the GSBPM as a business process model in INSTAT.

DESCRIPTION OF GENERIC STATISTICAL BUSINESS PROCESS MODEL - GSBPM

This model is intended for all activities undertaken by producers of official statistics at national and international level which result in data outputs. The GSBPM describes and defines the set of business processes needed to produce official statistics. It provides a standard framework and harmonised terminology to help statistical organisations to modernise their statistical production processes. It is designed to be independent of the data source, so it can be used for the description and quality assessment of processes based on surveys, censuses, administrative records, and other non-statistical or mixed sources. The GSBPM applies to cases where existing data are revised or time-series are re-calculated, either as a result of improved source data, or a change in methodology. In these cases, the input data are the previously published statistics, which are then processed and analysed to produce revised outputs. In such cases, it is likely that several sub-processes and possibly some phases (particularly the early ones) would be omitted. As well as being applicable for processes which result in statistics, the GSBPM can also be applied to the development and maintenance of statistical registers, where the inputs are similar to those for statistical production (though typically with a greater focus on administrative data), and the outputs are typically frames or other data extractions, which are then used as inputs to other processes².

The GSBPM is flexible to apply in all of the above scenarios and 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 a number of times forming iterative loops, particularly within the Process and Analyse phases.

GSBPM should be seen more as a matrix, through which there are many possible paths. In this way the GSBPM aims to be sufficiently generic to be widely applicable, and to encourage a standard view of the statistical business process, without becoming either too restrictive or too abstract and theoretical.

² See: <http://www1.unece.org/stat/platform/download/attachments/57835551/BR+meeting+paper2.doc>



The way the GSBPM has been adopted by agencies can be viewed on a continuum from agencies that have adopted GSBPM as is to those who have existing business process models are easily related to GSBPM. Four categories have been used. These are³:

- Agencies that have adopted GSBPM as is;
- Agencies that adopted a version of GSBPM that fits with their agency;
- Agencies that have an existing process model that can be mapped to GSBPM and
- Agencies that do not have an existing process model.

The structure of GSBPM suggested for the adoption by INSTAT is given in the Annex 1.

The GSBPM recognises several over-arching processes that apply throughout the production phases, and across statistical business processes. Two most important of them, the processes of quality management and metadata management are further elaborated in the Annexes 2 and 3, respectively.

Although the original aim of the GSBPM was to provide a basis for statistical organisations to agree on standard terminology to aid their discussions on developing statistical metadata systems and processes, this model can be used for many other purposes and some examples of use are given in Annex 4.

Relations between GSBPM and other models and standards as a result of modernisation initiatives, and similar activities outside the scope of official statistics are given in Annex 5.

A list of abbreviations and acronyms, which are included in the report is content of Annex 6.

CONCLUSIONS AND RECOMMENDATIONS

This report was prepared as a Project output 1a.1: The GSBPM as a business process model in INSTAT. The report was prepared using various existing documents provided by INSTAT and using external sources (UNECE and Eurostat on the first place) and following planned activities, objectives and scope of the project. All the information were gathered in consultations with INSTAT representatives and in respect to their requirements and expectations.

The GSBPM or other business process model is not a part of INSTAT's Official Statistics Programme for the period 2012-2016.

In Report on Limited Peer Review of the Institute of Statistics of Albania, 2015. under *Principle 4: Commitment to quality*, it is stated that INSTAT is well aware of the importance of the quality aspects of statistical products, although an over-arching quality management policy is still lacking and the co-ordination of quality work is still under consideration. It is also mentioned that more work should be devoted to overall frameworks and guidelines and quality reviews of statistical domains and regular communication about the level of quality in all areas of statistics should be enforced.

The deadline set for finalization of improvement action - *Quality guidelines describing in detail the implementation of quality management within the statistical production processes based on the*

³ National Implementations of the GSBPM – A Summary Based on METIS Case Studies, Prepared by Thérèse Lalor, Australian Bureau of Statistics, <http://www1.unece.org/stat/platform/display/GSBPM/>



Generic Statistical Business Process Model (GSBPM) or another equivalent process representation shall be developed was set for the fourth quarter of 2017.

The recommendations to be considered by INSTAT are as follows:

- Formal adoption of The GSBPM as a business process model in INSTAT,
- Updating of other procedures related to GSBPM in INSTAT,
- Implementing GSBPM in INSTAT through regular statistical practice. Some adjustments and improvements of the suggested model are possible during the implementation of the model
- Promotion of the GSBPM as business process model in INSTAT
- Consultations between the INSTAT and other producers of official statistics related to implementation of GSBPM in whole statistical system
- Regular evaluation of implementation GSBPM through making available adequate technical and human resources on behalf of the INSTAT and other involved institutions;

This documents is an input for upcoming Project activities. The blueprint lays down the reference model for the other activities to be undertaken in the other sub-components under Component 1 and is the basis for the IT systems developments as identified under Component 2.



ANNEX 1. GENERIC STATISTICAL BUSINESS PROCES MODEL - GSBPM⁴

The GSBPM was developed by the UNECE and the Conference of European Statisticians Steering Group on Statistical Metadata (better known as "METIS"). Since its release in April 2009, version 4.0 of this model has already been widely adopted by national and international statistical organisations around the world. It is intended to facilitate the convergence of statistical production processes, both within and between organisations. The Joint UNECE / Eurostat / OECD Work Sessions on Statistical Metadata (METIS) have prepared a Common Metadata Framework (CMF)⁵. Part C of this framework is entitled "Metadata and the Statistical Cycle". This part refers to the phases of the statistical business process (also known as the statistical value chain or statistical cycle) and provides generic terms to describe them. Since November 2013, this work has been taken over by the Modernisation Committee on Standards, under the High-Level Group for the Modernisation of Statistical Production and Services (HLG)⁶. The current version of the GSBPM (version 5.0) is the direct result of this work. The reader is therefore invited to check the website www.unece.org/stats/gsbpm to be sure of having the latest version.

Important:

- The GSBPM is not intended to be a rigid framework in which all steps must be followed in a strict order, but rather a model that identifies the steps in the statistical production process, and the inter dependencies between them. It aims to be sufficiently generic to be widely applicable, and to encourage a standard view of statistical production, without becoming either too restrictive or too abstract and theoretical. Different statistical production processes will follow different paths through the model, using different subprocesses, in different orders.
- 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 a number of times forming iterative loops, particularly within the Process and Analyse phases.
- GSBPM should therefore be seen more as a matrix, through which there are many possible paths. In this way the GSBPM aims to be sufficiently generic to be widely applicable, and to encourage a standard view of the statistical business process, without becoming either too restrictive or too abstract and theoretical.

The GSBPM comprises three levels:

- Level 0: statistical business process;
- Level 1: eight phases of the statistical business process;
- Level 2: sub-processes within each phase.

Level 1 assumes following phases:

1. Specify Needs,
2. Preparation and development of statistical methodologies,
3. Build necessary instruments for enforcement,

⁴ www.unece.org/stats/gsbpm

⁵ See: <http://www.unece.org/stats/cmfi/>

⁶ See: <http://www1.unece.org/stat/platform/display/hlgbas>



4. Data collection,
5. Data processing,
6. Analyse,
7. Dissemination, and
8. Evaluate.

A diagram showing the phases (level 1) and sub-processes (level 2) is presented as figure 1.1.

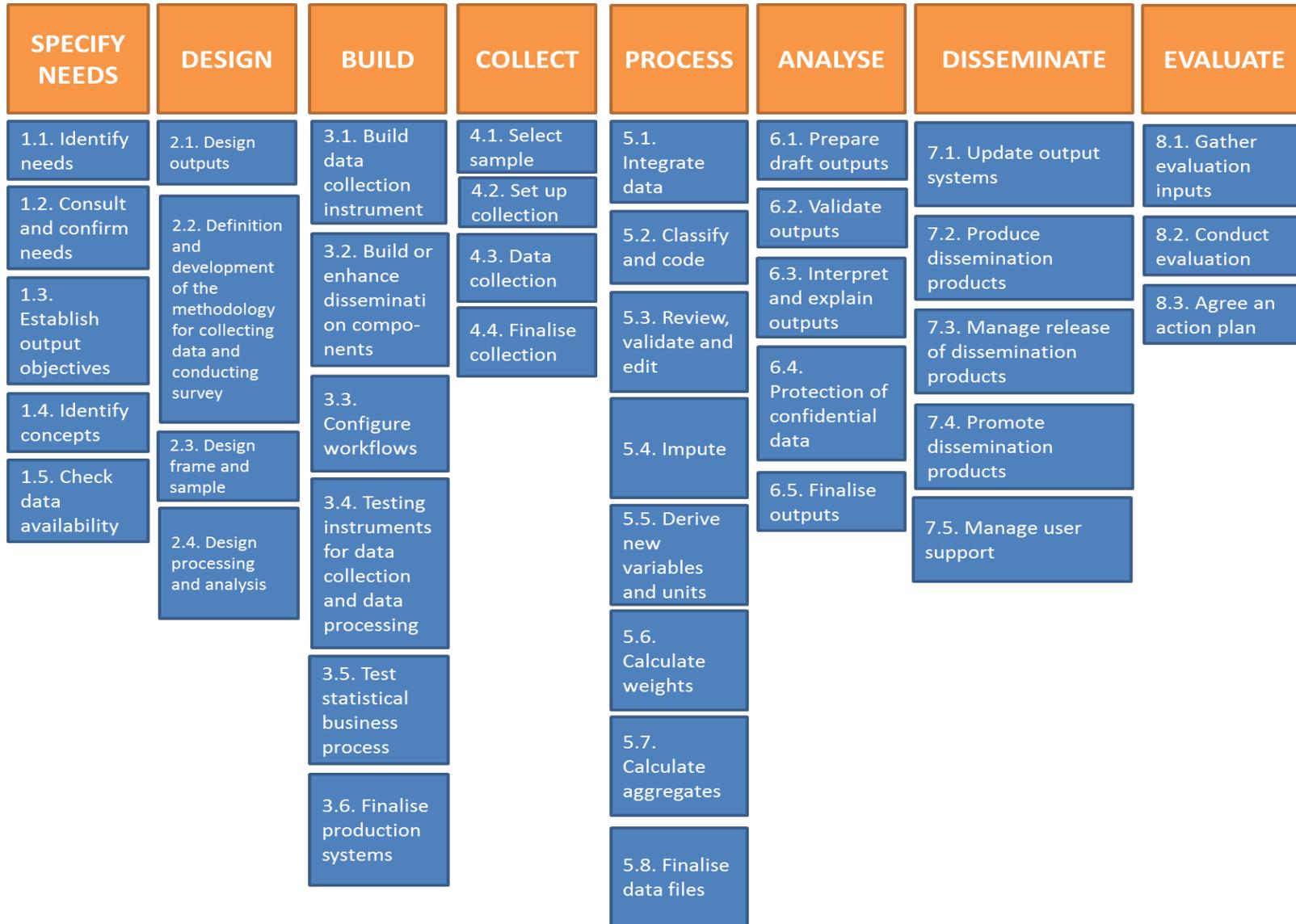


Figure 1.1. Phases (level 1) and sub-processes (level 2) in GSBPM

1. SPECIFY NEEDS PHASE

1. SPECIFY NEEDS PHASE

1.1. Identify needs

1.2. Consult and confirm needs

1.3. Establish output objectives

1.4. Identify concepts

1.5. Check data availability

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 (customers) to identify their detailed statistical needs, proposing high level solution options and preparing business cases to meet these needs.

In this phase INSTAT should take following sub-process:

- 1.1. Identifies the need for the statistics;
- 1.2. Confirms, in more detail, the statistical needs of the stakeholders;
- 1.3. Establishes the high level objectives of the statistical outputs;
- 1.4. Identifies the relevant concepts and variables for which data are required;
- 1.5. Checks the extent to which current data sources can meet these needs;
- 1.6. Prepares the business case to get approval to produce the statistics.

- **IDENTIFY NEEDS FOR THE STATISTICS**

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, an environmental change such as a reduced budget. It also includes consideration of practice amongst other (national and international) statistical organisations producing similar data, and in particular the methods used by those organisations. It may involve consideration of specific needs of different user communities, such as the disabled, or different ethnic groups.

- **CONSULT AND CONFIRM NEEDS**

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

- **ESTABLISH OUTPUT OBJECTIVES**

This sub-process identifies the statistical outputs that are required to meet the user needs identified in sub-process 1.2 (Consult and confirm 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.

- **IDENTIFY CONCEPTS**

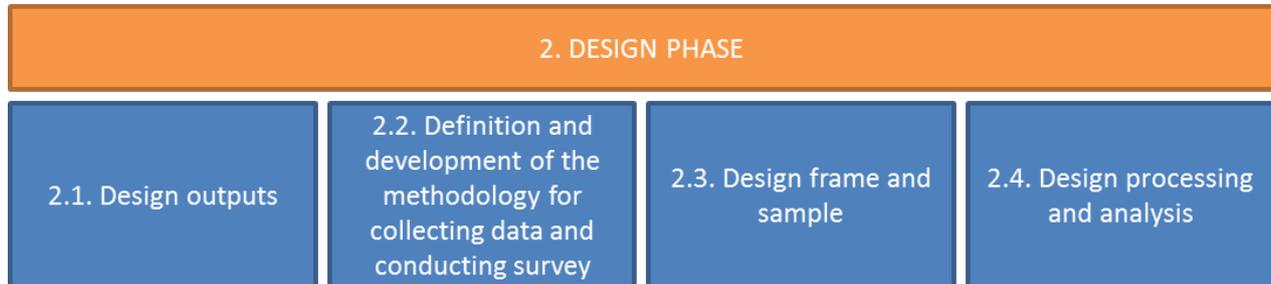
This sub-process clarifies the required concepts to be measured by the business process from the point of view of the user. At this stage the concepts identified may not align with existing statistical standards.



- **CHECK DATA AVAILABILITY**

This sub-process checks whether current data sources could meet user requirements, and the conditions under which they would be available, including any restrictions on their use. An assessment of possible alternatives would normally include research into potential administrative or other non-statistical data sources, to determine whether they would be suitable for use for statistical purposes. When existing sources have been assessed, a strategy for filling any remaining gaps in the data requirement is prepared. This sub-process also includes a more general assessment of the legal framework in which data would be collected and used, and may therefore identify proposals for changes to existing legislation or the introduction of a new legal framework.

2. DESIGN PHASE



This phase describes the development and design activities, and any associated practical research work needed to define the statistical outputs, concepts, methodologies, collection instruments⁷ 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 statistical 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 to comparability and usability of outputs.

- **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. Disclosure control methods, as well as 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, international standards, and information about practices in other statistical organisations.

- **DEFINITION AND DEVELOPMENT OF THE METHODOLOGY FOR COLLECTING DATA AND CONDUCTING SURVEY**

It includes making all the necessary methodologies (methods, instruments for data collection, variables, definitions, descriptions, instructions, agreements and contracts with providers of data, content of the questionnaire, dissemination plan, etc.).

Preparation of metadata descriptions of collected and derived variables and classifications is a crucial prerequisite for the next stages. It determines the most appropriate methods and instruments for data collection. Activities depend on the methods of data collection (CAPI, PAPI, CATI, CAWI) including testing instruments. They are made with all formal agreements on the delivery of data, such as memoranda of understanding and confirmation of the legal basis for data collection.

⁷ For GSBPM purposes, collection instruments are defined broadly to include any tool or routine to gather or extract data and metadata, from paper questionnaires to web-scraping tools. In GSIM version 1.1, collection instruments are "exchange channels" used for incoming information.



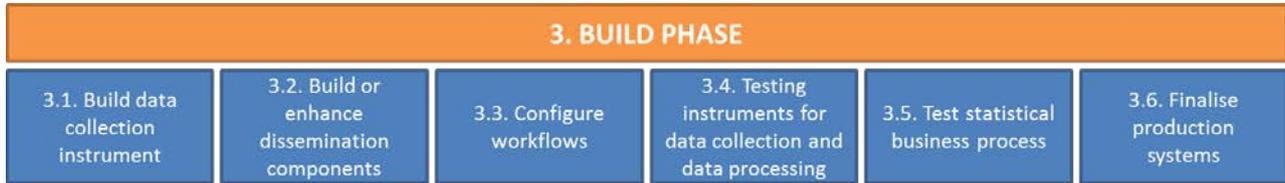
- **DESIGN FRAME AND SAMPLE**

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. This sub-process describes how these sources can be combined if needed. Analysis of whether the frame covers the target population should be performed.

- **DESIGN PROCESSING AND ANALYSIS**

This sub-process designs the statistical processing methodology to be applied during the "Process" and "Analyse" phases. This can include specification of routines for coding, editing, imputing, weighting, estimating, integrating, validating and finalizing data sets.

3. BUILD PHASE



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 direct the selection of reusable processes, instruments, information, and services that 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.

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

- **BUILD DATA COLLECTION INSTRUMENT**

This sub-process describes the activities to build the collection instruments to be used during the "Collect" phase. The collection instrument is generated or 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 hubs. Collection instruments may also be data extraction routines used to gather data from existing statistical or administrative data sets. This sub-process also includes preparing and pilot testing the contents and functioning of that instrument (e.g. testing the questions in a questionnaire).

- **BUILD OR ENHANCE DISSEMINATION COMPONENTS**

This sub-process describes the activities to build new and enhance existing components and services needed for the dissemination of statistical products. All types of dissemination components and services are included, from those that are used to produce traditional paper publications to those that provide web services, open data outputs, or access to micro-data.

- **CONFIGURE WORKFLOWS**

This sub-process describes actions to be taken for the construction of new and improvement of existing software components required for a business process, designed in the preparation phase. Components may include control tables and reports, databases, tables of results, tools for data transformation, data management tools and metadata.

- **TESTING INSTRUMENTS FOR DATA COLLECTION AND DATA PROCESSING**

This sub-process includes technical testing and approval of new programs and procedures. Includes testing the interaction between the components and ensures that the production system is



functioning as a coordinated set of components. Includes data collection for experimental studies in order to test instruments for data collection. This is followed by processing and analysing of the collected data. After experimental studies it might be required to return to previous steps and make adjustments.

- **TEST STATISTICAL BUSINESS PROCESS**

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

- **FINALISE PRODUCTION SYSTEMS**

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

- producing documentation about the process components, including technical documentation and user manuals
- training the business users on how to operate the process
- moving the process components into the production environment, and ensuring they work as expected in that environment.



4. COLLECT PHASE

4. COLLECT PHASE

4.1. Select sample

4.2. Set up
collection

4.3. Data collection

4.4. Finalise
collection

This phase collects or gathers all necessary information (data and metadata), using different collection modes (including extractions from statistical, administrative and other non-statistical registers and databases), and loads them into the appropriate environment for further processing.

- **SELECT SAMPLE**

This sub-process selects the sample for this iteration of the collection. It also includes the coordination of samples between instances of the same statistical business process (for example to manage overlap or rotation), and between different processes using a common frame or register (for example to manage overlap or to spread response burden). Quality assurance 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 sources). After sample selection follows the preparation of draft address lists and all other ancillary activities.

- **SET UP COLLECTION**

This sub-process ensures that the people, processes and technology are ready to collect data and metadata, in all modes as designed. Where the process is repeated regularly, some (or all) of these activities may not be explicitly required for each iteration. For one-off and new processes, these activities can be lengthy. This sub-process includes:

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

- **DATA COLLECTION**

This sub-process is where the collection is implemented, with the different instruments being used to collect or gather the information, which may include raw micro-data or aggregates produced at the source, as well as any associated metadata. It includes the initial contact with providers and any subsequent follow-up or reminder actions. When the collection meets its targets (which usually manifests in response rate), it is closed and a report on the collection is produced. Some basic validation 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. All validation of the content takes place in the Process phase.



- **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 take-on, for example using clerical staff or optical character recognition (OCR) tools to extract information from paper questionnaires, or converting the formats of files received from other organisations. It may also include analysis of the process metadata (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.

5. PROCESS PHASE



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

- **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 data sources, and a variety of collection modes, including extracts of administrative 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
- matching / record linkage routines, with the aim of linking micro or macro data from different sources
- 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. After integration, depending on the needs of data protection, data can be anonymised, i.e. the identifiers can be removed such as name and address in order to protect confidentiality. It includes the classification and coding of the input data.

- **CLASSIFY AND CODE**

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



- **REVIEW, VALIDATE AND EDIT**

This sub-process examines data to try to identify potential problems, errors and discrepancies such as outliers, item non-response and miscoding. It may be run iteratively, validating data against predefined 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. Some elements of validation may occur alongside collection activities, particularly for modes such as web collection.

- **IMPUTE**

Where data are considered incorrect, missing or unreliable, new values may be inserted 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:

- the determination of whether to add or change data;
- the selection of the method to be used;
- adding / changing data values;
- writing the new data values back to the data set, and flagging them as changed;
- the production of metadata on the editing and imputation process.

- **DERIVE NEW VARIABLES AND UNITS**

This sub-process derives data for variables and units that are not explicitly provided in the collection, but are needed to deliver the required outputs. It derives new variables by applying arithmetic formulae to one or more of the variables that are already present in the dataset, or applying different model assumptions. This activity may need to be iterative, 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.

- **CALCULATE WEIGHTS**

This sub process creates weights for unit data records. In the case of sample surveys, weights can be used to "gross-up" results to make them representative of the target population, or to adjust for



non-response in total enumerations. In other situations, variables may need weighting for normalisation purposes.

- **CALCULATE AGGREGATES**

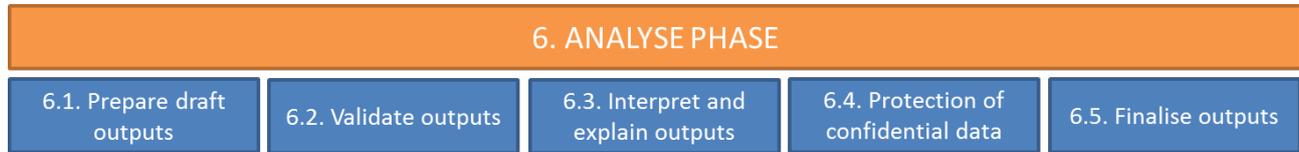
This sub process creates aggregate data and population totals from micro-data or lower-level aggregates. In the case of sample surveys, sampling errors may also be calculated in this sub-process, and associated to the relevant aggregates.

- **FINALISE DATA FILES**

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



6. ANALYSE PHASE



In this phase, statistical outputs are produced, examined in detail and made ready for dissemination.

- **PREPARE DRAFT OUTPUTS**

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

- **VALIDATE OUTPUTS**

This sub-process is where statisticians validate the quality of the outputs produced, in accordance with a general quality framework and with expectations. 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 and paradata (process metadata) are present and in line with expectations
- 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.

- **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 produced for this cycle by assessing how well the statistics reflect their initial expectations, viewing the statistics from all perspectives using different tools and media, and carrying out in-depth statistical analyses.



- PROTECTION OF CONFIDENTIAL DATA

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

- FINALISE OUTPUTS

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

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



7. DISSEMINATE PHASE



This phase manages the release of the statistical products to customers. It includes all activities associated with assembling and releasing a range of static and dynamic products via a range of channels. These activities support customers to access and use the outputs released by the statistical organisation.

For statistical outputs produced regularly, this phase occurs in each iteration.

- **UPDATE OUTPUT SYSTEMS**

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

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

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.

- **PRODUCE DISSEMINATION PRODUCTS**

This sub-process produces the products, as previously designed, to meet user needs. They could include printed publications, press releases and web sites. The products can take many forms including interactive graphics, tables, public-use micro-data sets and downloadable files. Typical steps include:

- preparing the product components (explanatory text, tables, charts, quality statements etc.);



- assembling the components into products;
- editing the products and checking that they meet publication standards.

- **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.

- **PROMOTE DISSEMINATION PRODUCTS**

This sub-process concerns the active promotion of the statistical products produced in a specific statistical business process, to help them reach the widest possible audience. It includes the use of customer relationship management tools, to better target potential users of the products, as well as the use of tools including web sites, wikis and blogs to facilitate the process of communicating statistical information to users.

- **MANAGE USER SUPPORT**

This sub-process ensures that customer queries and requests for services such as micro-data 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 over-arching quality management process, as they can indicate new or changing user needs.



8. EVALUATE PHASE



This phase manages the evaluation of a specific instance of a statistical business process. It logically takes place at the end of the instance of the process, but 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.

- **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, and makes them available for the person or team producing the evaluation.

- **CONDUCT EVALUATION**

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

- **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.

Other processes in GSBPM



The GSBPM also recognises several over-arching processes that apply throughout the eight phases, and across statistical business processes.

These can be grouped into two categories:

- 1) Those that have a **statistical component** (considered to be more important in the context of this model):
 - **Quality management** - includes quality assessment and control mechanisms. It recognises the importance of evaluation and feedback throughout the statistical business process;
 - **Metadata management** - Metadata are generated and processed within each phase, there is, therefore, a strong requirement for a metadata management system to ensure 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** - includes process-independent considerations such as general data security, custodianship and ownership, data quality, archiving rules, preservation, retention and disposal;
 - **Process data management** - includes the management of data and metadata generated by and providing information on all parts of the statistical business process.
 - **Knowledge management** - ensures that statistical business processes are repeatable, mainly through the maintenance of process documentation;
 - **Statistical framework management** - includes developing standards, for example methodologies, concepts and classifications that apply across multiple processes;
 - **Statistical program management** - includes systematic monitoring and reviewing of emerging information requirements and emerging and changing data sources across all statistical domains. It may result in the definition of new statistical business processes or the redesign of existing ones;
 - **Provider management** - includes cross-process burden management, as well as topics such as profiling and management of contact information (and thus has particularly close links with statistical business processes that maintain registers);
 - **Customer management** - includes general marketing activities, promoting statistical literacy, and dealing with non-specific customer feedback.

- 2) Those that are **more general**, and could apply to any sort of organization (recognised as they have (often indirect) impacts on several parts of the model):
 - Human resource management;
 - Financial management;
 - Project management;
 - Legal framework management;
 - Organisational framework management;
 - Strategic planning.

The processes of quality management and metadata management are further elaborated in Annexes.

ANNEX 2. QUALITY MANAGEMENT

The GSBPM recognises several over-arching processes that apply throughout the production phases, and across statistical business processes.

Quality concerns organisations, processes and products. In the present framework, quality management over-arching process refers mainly to product and process quality. The main goal of quality management within the statistical business process is to understand and manage the quality of the statistical products. There is general agreement among statistical organisations that quality should be defined according to the ISO 9000-2005 standard: "The degree to which a set of inherent characteristics fulfils requirements"⁸. Thus, product 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 the product quality, quality management should be present throughout the statistical business process model. It is closely linked to Phase 8 (Evaluate), which has the specific role of post-evaluating individual instances of a statistical business process. 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.

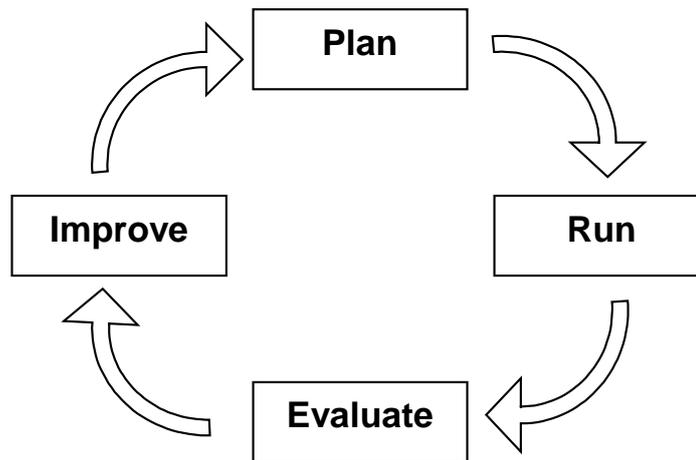
In addition, a fundamental role in quality management is played by the set of quality control actions that should be implemented within the sub-processes to prevent and monitor errors. The strategy could be reported in a quality assurance plan.

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 period⁹.

⁸ ISO 9000:2005, Quality management systems -- Fundamentals and vocabulary. International Organization for Standardization, www.iso.org/iso/home

⁹ A suitable global framework is the National Quality Assurance Framework developed by a global expert group under the United Nations Statistical Commission. See: <http://unstats.un.org/unsd/dnss/QualityNQAF/nqaf.aspx>

Broadening the field of application of the quality management over-arching 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 (figure



2.1).

Figure 2.1. Quality loop

Examples of quality management activities include:

- Setting and maintaining of the quality framework;
- Setting of global quality criteria;
- Setting process quality targets and monitoring compliance;
- Seeking and analysing user feedback;
- Reviewing operation and documenting lessons learned;
- Examining process metadata and quality indicators;
- Internal or external auditing on statistical processes.

Quality management also involves institutional and organisational factors. Such factors are included in other GSBPM over-arching processes (e.g. Human resources management, Statistical programme management) although they can have an impact on quality.

ANNEX 3. METADATA MANAGEMENT

Good metadata management is essential for the efficient operation of statistical business processes. Metadata are present in every phase, either created or carried forward from a previous phase. In the context of this model, the emphasis of the over-arching process of metadata management is on the creation, 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 envisaged in Part A of the Common Metadata Framework¹⁰ "Statistical Metadata in a Corporate Context", that is, that metadata should uniquely and formally define the content and links between objects and processes in the statistical information system.

Part A of the Common Metadata Framework also identifies the following sixteen core principles for metadata management, all of which are intended to be covered in the over-arching Metadata Management process, and taken into the consideration when designing and implementing a statistical metadata system. The principles are presented in four groups:

- | | |
|---------------------------|---|
| Metadata handling | <ul style="list-style-type: none">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 reasonsiv. Versions: Preserve history (old versions) of metadata. |
| Metadata Authority | <ul style="list-style-type: none">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 |

¹⁰ See: <http://www.unece.org/stats/cmf/PartA.html>

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).

ANNEX 4. OTHER USES OF GSBPM

The original aim of the GSBPM was to provide a basis for statistical organisations to agree on standard terminology to aid their discussions on developing statistical metadata systems and processes. However, as the model has developed, it has become increasingly apparent that it can be used for many other purposes, in particular related to modernisation of official statistics. A number of papers describing actual and potential uses of the GSBPM are available on the UNECE wiki platform¹¹. The list below aims to highlight some current uses, and to inspire further ideas on how the GSBPM can be used in practice.

- Harmonizing statistical production architectures - The GSBPM can be seen as a model for an operational view of statistical computing architecture. It identifies the key components of the statistical business process, promotes standard terminology and standard ways of working across statistical business processes. It is a key enabler of the Common Statistical Production Architecture¹².
- Facilitating the sharing of statistical software - Linked to the point above, the GSBPM defines the components of statistical processes in a way that not only encourages the sharing of software tools between statistical business processes, but also facilitates sharing between different statistical organisations that apply the model. It has been used to "classify" software available for sharing in the inventory compiled by the Sharing Advisory Board¹³.
- Describing which standards are or could be used for different phases of the statistical production process. For example, Annex 2 of the SDMX 2.1 User Guide¹⁴ explores how SDMX applies to statistical work in the context of a business process model.
- Providing a framework for process quality assessment and improvement - If a benchmarking approach to process quality assessment is to be successful, it is necessary to standardise processes as much as possible. The GSBPM provides a mechanism to facilitate this.
- Better integrating work on statistical metadata and quality - Linked to the previous point, the common framework provided by the GSBPM can help to integrate international work on statistical metadata with that on data quality by providing a common framework and common terminology to describe the statistical business process.
- Providing the underlying model for methodological standards frameworks - Methodological standards can be linked to the phase(s) or sub-process(es) they relate to and can then be classified and stored in a structure based on the GSBPM.

¹¹ See: <http://www1.unece.org/stat/platform/display/metis/Papers+about+the+GSBPM>

¹² See:

<http://www1.unece.org/stat/platform/display/CSPA/Common+Statistical+Production+Architecture+Home>

¹³ See: <http://www1.unece.org/stat/platform/display/msis/Software+Inventory>

¹⁴ See: http://sdmx.org/index.php?page_id=38



- Providing a structure for documentation of statistical processes - The GSBPM can provide a structure for organizing and storing documentation within an organisation, promoting standardisation and the identification of good practices.
- Providing a framework for building organisational capability - The GSBPM can be used to develop a framework assess the knowledge and capability that already exists within an organisation, and to identify the gaps that need to be filled to improve operational efficiency.
- Providing an input to high-level corporate work planning.
- Developing a business process model repository - Statistics New Zealand has developed a database to store process modelling outputs and allow them to be linked to their statistical business process model. They also plan to develop a Business Process Modelling Community of Practice - i.e. a regular forum to build knowledge of process modelling, to promote their business process model and increase understanding of it, and to discuss process modelling and models as enablers for process improvement.
- Measuring operational costs - The GSBPM can be used as a basis for measuring the costs of different parts of the statistical business process. This helps to target modernisation activities to improve the efficiency of the parts of the process that are most costly.
- Measuring system performance - Related to the point above on costs, the GSBPM can also be used to identify components that are not performing efficiently, that are duplicating each other unnecessarily, or that require replacing. Similarly it can identify gaps for which new components should be developed.

ANNEX 5. RELATIONSHIPS WITH OTHER MODELS AND STANDARDS

The GSBPM was originally developed based on the Generic Business Process Model from Statistics New Zealand, supplemented by input from other statistical organisations with experience of statistical process modelling. However, a number of other related models and standards exist, both as a result of modernisation initiatives, and similar activities outside the scope of official statistics.

CSPA

Turning first to the links between the GSBPM and the other frameworks and standards needed for statistical modernisation, it can be helpful to consider them in the context of enterprise architecture. Enterprise architecture relates business functions and processes to the information, applications and technology needed to run them. It is a way of describing what an organisation does, and how it does it, to try to identify how the organisation could improve quality and efficiency.

In the context of statistical modernisation, the aim is to align the enterprise architectures of different organisations, creating an “industry architecture” for the whole “official statistics industry”. This approach is intended to facilitate collaboration, sharing and joint development of the components and services that are needed for the different parts of the statistical business process (defined in relation to the GSBPM). The result is the Common Statistical Production Architecture (CSPA), first released at the end of 2013¹⁵.

GSIM

The Generic Statistical Information Model (GSIM) is a reference framework for statistical information, designed to play an important part in modernising and streamlining 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 in the design and production of statistics. 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. The GSIM is designed to allow for innovative approaches to statistical production to the greatest extent possible; for example, in the area of dissemination, where demands for agility and innovation are increasing. It also supports current approaches of producing statistics.

¹⁵ See:

<http://www1.unece.org/stat/platform/display/CSPA/Common+Statistical+Production+Architecture+Home>

The GSIM identifies around 110 information objects, examples include data sets, variables, statistical classifications, units, populations as well as the rules and parameters needed for production processes to run (for example, data editing rules).

The GSIM and the GSBPM are complementary models for the production and management of statistical information. As shown in the diagram below, the GSIM helps describe 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 GSIM (figure 5.1.).



Figure 5.1 The GSIM and the GSBPM as complementary models

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. Similarly, both models support the implementation of CSPA, but can be applied regardless of whether that architectural framework is used or not.

In the same way that individual statistical business processes do not use all of the sub-processes described within the GSBPM, it is very unlikely that all information objects in the GSIM will be needed in any specific statistical business process.

Applying the GSIM and GSBPM together can facilitate the building of efficient metadata driven systems, and help to harmonise statistical computing infrastructures.

GLBPM

Looking outside the domain of official statistics, the social survey research community has developed the Generic Longitudinal Business Process Model (GLBPM)¹⁶ "to provide a generic model that can serve as the basis for informing discussions across organisations conducting longitudinal data collections, and other data collections repeated across time".

Like the GSBPM, the GLBPM is a reference model against which actual business processes can be mapped. It is a reference model of the process of longitudinal and repeat cross-sectional data collection for research organisations. It describes the activities undertaken and maps these to their typical inputs and outputs.

¹⁶ See: <http://dx.doi.org/10.3886/DDILongitudinal05>



The GLBPM has many similarities to the GSBPM, although it differs in some specific activities reflecting the different needs and practices of the statistical and social science research communities. The GLBPM, takes the approach of having a non-linear path through a matrix of alternatives directly from the GSBPM.



ANNEX 6: ABBREVIATIONS AND ACRONYMS

List of abbreviations and acronyms, included in the report.

CMF - Common Metadata Framework: A set of resources relating to the use of metadata by statistical organisations, including information on standards and best practices. See <http://www.unece.org/stats/cm/>.

CSPA - Common Statistical Production Architecture: An industry architecture which brings together the GSBPM and the GSIM, in addition to new frameworks about Statistical Services to create an agreed top level description of the 'system' of producing statistics which is in alignment with the modernisation initiative.

DDI - Data Documentation Initiative: An international standard for describing data from the social, behavioural, and economic sciences.

GLBPM - Generic Longitudinal /Business Process Model: A model based on the GSBPM, developed by the social survey research community.

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

GSIM - Generic Statistical Information Model: A reference framework of information objects, which enables generic descriptions of the definition, management and use of data and metadata throughout the statistical production process.

HLG - The High-Level Group for the Modernisation of Statistical Production and Services

INSTAT - Institute of Statistics of Albania

METIS - The "brand name" for work on Statistical Metadata under the Conference of European Statisticians.

OECD - Organisation for Economic Cooperation and Development

SDMX - Statistical Data and Metadata eXchange: A set of technical standards and content-oriented guidelines, together with an IT architecture and tools, to be used for the efficient exchange and sharing of statistical data and metadata.

UNECE - United Nations Economic Commission for Europe

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



ANNEX 7: REFERENCES

Adapted Global Assessment of the National Statistical System of Albania, 2010

Applying the Generic Statistical Business Process Model to business register maintenance
<http://www1.unece.org/stat/platform/download/attachments/57835551/BR+meeting+paper2.doc>

Generic Statistical Business Process Model, GSBPM, (Version 5.0, December 2013), United Nations Economic Commission for Europe (UNECE)
<http://www1.unece.org/stat/platform/display/GSBPM/GSBPM+v5.0>

IPA 2014 Multi-beneficiary statistical cooperation programme, MISSION REPORT - DRAFT, Pilot Project 18, Developing a Metadata System, May, 2016

ISO 9000:2005, Quality management systems -- Fundamentals and vocabulary.
International Organization for Standardization, www.iso.org/iso/home

Light Peer Review of the Implementation of the European Statistics Code of Practice in the Republic of Albania - Final Report, 2014

National Implementations of the GSBPM – A Summary Based on METIS Case Studies, Prepared by Thérèse Lalor, Australian Bureau of Statistics,
<http://www1.unece.org/stat/platform/display/GSBPM/>

Official Statistics Programme for the period 2012-2016, INSTAT, www.instat.gov.al

Report on Limited Peer Review of the Institute of Statistics of Albania, 2015

Links:

www.unece.org/stats/gsbpm

<http://www.unece.org/stats/cmf/>

<http://www1.unece.org/stat/platform/display/GSBPM/GSBPM+v5.0>

<http://www1.unece.org/stat/platform/display/hlqbas>

<http://www1.unece.org/stat/platform/download/attachments/57835551/BR+meeting+paper2.doc>

<http://unstats.un.org/unsd/dnss/QualityNQAF/nqaf.aspx>

<http://www.unece.org/stats/cmf/PartA.html>

<http://www.instat.gov.al>

<http://www1.unece.org/stat/platform/display/metis/Papers+about+the+GSBPM>



<http://www1.unece.org/stat/platform/display/CSPA/Common+Statistical+Production+Architecture+Home>

<http://www1.unece.org/stat/platform/display/msis/Software+Inventory>

http://sdmx.org/index.php?page_id=38

<http://www1.unece.org/stat/platform/display/CSPA/Common+Statistical+Production+Architecture+Home>

<http://dx.doi.org/10.3886/DDILongitudinal05>